Title	Semiparametric Transformation for Non-Linear Regression Model
Author(s)	伊藤, 雅憲
Citation	
Issue Date	
oaire:version	VoR
URL	https://hdl.handle.net/11094/2034
rights	
Note	

Semiparametric Transformation for Non-Linear Regression Model

Masanori Ito

MARCH 2009

Semiparametric Transformation for Non-Linear Regression Model

A dissertation submitted to
THE GRADUATE SCHOOL OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN ENGINEERING

BY

Masanori Ito

MARCH 2009

謝辞

本論文の作成におきましては，多くの方々にご指導・ご支援を頂戴いたしました．ここに，深くお礼を申しあげます.

指導教官の白旗慎吾先生には，本論文の全体を通して，多大なご教示をいただきました．いつも優しいお言葉で，かつ深いご経験に基づいた含蓄のあるご指導をいただきました．諸種の会合でお会いしたときには，いつも場を和ませる暖かいお話を聴かせていただきました．心よりお礼を申しあげるとともに，今後もますますのご高配のほど，よろしくお願い申しあげます．大阪大学教授（大学院基礎工学研究科 数理計量ファイナンス講座）の長井英生先生には，本論文の初稿をお受け取りいただき，丁寧に査読いただきました．誠にありがとうございました．大阪大学教授（大学院基礎工学研究科 統計数理講座）の狩野 裕先生には，本論文の初稿を査読いただき，また公聴会では貴重なご指摘を賜りました．ご指摘のなかには理論的なものだけでなく，データの解釈についての実践的なお話もあり，今後の研究に対して大変に刺激を受けました．大阪大学准教授の坂本 亘先生には，本論文の構成から研究内容の詳細まで，多大なご指導を賜りました．とくに，ご自身のご研究の成果のうちの一つをご伝授いただき，さらに適用の際の留意点を丁寧にご指導いただきました．本研究におけるノンパラメトリック変換の平滑化パラメータの選定問題について一つの解決策を提示することができたのは，坂本先生のご指導の賜物でございます，心よりお礼を申しあげます.

本論文の作成にあたり，医学統計研究会（通称，BRA）の皆様には多大なご支援と心温まるご配慮をいただきました．後藤文彦さん（MPC 株式会社社長）には，お酒の席をご一緒させていただいた折には,「(学位を）はやくとりなさい」との激励のお言葉をいつも頂戴いたしました．イーピーエス株式会社　魚井 徹博士には，学会や懇親会等でお会いした際に，筆者の発表内容に対するご指摘や，励ましのお言葉を頂戴いたしました．臨床研究情報センター [財団法人 先端医療振興財団] の松原義弘博士には，遊学の場面でお会いした際にはいつも煙草をお供に，有益なお話をお聴きすることができました．フィールドワークス（株）の木田義之さんには，BRA 東京定例会後の課題検討会等でお会いした際に，幾度も激励のお言葉をおかけいただき，またご自身のご趣味（ゴルフ）のお話でいつも盛り上がりました．株式会社ソリューションラボの志賀 功さんには，筆者が大分に訪れた際にはいつも大変にお世話になりました．ご自身の会社に連れて行っていただき，

医学統計学習塾の開発についての討議に参加させていただき，大変に啓発されたこともございました．(株）富士通ソフトウェア・ラボラトリの衛藤俊寿博士には，大分統計談話会の折にはいつもお世話になり，発表後の筆者をねぎらってくださいました．ファイザー株式会社の栗林和彦博士には，BRA の諸種の会合や計算機統計学会スタディ・グループを通して，幾度も貴重なコメント・ご指導をいただきました．ご自身の業務および学問に対する姿勢は筆者にとってあこがれであり，大変に啓発されました．第一三共株式会社の佐藤俊之博士には，BRA 諸会合でお会いするたびに，勇気付けていただきました．ご自身が，学会やシンポジウムの仕掛け人として活躍されるお姿に，大変に啓発されました．小野薬品工業株式会社の冨金原 悟博士には，学会の終了後やご自身が上京された折など，筆者にご馳走していただきました．遊学ともに，大変にお世話になりました．ファイザー株式会社の河合統介博士には，公私を問わず，様々な遊学の場面で数えきれないほどのご助言，叱咤激励をいただきました．いつも，筆者が困ったり，迷ったりした際には，メールしたり，お酒の席でご相談をいたしましたが，河合さんは，ときには厳しい叱責で，ときには心温まるご助言で，必ず応えてくださいました，心より感謝いたします．あすか製薬（株）の藤澤正樹博士には，ご自身が学位論文を作成された折のご経験に基づいて，筆者に幾度も貴重なアドバイスをしてくださいました．ご自身が，BRA および大愚の会の運営にかける姿勢，またその姿勢に対して多くのお仲間の方々が賞賛されていることを鑑みて，いつも尊敬の念を抱いておりました．エーザイ（株）の高瀬貴夫さんには，公私を問わず，色々とご相談に乗っていただき，貴重なご助言を沢山いただきました．協和発酵キリン（株）の古川泰伸さんには，とくに BRA 東京定例会の折に，貴重なコメントを幾度もいただきました．ファイザー株式会社の山邉太陽さんには，学問のアドバイスだけでなく，勤行（業務）のことについても先見性のあるお話を聴かせていただきました．ノバルティス株式会社の尼ヶ崎太郎さんには，ご自身の研究への姿勢，とくに数理的な問題に真正面から挑む所に啓発されました．これらの方々に重ねてお礼を申しあげます．

長崎大学教授の柴田義貞先生には，BRA の諸会合や業務でお話する機会を通して，大変に啓発されました．学会や研究会での鋭いご発言は，その内容がいつもぶれずに一貫しておられ，科学者としてのご姿勢を崩されない雰囲気が，筆者にとってはあこがれでした．お酒の席であっても，筆者の拙い質問にいつも真剣に応えてくださいました．期待値の概念についての分かり易い事例を交えてのお話は，いまでも強く印象に残っております．大分大学教授の越智義道先生には，BRA 年会・納会および大分統計談話会等でお会いした際に，筆者の発表内容についてご助言をいただきました．とくに，筆者が修士課程の頃よ

り事例研究を行っていた「スキーナ川の鮭データ」を未だに解剖していることについて，「データ解析においては，しつこさの姿勢が大切です」と激励いただいたことが，最も心に残っております．鹿児島高等専門学校教授の藤崎恒晏先生には，BRA の会合，日本計算機統計学会のシンポジウムや大会，大分統計談話会の会合でお会いするたびに元気づけていただき，温かく接してくださいました．大阪大学准教授の濱﨑俊光先生には，本論文の中心となった筆者の計算機統計学会欧文誌への投稿論文について丁寧に査読いただき，研究内容だけでなく英語の表現についても厳しくご指導をいただきました．また，統計的変換論に纏わる新しいアイデアや，それに伴う成書についてもお貸しいただき，大変にお世話になりました．大阪大学助教の杉本知之先生には，BRA の諸会合でお会いするたびに温かく勇気づけてくださいました．ご自身の，研究に対する姿勢および教育への心構えには，大変に啓発されました．兵庫医科大学講師の大門貴志先生には，学会などでお会いするたびにお声をかけていただき，励ましのお言葉をいただきました．国際学会などでご自身の迫力のあるご発表を拝聴し，大変に刺激を受けました．山梨大学 助教の下川敏雄先生には，いつも統計科学のホット・トピックについて聴かせていただき，大変に勉強になりました．ご自身の，色々な分野に挑戦していくお姿をみて啓発されました．

ノバルティス株式会社の池田公俊さんには，研究内容についてご助言いただいただけでなく，筆者の急な遊学の誘いに快くおつき合いいただきました．アスビオファーマ株式会社の永久保太士さんには，研究の話につき合っていただいただけでなく，筆者をねぎらうために楽しいお酒の場を提供してくださいました．大阪大学 大学院基礎工学研究科 博士前期課程 2 年の中村将俊さん，大江基貴さん，五十川直樹さんには，筆者が帰阪した際には色々とお世話いただきました．お酒の席では，若々しく，楽しいお話で場を盛り上げてくださいました．

アステラス製薬株式会社 開発本部 データサイエンス部 部長の向井満利さんには，大阪大学 大学院基礎工学研究科 博士後期課程（社会人コース）に入学する機会を与えていただき，また入学してからも様々なご配慮をしてくださいました．同部 次長の廣岡秀樹さんには，本研究を進めるにあたって，業務との両立について諸種のご配慮をいただき，国際学会発表などの沢山の機会を与えてくださいました．2 人で休日出勤をしていた折に，お休み中であるにもかかわらず，筆者の薬物動態モデルに基づく事例研究の場面において貴重なご助言をいただき，ご自身で即座にプログラムを作成くださり，筆者の数値結果の検証をいただいたことが一番心に残っております．お 2 人のご配慮に心よりお礼を申しあげます．

医学統計研究会の理事長である後藤昌司先生には，本研究の主題をご提示いただくとともに，本論文の全体およびそれに纏わる公表論文作成，学会発表などのすべてについて本当にたくさんのことを教えていただきました．後藤先生には，学問だけでなく,「掃除・勤行・学問」の実践の教えに基づいて熱意あるご指導をいただき，筆者の社会人として欠けている部分を徹底的に直していただき，さらに筆者の「人間的魅力の醸成」においても大変に貴重なご指導をいただきました．国際学会やシンポジウムの折には様々な「遊学」に同行させていただき，見聞を深めることができました．本論文の執筆の過程で，筆者の研究に対する「甘さ」がみられた折には，厳しく叱咤してくださいました．修士課程の研究でご指導を頂戴してから数えて 9 年間，教えていただいた沢山の貴重な財産を糧として，今後も歩みを進めたく存じております．後藤先生とお会いしてから筆者の人生は大きく変わりましたが，9 年前の「善縁」は筆者の強運であったと深く感じております．今までのご指導に心より感謝いたします，ありがとうございました．

筆者は，博士後期課程（社会人コース）の 3 年間にわたり，アステラス製薬株式会社から経済的な援助を受けました．ここにその援助に対し深謝いたします．

最後に絶えず，著者の身を心配し，励ましてくれた妻と両親に感謝いたします．

Abstract

All phenomena in the natural world occur as a consequence of intertwining of many factors in the background. A *system* can be seen a kind of operator which gives a signal a certain action in a certain target, and a function to create output from a kind of input. The function that relates the output to the input obtained by this formulation is called a *model*. The theoretical model, which is built based on existing theories and knowledge, has deviation from observed data because it does not consider the generation-mechanism of data. A common approach to deal with errors is the power-transformation approach. For nonlinear regression models, we can use the Power Transform-Both-sides (PTB) approach. This approach tries to achieve normality and homoscedasticity of the error by transformation. However, it is difficult to achieve these two aims simultaneously by a power transformation with one transformation parameter like PTB. In particular, PTB is insufficient to stabilize the error variance. Then, we suggested the Power Transform-Both-sides and Weighted Least Squares (PTBWLS) approach that implements a power-weighted transformation (PWT) to PTB. The most important problem of the above parametric transformation approaches is that they are too sensitive to data. To tackle this problem we provided the Nonparametric Transform-Both-sides (NTB) approach, which uses a cubic spline curve as a transformation function. The spline function in this approach is identified by maximizing the penalized likelihood. Furthermore, combining PTBWLS with NTB together, we proposed the Nonparametric Transform-Both-sides and Weighted Least Squares (NTBWLS) approach. The NTBWLS is designed to implement both nonparametric estimation of the transformation function and parametric estimation of the power-weighted function. We conducted some case studies, a numerical investigation in which data were generated from a 1-compartment model, and a couple of simulation experiments. From these results, we concluded that NTBWLS is superior to the other existing approaches in the situation where data have problematic heteroscedasticity and non-normality.

Notations

Notations	Definitions and examples	Remarks
General		
$Y = f(\boldsymbol{X};\boldsymbol{\beta}) + \varepsilon$		Non-linear model
Y		Response
y_n		The nth observation of response
$\boldsymbol{X}$	$\boldsymbol{X} = (X_1, ..., X_{p_0})^{\mathrm{T}}$	
Predictor vector $\boldsymbol{x}_p$ $(p = 1, 2, ..., p_0)$ with n observations		The $n \times 1$ vector of the pth predictor vector
$\boldsymbol{\beta}$	$\boldsymbol{\beta} = (\beta_1, ..., \beta_I)^{\mathrm{T}}$	$I \times 1$ parameter vector
$f(\boldsymbol{X};\boldsymbol{\beta})$		Non-linear function
ε		Error
$L(\cdot)$	$L(\boldsymbol{\beta}, \sigma)$	Likelihood function
$l(\cdot)$	$l(\boldsymbol{\beta}, \sigma)$	Log-likelihood function
Distributions		
$\mathrm{N}(\cdot, \cdot)$	$\mathrm{N}(0, \sigma^2)$	Normal distribution
$\mathrm{PND}(\cdot, \cdot, \cdot)$	$\mathrm{PND}(\lambda, \mu, \sigma^2)$	Power normal distribution
$\mathrm{E}(\cdot)$	$\mathrm{E}(Y)$	Expectation
$\mathrm{Var}(\cdot)$	$\mathrm{Var}(Y)$	Variance
ψ		Standard normal probability density function
Ψ		Cumulative distribution function of ψ
PTB		
$H_{\mathrm{P}}(\cdot, \cdot)$	$H_{\mathrm{P}}(Y, \lambda)$	Power transformation function
λ		Power transformation parameter
ϕ		Power weighted parameter
ε_{P}		Error on power transforming both-sides
NTB		
$H_{\mathrm{S}}(\cdot)$	$H_{\mathrm{S}}(u)$	Smooth transformation function
$h_{\mathrm{S}}(\cdot)$	$h_{\mathrm{S}}(u)$	Log-derived function of $H_{\mathrm{S}}(\cdot)$
$J(\cdot)$	$J(H_{\mathrm{S}}(u))$	Roughness penalty
ρ		Smoothing parameter
$L_{\mathrm{P}}(\cdot)$	$L_{\mathrm{P}}(\boldsymbol{\beta}, \sigma, h_{\mathrm{S}}(u))$	Penalized likelihood
$l_{\mathrm{P}}(\cdot)$	$l_{\mathrm{P}}(\boldsymbol{\beta}, \sigma, h_{\mathrm{S}}(u))$	Penalized log-likelihood
$H_{\mathrm{K}}(\cdot)$	$H_{\mathrm{K}}(t)$	Kernel function
$g(\cdot)$	$g(u)$	Probability density function
w		Band width

Abbreviations

OLS	Ordinary Least Squares
PTB	Power Transformation-Both-sides
NTB	Nonparametric Transformation-Both-sides
PWT	Power-Weighted Transformation
PTBWLS	Power Transformation-Both-sides and Weighted Least Squares
NTBWLS	Nonparametric Transformation-Both-sides and Weighted Least Squares
SRC	Spearman Rank Correlation

Acknowledgment

The author would like to acknowledge the advice and the kind helps of many people to the completion of this thesis. First the author wishes to thank Professor Shingo Shirahata of Osaka University, who has provided adequate and useful suggestions throughout writing this thesis. The author is also deeply grateful to Professor Hideo Nagai and Professor Yutaka Kano of Osaka University for reviewing the first draft of the thesis carefully and providing valuable suggestions. Associate Professor Wataru Sakamoto of Osaka University has kindly clarified his thinking and given him helpful ideas on numerous occasions. Especially, he has provided many insights into the selection of the smoothing parameter.

The author wishes to thank the following members of the Biostatistical Research Association (BRA) because they have provided valuable motivation, feedback, and ideas: Mr. Fumihiko Goto, Dr. Tohru Uwoi, Dr. Yoshihiro Matsubara, Dr. Kazuhiko Kuribayashi, Dr. Norisuke Kawai, and Dr. Masaki Fujisawa. The author shows his special gratitude to Professor Yoshisada Shibata of Nagasaki University and Professor Yoshimichi Ochi of Oita University for giving valuable comments and suggestions. The author also expresses his gratitude to Associate Professor Toshimitsu Hamasaki and Dr. Tomoyuki Sugimoto of Osaka University for reading the publications and giving helpful suggestions. The author wishes to thank Dr. Jun Takeda, who has given him many of the wonderful ideas.

The author has been able to devote so much time to his thesis because of the cooperation of his superiors, Vice President Mr. Mitsutoshi Mukai and Senior Director Mr. Hideki Hirooka, in Astellas Pharma Inc.

Dr. Masashi Goto, who is the representative of the Biostatistical Research Association (BRA), has provided precious advise not only on the research but on knowledge for living as members of society. The author would especially like to thank him for giving valuable ideas and motivations for this work.

Finally the author is grateful to his wife and parents for warm-hearted encouragement and kind support.

Contents

1. Introduction

1.1 Background

All phenomena in the natural world occur as a consequence of intertwining of many factors in the background. It is difficult individually to identify and to interpret all these factors on the process of investigating the mechanism where the phenomena is generated. Then, it is tried usually to assume or to remove only the main factor and to simplify the phenomenon. The selection of the factor and the process of compression are included there. As a result, even if the phenomenon will not be completely described, "Mechanism (system)" of substitution that approximately simplifies the phenomenon is composed. That is, in the science field, the system is composed by the representative characteristic that controls a peculiar theory. A system can be seen a kind of operator which gives a signal a certain action in a certain target, and a function to create output from a kind of imput. That is, it is a process of conversion from the input to the output (Howard, 1963; Ohta *et al.*, 1968; Goto *et al.*, 1968a and 1968b). Therefore, it is substituted to clarify the phenomena or the structure in the background by two signals of the process of conversion in the system, that is, constructing the relation between the input and the output. It is called "System identification" (Kume, 1971) and actually the system is formulated by expressing it in the form of any functions for the relation between two signals. The function that relates the output to the input obtained by this formulation is called a model. That is, obtaining the model is intended with a system identification. However, It is necessary for the output to predict or to control using the model in a statistical science. It can be thought that the input is an explanation factor (variable) to complicate the phenomena, and the output is a response (data) obtained by observing the phenomena based on a statistical perspective. In this context, The model obtained by identifying the system will work as a tool of the prediction and the control in practice by moving forward with a phased approach through the process of a statistical inference, evaluation and diagnosis. However, it is more common that it is difficult to identify the system. In this situation, it is either whether to obtain a principle experienced to achieve the result of a priori inference on the theoretical research or repeat the phenomenological observation by the experiment and the observation. In this paper, we focus on the former, that is, a statistical inference on the theoretical model.

1.2 Objectives of our studies

Important objectives of regression analysis are ordinarily 1) the prediction of the response variable with variability of the exploratory factors based on the model, 2) the control of the response variable by handling the exploratory factors, and 3) the calibration of the exploratory factors corresponding to controlling the response variable. We deal with the theoretical model as the model which gives a relationship between the response and the exploratory variables. The theoretical model is derived by the system based on a characteristic theory in a science field, and has an almost of that complex and non-linear structure. However, the theoretical model has the unbridgeable gaps between actual phenomenon and the model because the theoretical model only approximates simplification of the system even if composed exquisitely to adjust to a characteristic phenomenon the theoretical model. Also, the theoretical model, which is built based on existing theories and knowledge, has deviation from observed data because it does not consider the generation-mechanism of data (Goto, 1974; Goto and Daimon, 2000). That is, it is "Error" that shows the gap between data and the model to perform a big role in the inference on theoretical model.

The objectives in this paper are to design the error of the theoretical model statistically, and to provide "Bridge" between the model and data. In other words, it is to satisfy the symmetry of the error distribution (if possible, normality) and the homoscedasticity of the error. A common approach to deal with errors is the power-transformation approach. The power-transformation approach for a response has easiness of the interpretation and flexibility of application as inclusion type of the log-transformation, so it exists as a typical approach for an inference on the linear models (Box and Cox, 1964; Atkinson, 1985; Goto *et al.*, 1991). For non-linear regression models, we can use the Power Transform-Both-sides (PTB) approach which has been proposed by Carroll and Ruppert (1984). This approach is to transform both the response and predictive function (non-linear function expressed by some predictors and parameters) while paying attention to the immutability of the model before and after transformation. This approach tries to achieve normality and homoscedasticity of the error by transformation. However, it is difficult to achieve these two aims simultaneously by a power transformation with one transformation parameter like PTB (Goto, Inoue and Tsuchiya, 1987 : Goto, 1992, 1995 : Goto, Isomura and Hamasaki, 2000). In particular, PTB is insufficient to stabilize the error variance (Carroll and Ruppert, 1988). Goto (1992) provides three types of double power-transformation approaches and clarifies the assumptions and objectives of the transformations (see also Goto (1995) and Goto *et al.*(2000)). The Double Power Weighted Transformation (DPWT) involves two separate transformation parameters, namely, one is the parameter to induce the normality of the errors and the other is to estimate an appropriate weight which stabilizes the error variance. Then, we suggest the Power Transform-Both-sides and Weighted Least Squares (PTBWLS) approach as an analogy of DPWT. PTBWLS implements a power weighted transformation (PWT) provided by Box and Hill (1974) to PTB.

The most important problem of the above parametric transformation approaches is that they are too sensitive to data. To tackle this problem we provide the Nonparametric Transform-Both-sides (NTB) approach, which uses a cubic spline curve as a transformation function. It has been discussed by Nychka and Ruppert (1995) and Ito and Goto (2004). In the research of Ito and Goto (2004), we introduced NTB as an alternative approach of the PTB, in the inference of theoretical models. As for the estimation of the parameters in the theoretical models, we presented the method which represents the function of one of the methods of the transformation by the cubic spline curve. From the investigation of two examples, we suggested that the NTB could be an index for the validation of the PTB and was more robust than PTB for outliers. Furthermore, we verified these results by three simulation experiments. In the methodology for fitting of the empirical model, we introduced Alternating Conditional Expectation (ACE) provided by

Breiman and Friedman (1985) and Additivity VAriance Stabilization (AVAS) provided by Tibshirani (1988) as two nonparametric transformation approaches that optimize relationship between the response and explanatory variables. We examined the validity of the theoretical models by fitting the empirical models via ACE and AVAS to the example data. As a result, both methods of ACE and AVAS improved the normality and homoscedasticity of the error.

In NTB, the spline function is identified by maximizing the penalized likelihood. Furthermore, combining PTBWLS with NTB together, we propose the Nonparametric Transform-Both-sides and Weighted Least Squares (NTBWLS) approach. The NTBWLS is designed to implement both nonparametric estimation of the transformation function and parametric estimation of the power-weighted transformation function. In the research of Ito and Goto (2006), through the numerical investigation of one example using data generated from a non-linear model, we conclude that PTB and PTBWLS induce normally distributed additive errors and stabilize the error variance, and NTBWLS improves the degrees of normality and homoscedasticity of the error more than PTB and PTBWLS. However, There were problems for the identification of the optimal nonparametric transformation function in NTBWLS. In the estimation of the spline function, it is need to choose a appropriate value for the smoothing parameter based on a given data. One computationally intensive strategy is to estimate the smoothing parameter on the basis of cross-validation. However, the idea of cross-validation is to optimize on predicting responses, which does not match to a primary objective of the nonparametric regression (Sakamoto, 2007). In this paper, we use a maximizing marginal likelihood approach to select the smoothing parameter. The smoothing parameter, which govern global nonlinear regression structure, are estimated with the maximum marginal likelihood estimation, or the empirical Bayes method.

1.3 Outline of datasets

In this section, we show some data of case studies used to investigate in later section.

[**Data set No.1: Shortleaf pine data**($N = 70$): Bruce and Schumacher, 1935]

The girth and to a lesser extent the height, are easily measured, but it is the volume of usable timber that determines the value of a tree. The aim is therefore to find a formula for predicting volume from the other two measurements. Table 1.1 contains 70 observations on the volume in cubic feet of shortleaf pine, from Bruce and Schumacher (1935) together with x_1, the girth of each tree, that is, the diameter at breast height, in inches and x_2, the height of the tree in feet. Atkinson and Rinai (2000) suggests a conical model

$$f(\boldsymbol{x};\beta_1) = \beta_1 x_1^2 x_2. \tag{1.1}$$

They use PTB approach for fitting the conical model and calculate scoring test statistics on the null hypothesis $\mathrm{H}_0 : \beta_1 = 0$ in order to investigate a sensitivity for the estimates of the model parameters. Finally, it is selected for power-transformation to handle log-transforming as giving a good result.

[**Data set No.2: Skeena salmon data**($N = 70$): Bruce and Schumacher, 1935]
Ricker and Smith (1975) give numbers of spawners and recruits from 1940 until 1967 for the Skeena River sockeye salmon stock. Their data are given in Table 1.2. Let x denote the number of spawning salmon in a given year and let y be the number of recruited salmon associated with the same year. Ricker (1954) derived the theoretical deterministic model

$$f(x;\boldsymbol{\beta}) = \beta_1 x \exp(-\beta_2 x) \tag{1.2}$$

Ricker's model is widely used for salmon stocks and appears to fit them well. This function is taken to be the parametric regression function for the median of the distribution of recruited salmon given a particular number of spawning fish. A scatter plot of these data suggest that, although the Ricker model is a reasonable choice for the median response, the variance of recruit salmon does not appear to be constant and the response is right skewed. A second model was derived by Beverton and Holt (1957), namely

$$f(x;\boldsymbol{\beta}) = \frac{1}{\beta_1 + \beta_2/x}, \quad \beta_1 \geq 0, \quad \beta_2 \geq 0. \tag{1.3}$$

When fit to the same dataset, the Ricker and Beverton-Holt models are often similar over the range of spawner values in the data, despite qualitatively different behavior as the

Table 1.1: Shortleaf pine data

number	volume (y)	girth (x_1)	height (x_2)	number	volume (y)	girth (x_1)	height (x_2)
1	4.6	33	2.2	36	11.0	71	25.8
2	4.4	38	2.0	37	11.1	81	32.8
3	5.0	40	3.0	38	11.2	91	35.4
4	5.1	49	4.3	39	11.5	66	26.0
5	5.1	37	3.0	40	11.7	65	29.0
6	5.2	41	2.9	41	12.0	72	30.2
7	5.2	41	3.5	42	12.2	66	28.2
8	5.5	39	3.4	43	12.2	72	32.4
9	5.5	50	5.0	44	12.5	90	41.3
10	5.6	69	7.2	45	12.9	88	45.2
11	5.9	58	6.4	46	13.0	63	31.5
12	5.9	50	5.6	47	13.1	69	37.8
13	7.5	45	7.7	48	13.1	65	31.6
14	7.6	51	10.3	49	13.4	73	43.1
15	7.6	49	8.0	50	13.8	69	36.5
16	7.8	59	12.1	51	13.8	77	43.3
17	8.0	56	11.1	52	14.3	64	41.3
18	8.1	86	16.8	53	14.3	77	58.9
19	8.4	59	13.6	54	14.6	91	65.6
20	8.6	78	16.6	55	14.8	90	59.3
21	8.9	93	20.2	56	14.9	68	41.4
22	9.1	65	17.0	57	15.1	96	61.5
23	9.2	67	17.7	58	15.2	91	66.7
24	9.3	76	19.4	59	15.2	97	68.2
25	9.3	64	17.1	60	15.3	95	73.2
26	9.8	71	23.9	61	15.4	89	65.9
27	9.9	72	22.0	62	15.7	73	55.5
28	9.9	79	23.1	63	15.9	99	73.6
29	9.9	69	22.6	64	16.0	90	65.9
30	10.1	71	22.0	65	16.8	90	71.4
31	10.2	80	27.0	66	17.8	91	80.2
32	10.2	82	27.0	67	18.3	96	93.8
33	10.3	81	27.4	68	18.3	100	97.9
34	10.4	75	25.2	69	19.4	94	107.0
35	10.6	75	25.5	70	23.4	104	163.5

Table1.2: Skeena salmon data

Year	Spawners (x)	Recruits (y)	Year	Spawners (x)	Recruits (y)
1940	963	2215	1954	511	1393
1941	572	1334	1955	87	363
1942	305	800	1956	370	668
1943	272	438	1957	448	2067
1944	824	3071	1958	819	644
1945	940	957	1959	799	1747
1946	486	934	1960	273	744
1947	307	971	1961	936	1087
1948	1066	2257	1962	558	1335
1949	480	1451	1963	597	1981
1950	393	686	1964	848	627
1951	176	127	1965	619	1099
1952	237	700	1966	397	1532
1953	700	1381	1967	616	2086

(Units are thousands of fish)

number of spawners increases to infinity.

[**Data set No.3: Acetaminophen data**($N = 13$): Channer and Roberts, 1985]
Channer and Roberts (1985) studied the effect of delayed esophageal transit on the absorption of acetaminophen. Patients awaiting cardiac catheterization took a single 500-milligram tablet containing acetaminophen and barium sulfate. Table 1.4 lists the average plasma acetaminophen data obtained 6 hr after swallowing the tablet. The blood drug concentration in the systematic circulation compartment (non-linear predictive function) is

$$f(t;\boldsymbol{\beta}) = \frac{500K_{12}}{\mathscr{V}_1(K_{12}-K_{20})}\left\{\exp(-K_{20}t)-\exp(-K_{12}t)\right\}, \tag{1.4}$$

where t is the time following administration, $\mathscr{V}_1$ is the volume of distribution, K_{12} is the first-order absorption rate constant, K_{20} is the first-order elimination rate constant and $\boldsymbol{\beta} = (\mathscr{V}_1, K_{12}, K_{20})^{\mathrm{T}}$. We generate random numbers for the parameter estimation of the 1-compartment model in the example data. The goals are to assess how much each method can improve non-normality and heteroscedasticity.

Table1.3: Average plasma acetaminophen data

Time (min)	Concentration (mg/l)
0	0
10	2.1
20	5.6
30	5.8
40	6.3
50	4.7
60	4.1
90	3.5
120	2.8
150	2.2
180	1.7
210	1.8
240	1.5
360	0.75

2. Various Types of Approaches for Inference on Models

In this chapter, we suggest some approaches for fitting theoretical models. First we introduce the standard approaches. Then, we expand it to some parametric transformation approaches. Next, as an alternative to the parametric approaches, we provide the two nonparametric transformation approaches. Finally, we propose the semiparametric transformation approach, which designed to implement both nonparametric estimation of the transformation function and parametric estimation of the power-weighted transformation function.

2.1 Inference on theoretical models

In general, a non-linear regression model can be expressed by

$$Y = f(\boldsymbol{X}; \boldsymbol{\beta}) + \varepsilon, \tag{2.1}$$

where $\boldsymbol{X}$ is the $p_0 \times 1$ predictor vector X_p ($p = 1, 2, ..., p_0$), and $f(\boldsymbol{X}; \boldsymbol{\beta})$ is the known function with the parameter β_i ($i = 1, 2, ..., I$) ($f(\boldsymbol{X}; \boldsymbol{\beta}) > 0$), ε is the error to be normally distributed with zero mean. Y is the positive response (random variable) corresponding to $f(\boldsymbol{X}; \boldsymbol{\beta})$.

2.1.1 Standard approaches

We assume that the observations $(\boldsymbol{x}_n, y_n)$ ($n = 1, 2, ..., N$) are given. The ordinary least squares (OLS) approach is often used by estimating $\hat{\beta}$ regardless of the linearity of the model. If we have $\boldsymbol{b} = (b_1, ..., b_I)$ as any estimate of $\boldsymbol{\beta}$, we can set $\boldsymbol{b}$ which satisfied

$$SSE(\boldsymbol{b}) = \sum_{n=1}^{N} \{y_n - f(\boldsymbol{x}_n; \boldsymbol{b})\}^2 \tag{2.2}$$

and

$$\frac{\partial SSE(\boldsymbol{b})}{\partial b_i} = 0, \quad i = 1, ..., I$$

as $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, then this is a least square estimate of $\boldsymbol{\beta}$. The error is distributed as normal in above assumption, so $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ consists with the maximum likelihood estimate. It is usually calculated by Gauss-Newton algorithm based on approximation of Taylor expansion, because $f(\boldsymbol{X}; \boldsymbol{\beta})$ is a non-linear function for $\boldsymbol{\beta}$ thus it is difficult that we derive $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ analytically.

Another standard approach is a maximum likelihood estimation method. In the model (2.1), if the simultaneous distribution of the error is known, the maximum likelihood estimate of $\boldsymbol{\beta}$ can be obtained by maximizing the likelihood function. As the error is assumed normality here, the log-likelihood function for the observations $(\boldsymbol{x}_n, y_n)$ $(n = 1, 2, ..., N)$ is

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma) &= -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \{y_n - f(\boldsymbol{x}_n; \boldsymbol{\beta})\}^2 \\ &= -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} SSE(\boldsymbol{\beta}). \end{aligned} \tag{2.3}$$

The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ of $\boldsymbol{\beta}$ and σ^2 respectively which maximize (2.3) are the maximum likelihood estimates. Incidentally, under the fixed σ^2, the $\boldsymbol{\beta}$ maximizing (2.3) consists with the least square estimate $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$. Then, we have the maximum likelihood estimate of σ^2

$$\hat{\sigma}^2 = \frac{1}{N} SSE(\hat{\boldsymbol{\beta}}) = \frac{1}{N} \sum_{n=1}^{N} \{y_n - f(\boldsymbol{x}_n; \hat{\boldsymbol{\beta}})\}^2. \tag{2.4}$$

In fact, $\hat{\boldsymbol{\beta}}$ can be obtained by calculating iterated based on Taylor expansion approximation. Under $\boldsymbol{\beta}_0$ is given as initial value (vector) of $\boldsymbol{\beta}$, we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial \boldsymbol{\beta}} L(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}} \\ &\approx \frac{\partial}{\partial \boldsymbol{\beta}} L(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} + \left\{ \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\mathrm{T}}} L(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} \right\} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \end{aligned}$$

by using Taylor expansion in the first order. So we can approximate as

$$\hat{\boldsymbol{\beta}} \approx \boldsymbol{\beta}_0 - \left\{ \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\mathrm{T}}} L(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} \right\}^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} L(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0}. \tag{2.5}$$

Newton-Raphson method iterates till we finish converging the parameter estimates by updating with (2.5). In addition, it is easier and more stable as converging to calculate second term of (2.5)

$$\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} L(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

directly than to calculate an expectation Fisher information

$$\mathrm{E}\Big\{\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} L(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\Big\} = \Big\{\frac{\partial}{\partial\boldsymbol{\beta}} L(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\Big\}^2.$$

It is well known as Fisher's scoring algorithm.

2.1.2 Power Transform-Both-sides approach

A power-transformation approach aims symmetry (or normality) of the error, homoscedasticity of the error, additivity of the model and obtaining independent observations. In (2.1), one way to give symmetry or homoscedasticity for the error is to power-transform the response. The power transformation with parameter λ for variable t ($t > 0$) is

$$H_{\mathrm{P}}(t;\lambda) \;=\; \begin{cases} (t^{\lambda}-1)/\lambda & \lambda \neq 0\ , \\ \log t & \lambda = 0\ , \end{cases}$$

and it is usually restricted to the response (Box and Cox, 1964). However, in transforming only the response, there is a question about the implications of breaking the known relationship between the response Y and the prediction function $f(\boldsymbol{X};\boldsymbol{\beta})$. The natural setting for this problem is to give identical power transformation for the response Y and the prediction function $f(\boldsymbol{X};\boldsymbol{\beta})$, namely, to use PTB approach. Therefore, for the model (2.1), we have

$$H_{\mathrm{P}}(Y;\lambda) = H_{\mathrm{P}}\{f(\boldsymbol{X};\boldsymbol{\beta});\lambda\} + \varepsilon_{\mathrm{P}} \tag{2.6}$$

(Carroll and Ruppert, 1984). This handling aims to make the error variance constant and normality. However, it is difficult to achieve normality and homoscedasticity of the error after the transformation (Goto *et al.*, 1987; Goto, 1992, 1995, 2000; Jimura and Goto, 1997).Goto (1992) provides three types of double power-transformation approaches and clarifies the assumptions and objectives of the transformations (see also Goto (1995) and Goto *et al.* (2000)). PTB aims to make the error variance constant, but leaves the error distribution unchanged. We assume that the error ε and ε_{P} are distributed as $\mathrm{N}(0, \sigma_n^2)$ and $\mathrm{N}(0, \sigma^2)$ respectively. In the framework of transform-both-sides, we can estimate $\boldsymbol{\beta}, \sigma^2$ and λ, by maximizing the log-likelihood

$$\begin{aligned} L_{\mathrm{P}}(\boldsymbol{\beta}, \sigma^2, \lambda) &= \sum_{n=1}^{N} \Big(-\frac{1}{2}[H_{\mathrm{P}}(y_n; \lambda) - H_{\mathrm{P}}\{f(\boldsymbol{x}_n; \boldsymbol{\beta}), \lambda\}]^2/\sigma^2 + \log \frac{d}{dt} H_{\mathrm{P}}(y_n; \lambda) \\ &\quad -\frac{1}{2} \log \sigma^2 \Big) + C_0 \end{aligned} \tag{2.7}$$

for the observations $\{(\boldsymbol{x}_n, y_n),\ n = 1, 2, ..., N\}$ (Carroll and Ruppert, 1984, 1988), where C_0 is a constant including the coefficient of the probability density function.

2.1.3 Power Weighted Transformation approach

PWT is presented here for obtaining approximate weights in a weighted least squares analysis when the variance of the fitted dependent variable is a function of its expected value. The method is applicable both for linear and non-linear least squares analysis, and whether or not inhomogeneity of variance exists initially or is induced by transformation of the data (Box and Hill, 1974). In the model (2.1), we assume that the error ε are distributed as $\mathrm{N}(0, \sigma_n^2)$. The power weighted transformation function can be expressed as $H_{\mathrm{P}}(y_n; \phi)$ with power weighted parameter ϕ as well as PTB. The variance of $H_{\mathrm{P}}(y_n; \phi)$ is expressed as

$$V(H_{\mathrm{P}}(y_n; \phi)) = \sigma^2. \tag{2.8}$$

An approximate variance expression is now developed for Y using Bartlett's method for stabilizing variance. That is,

$$\begin{aligned} V(y_n) &\approx V[H_{\mathrm{p}}(y_n; \phi)][dy_n/dH_{\mathrm{p}}(y_n; \phi)|_{y_n = \mathrm{E}(y_n)}]^2 \\ &= V(y_n)[\mathrm{E}(y_n)]^{2-2\phi} \\ &= \sigma^2[f(\boldsymbol{x}_n; \boldsymbol{\beta})]^{2-2\phi}. \end{aligned} \tag{2.9}$$

For the weight $\omega_n = [f(\boldsymbol{x}_n; \boldsymbol{\beta})]^{2-2\phi}$, the variance can be expressed

$$V(\sqrt{\omega_n} y_n) \approx \sigma^2.$$

The unknown weighting parameter ϕ will be estimated here by maximizing a likelihood estimate. The log-likelihood for observations $\{(\boldsymbol{x}_n, y_n),\ \ n = 1, 2, ..., N\}$ is

$$L_{\mathrm{W}}(\boldsymbol{\beta}, \sigma^2, \phi) = \sum_{n=1}^{N}\Big(-\frac{\omega_n}{2}[y_n - f(\boldsymbol{x}_n; \boldsymbol{\beta})]^2/\sigma^2 - \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log\omega_n\Big) + C_0. \qquad (2.10)$$

The maximum likelihood estimates $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$ and $\hat{\phi}$ of $\boldsymbol{\beta}, \sigma^2$ and ϕ can be obtained by maximizing the log-likelihood (2.10).

2.2 Inference on theoretical models based on the nonparametric transformation

2.2.1 Transformation based on the smoothing spline function

NTB intends to adjustment for "roughness" of the nonparametric transformation function and estimates the transform-both-sides function and then the parameters of the model. That is, NTB substitutes a nonparametric transformation function for the power transformation function in PTB. We define a penalized likelihood similar to (2.7) in Section 2.1.2, which includes a penalty term. The nonparametric transformation function $H_{\mathrm{S}}(u)$, the parameters $\boldsymbol{\beta}$ and the variance parameter σ^2 are estimated by maximizing the penalized likelihood, where $H_{\mathrm{S}}(u)$ is a smooth function satisfying narrowly-defined monotonicity, and it corresponds to the power transformation function $H_{\mathrm{P}}(t; \lambda)$ of (2.6) in Section 2.1.2. The penalized likelihood function can be written as

$$\begin{aligned} L_{\mathrm{N}}(\boldsymbol{\beta}, \sigma^2, H_{\mathrm{S}}(u)) = -\frac{1}{2}\Big[H_{\mathrm{s}}(y_n) - H_{\mathrm{s}}(f(\boldsymbol{x}_n; \boldsymbol{\beta}))\Big]^2/\sigma^2 + \log\frac{d}{dt}H_{\mathrm{s}}(y_n) \\ -\frac{1}{2}\log\sigma^2 - \frac{\rho}{2}\sum J(H_{\mathrm{S}}(u)),\ \ \rho > 0, \end{aligned} \qquad (2.11)$$

where $L(\boldsymbol{\beta}, \sigma^2, H_{\mathrm{S}}(u))$ is the log-likelihood in the case of replacing $H_{\mathrm{P}}(t; \lambda)$ of (2.6) by $H_{\mathrm{S}}(u)$, and $J(H_{\mathrm{S}}(u))$ is the roughness penalty defined by

$$J(H_{\mathrm{S}}) = \int_{u_{\mathrm{L}}}^{u_{\mathrm{U}}}\Big(\frac{d^2 h_{\mathrm{S}}(u)}{du^2}\Big)^2 du,$$

where u_{L} and u_{U} are chosen as $\{y_n\} \in [u_{\mathrm{L}}, u_{\mathrm{U}}]$, ρ is a constant to adjust the effect of the roughness penalty on the penalized log-likelihood, called "smoothing parameter", $h_{\mathrm{S}}(u)$ is a log derivative of $H_{\mathrm{S}}(u)$, namely $h_{\mathrm{S}}(u) = \log(dH_{\mathrm{S}}(u)/du)$, and

$$H_{\mathrm{S}}(u) = \int \exp[h_{\mathrm{S}}(u)]du,$$

so $H_{\mathrm{S}}(u)$ is strongly limited by the restriction of narrowly-defined monotonicity. We estimate $H_{\mathrm{S}}(u)$ and the parameters $\boldsymbol{\beta}$ by maximizing $L_{\mathrm{N}}(\boldsymbol{\beta}, \sigma^2, H_{\mathrm{S}}(u))$. See Nychka and Ruppert (1995), Ito and Goto (2004) for more details on the estimation method of NTB.

2.2.2 Transformation based on the kernel function

Another method in the NTB approach is to use a kernel density estimator. The kernel density estimation approach is a method of using the frequency of the observations of the neighborhood of the respect to estimate the probability density function in any points. Here, the kernel density estimation method is applied by using the transform function to be a probability density function. The transform-both-sides model is

$$H_{\mathrm{K}}(Y) = H_{\mathrm{K}}[f(\boldsymbol{X}, \boldsymbol{\beta})] + \sigma^2 \varepsilon_{\mathrm{K}} \tag{2.12}$$

, where $H_{\mathrm{K}}(t)$ is a smooth function, σ^2 is a error variance and ε_{K} is distributed a standard normal. We assume the following conditions: 1) $H_{\mathrm{K}}(t)$ is strictly increasing and 2) $H_{\mathrm{K}}(\mathrm{E}(Y)) = 0$. Condition 1) is needed because we want $H_{\mathrm{K}}(t)$ to be invertible, so the correspondence between Y and $f(\boldsymbol{X}, \boldsymbol{\beta})$ can be identified. Conditions 2) is needed to ensure the uniqueness of the solution. It states that $H_{\mathrm{K}}(t)$ passes through a fixed point, $(\mathrm{E}(Y), 0)$. We assume that g_{uy}, g_u and $g_{\varepsilon_{\mathrm{K}}}$ are the probability density functions of $(U, Y), U$ and ε_{K} respectively, where $U = f(\boldsymbol{X}, \boldsymbol{\beta})$ and $g_{\varepsilon_{\mathrm{K}}}$ is a standard normal density function. Then, we can have

$$\begin{aligned} g_{uy} &= g_u(u) g_{uy}(u, y|u) \\ &= g_u(u) g_{\varepsilon}(\varepsilon). \end{aligned}$$

We can also define $\varepsilon_{\mathrm{K}} = H_{\mathrm{K}}(\varepsilon)$, then

$$g_{\varepsilon_{\mathrm{K}}}(\sigma^2 \varepsilon) = g_{\varepsilon_{\mathrm{K}}}[\{H_{\mathrm{K}}(y) - H_{\mathrm{K}}(u)\}/\sigma^2] \frac{dH_{\mathrm{K}}(y)}{dy} / \sigma^2$$

by using the transformation for the variables. Hence we have

$$\frac{dH_{\mathrm{K}}(y)}{dy} / \sigma^2 = \frac{g_{uy}(u, y)}{g_u(u) g_{\varepsilon_{\mathrm{K}}}\{[H_{\mathrm{K}}(y) - H_{\mathrm{K}}(u)]/\sigma^2\}}. \tag{2.13}$$

Where $u = y$ because $g_{\varepsilon_{\mathrm{K}}}(0) = (2\pi)^{-1/2}$, then

$$\frac{dH_{\mathrm{K}}(y)}{dy} / \sigma^2 = (2\pi)^{-1/2} \{g_{uy}(y, y)/g_u(y)\}. \tag{2.14}$$

So if we set $\tilde{H}_{\mathrm{K}}(y) = \dfrac{dH_{\mathrm{K}}(y)}{dy}$, we have

$$\begin{aligned}\int_{y_0}^{y} \tilde{H}_{\mathrm{K}}(t)dt &= \int_{y_0}^{y} g_{uy}(t,t)/\{(2\pi)^{-\frac{1}{2}} g_u(t)\}dt \\ &= B_1 H_{\mathrm{K}}(y) + B_0 \end{aligned} \tag{2.15}$$

for the constant y_0, where

$$B_1 = 1/\sigma_0, \quad B_0 = \int_{y_0}^{\mathrm{E}(Y)} g_{uy}(t,t)/\{(2\pi)^{-\frac{1}{2}} g_u(t)\}dt.$$

This suggests estimating $H_{\mathrm{K}}(y)$ in the following way. The first step is to obtain a preliminary $n^{1/2}-$consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0^*$. We then just replace $U_n = f(\boldsymbol{X}_n, \boldsymbol{\beta})$ by $V_n = f(\boldsymbol{X}_n, \hat{\boldsymbol{\beta}})$ by setting $\boldsymbol{\beta} = \boldsymbol{\beta}_0^*$ and giving the observations $(\boldsymbol{x}_n, y_n)$, $\quad n = 1, 2, ..., N$. Wang and Ruppert (1995) use the LAD estimator (denoted by $\boldsymbol{\beta}_{\mathrm{LAD}}$) as the preliminary estimator, $\hat{\boldsymbol{\beta}}$. Therefore, $\boldsymbol{\beta}_{\mathrm{LAD}}$ is $n^{1/2}$ consistent by a M-estimator argument. Note that because of the structure of model (2.12), the consistency of the least squares estimator will depend on the form of the unknown g. In model (2.12), $f(\boldsymbol{X}_n, \boldsymbol{\beta})$ is the conditional median of Y_n given $\boldsymbol{X}_n$ but is not the conditional mean (except for special g). Therefore the least squares estimator is not in general consistent, so we prefer using $\boldsymbol{\beta}_{\mathrm{LAD}}$ to using the least squares estimator.

2.3 Inference on theoretical models based on the semi-parametric transformation

2.3.1 Power Transform-Both-sides and Weighted Least Squares approach

In PTBWLS, we implement the power weighted transformation parameter for PTB. In (2.6), we assume that the distribution of the response is non-normal and that by transforming both sides, the response is distributed normally with inconstant variance $\sigma_n^2, n = 1, 2, ..., N$. We attempt to attain homoscedasticity of the response after transforming by implementing the power weighted transformation parameter ϕ. For the power transformation $H_{\mathrm{P}}(t; \phi)$, using Bartlett's methods (Bartlett, 1947), we have the first order approximation of the variance of $H_{\mathrm{P}}(y_n;\lambda)$

$$\begin{aligned} V(H_{\mathrm{P}}(y_n;\lambda)) &\approx V[H_{\mathrm{P}}(H_{\mathrm{P}}(y_n;\lambda);\phi)]\Big[\frac{dH_{\mathrm{P}}(y_n;\lambda)}{dH_{\mathrm{P}}(H_{\mathrm{P}}(y_n;\lambda);\phi)}\Big|_{H_{\mathrm{P}}(y_n;\lambda)=\mathrm{E}[H_{\mathrm{P}}(y_n;\lambda)]}\Big]^2 \\ &= V[H_{\mathrm{P}}(y_n;\lambda)](\mathrm{E}[H_{\mathrm{P}}(y_n;\lambda)])^{2-2\phi} \\ &= V[H_{\mathrm{P}}(y_n;\lambda)][H_{\mathrm{P}}(f(\boldsymbol{x}_n;\boldsymbol{\beta});\lambda)]^{2-2\phi}. \end{aligned}$$

For the weight $\omega_n = [H_{\mathrm{P}}(f(\boldsymbol{x}_n;\boldsymbol{\beta});\lambda)]^{2\phi-2}$, the variance can be expressed

$$V[\sqrt{\omega_n}H_{\mathrm{P}}(y_n;\lambda)] \approx \sigma^2.$$

The weighted parameter ϕ is chosen to make the variance constant using this relationship (Box and Hill, 1974). Therefore, the log-likelihood for observations $\{(\boldsymbol{x}_n, y_n), n = 1, 2, ..., N\}$ is

$$\begin{aligned} L_{\mathrm{PW}}(\boldsymbol{\beta},\sigma^2,\lambda,\phi) = \sum_{n=1}^{N}\Big(&-\frac{\omega_n}{2}\{H_{\mathrm{P}}(y_n;\lambda) - H_{\mathrm{P}}[f(\boldsymbol{x}_n;\boldsymbol{\beta}),\lambda]\}^2/\sigma^2 \\ &+\log\frac{d}{dt}H_{\mathrm{P}}(y_n;\lambda) - \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log\omega_n\Big) + C_0. \end{aligned} \tag{2.16}$$

The maximum likelihood estimates $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}$ and $\hat{\phi}$ of $\boldsymbol{\beta}, \sigma^2, \lambda$ and ϕ can be obtained by maximizing the log-likelihood (2.16). In (2.7), if we have S_{PTB} as the term for a sum of squares in the log-likelihood, it can be written as

$$S_{\mathrm{PTB}} = \sum_{n=1}^{N}[(y_n^{\lambda} - f_n^{\lambda})/\lambda]^2,$$

where $f_n = f(\boldsymbol{x}_n;\boldsymbol{\beta})$.Then, we have

$$S_{\mathrm{PTB}} \approx \sum_{n=1}^{N}\{f_n^{\lambda-1}(y_n - f_n) + 1/(2\lambda)[f_n^{\lambda-1} + \lambda(\lambda-1)f_n^{\lambda-2}](y_n - f_n)^2\}^2$$

by Taylor expansion for y_n^{λ} around f_n in the second order. If we ignore the fourth order term about $(y_n - f_n)$, we have

$$S_{\mathrm{PTB}} \approx \sum_{n=1}^{N} f_n^{2\lambda-2}(y_n - f_n)^2 + \sum_{n=1}^{N} f_n^{2\lambda-3}(f_n\lambda^{-1} + \lambda - 1)(y_n - f_n)^3.$$

In this expression the first term $\sum_{n=1}^{N}\omega_n(y_n - f_n)^2$, corresponds to the sum of squares in the power weighted transformation approach of Box and Hill (1974). Then, the second term $(y_n - f_n)^3$ stands for the third moment corresponding to the skewness of the error distribution. So, we can examine how minimizing the sum of squares in PTB can correct for not only the heteroscedasticity of the error, but the skewness of the error distribution.

2.3.2 Nonparametric Transform-Both-sides and Weighted Least Squares approach: brief overview

NTBWLS intends to estimate the transform-both-sides function, the parameters of the model and the power weighted transformation parameter simultaneously based on a penalized likelihood. That is, NTBWLS substitutes nonparametric transformation function for the power transformation function in PTBWLS. We define a penalized likelihood similar to (2.16) in Section 2.3.1 with a penalty term. $H_{\mathrm{S}}(u)$, $\boldsymbol{\beta}$, ϕ and σ^2 are estimated by maximizing the penalized likelihood. It can be written as

$$L_{\mathrm{NW}}(\boldsymbol{\beta}, \phi, \sigma^2, H_{\mathrm{S}}(u)) = L(\boldsymbol{\beta}, \phi, \sigma^2, H_{\mathrm{S}}(u)) - \frac{\rho}{2} J(H_{\mathrm{S}}(u)), \ \rho > 0, \tag{2.17}$$

where $L_{\mathrm{NW}}(\boldsymbol{\beta}, \phi, \sigma^2, H_{\mathrm{S}}(u))$ is the log-likelihood when we change $H_{\mathrm{P}}(t;\lambda)$ in (2.16) to $H_{\mathrm{S}}(u)$. The maximum penalized likelihood estimates $\hat{H}_{\mathrm{S}}(u)$, $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\hat{\phi}$ of $H_{\mathrm{S}}(u)$, $\boldsymbol{\beta}$, σ^2 and ϕ can be obtained by maximizing (2.17). Practically, we obtain the estimates by iterating over the next three steps:

(step 1) Under the fixed $\boldsymbol{\beta}$ and ϕ, estimate $H_{\mathrm{S}}(u)$ maximizing $L_{\mathrm{P}}(\boldsymbol{\beta}, \sigma^2, H_{\mathrm{S}}(u))$.

(step 2) Under transformation of both sides by $\hat{H}_{\mathrm{S}}(u)$, estimate ϕ maximizing $L_{\mathrm{P}}(\boldsymbol{\beta}, \phi, \sigma^2, \hat{H}_{\mathrm{S}}(u))$.

(step 3) Under transformation of both sides by $\hat{H}_{\mathrm{S}}(u)$ and estimated $\hat{\phi}$, estimate $\boldsymbol{\beta}$ maximizing $L_{\mathrm{P}}(\boldsymbol{\beta}, \hat{\phi}, \sigma^2, \hat{H}_{\mathrm{S}}(u))$.

This algorithm is performed under fixed smoothing parameter ρ. In this paper, we set some smoothing parameter and examine the relation to homogeneity and normality of the error variance after transforming. $H_{\mathrm{S}}(u)$ is estimated by using a cubic smoothing spline. We change (2.16) to a penalized likelihood relating to $h_{\mathrm{S}}(u)$ and estimate the parameters by using non-restrictive optimization. As the distribution of the response before and after transformation is assumed to be no different, unlike PTBWLS, the penalized likelihood is

$$\begin{aligned} L_{\mathrm{NW}}(\boldsymbol{\beta}, \phi, \sigma^2, h_{\mathrm{S}}(u)) &= \frac{1}{2}\sum_{n=1}^{N}\Big\{-\Big(\omega_n \int_{y_n}^{f_n} \exp h_{\mathrm{S}}(u)du\Big)^2 \Big/ \sigma^2 + 2h_{\mathrm{S}}(y_n) - \log\sigma^2 \\ &\quad + \log\omega_n\Big\} - \frac{\rho}{2}\int_{u_{\mathrm{L}}}^{u_{\mathrm{U}}} \Big(\frac{d^2 h_{\mathrm{S}}(u)}{du^2}\Big)^2 du + C_0, \end{aligned} \tag{2.18}$$

where y_n $(n = 1, 2, ..., N)$ are observations of response, $f_n = f(\boldsymbol{x}_n; \boldsymbol{\beta})$, ω_n are the weights determined by ϕ: $\omega_n = H_{\mathrm{S}}[f(\boldsymbol{x}_n; \boldsymbol{\beta})]^{2\phi-2}$.

2.3.3 Estimation of the nonparametric transformation function

In order to optimize (2.18), the estimation of $h_{\mathrm{S}}(u)$ is necessary. We estimate $h_{\mathrm{S}}(u)$ by using a weighted cubic smoothing spline. Then, under $J \geq 3$, for any j ($j = 1, 2, ..., J$), given (u_j, Z_j), we put

$$Z_j = h_{\mathrm{S}}(u_j) + \nu_j,$$

where $u_1, ..., u_J$ are the points in $[u_{\mathrm{L}}, u_{\mathrm{U}}]$, which satisfy $u_{\mathrm{L}} < u_1 < \cdots < u_J < u_{\mathrm{U}}$,and are chosen to become $\{y_n\} \in [u_1, u_J]$. In addition, the error ν_j is distributed $\mathrm{N}(0, 1/\eta_j)$ with variance $1/\eta_j$. Next, we define the two functional spaces: $\mathscr{S}_1[u_{\mathrm{L}}, u_{\mathrm{U}}]$ is the function in $[u_{\mathrm{L}}, u_{\mathrm{U}}]$. It is all of the functional space that is differentiable and absolutely continuous. $\mathscr{S}_2[u_{\mathrm{L}}, u_{\mathrm{U}}]$ is the function in $[u_{\mathrm{L}}, u_{\mathrm{U}}]$. It is all of the functional space that has a continuous second derivative. In this case, we define a penalized sum of squares

$$\mathscr{S}(h_{\mathrm{S}}(u)) = \sum_{j=1}^{J} \{Z_j - h_{\mathrm{S}}(u_j)\}^2 \eta_j + \rho \int_{u_{\mathrm{L}}}^{u_{\mathrm{U}}} \Big(\frac{d^2 h_{\mathrm{S}}(u)}{du^2}\Big)^2 du, \ \ \rho > 0, \tag{2.19}$$

where η_j is the weight. In addition, it is assumed that the estimation equation produces on $h_{\mathrm{S}}(u)$ that minimizes $\mathscr{S}(h_{\mathrm{S}}(u))$ in a set of all curves smooth enough, and $\mathscr{S}_2[u_{\mathrm{L}}, u_{\mathrm{U}}]$ is $\hat{h}_{\mathrm{S}}(u)$. In this case, $\hat{h}_{\mathrm{S}}(u)$ is the (natural) cubic spline with knots at u_j (O'sullivan *et al.*,1986).

We extend the maximizing penalized log-likelihood algorithm of Nychka and Ruppert(1995), and build the algorithm with the power weighted transformation parameter and estimate $h_{\mathrm{S}}(u)$. In practice, on the basis of choosing $u_1, ..., u_J$ as including $\{y_n\}$ and $\{f_n\}$, we approximate $\{Z_j\}$ by the integral in the first term of (2.18), namely

$$\omega_n \int_{y_n}^{f_n} \exp h_{\mathrm{S}}(u) du,$$

and we maximize the penalized log-likelihood. More specifically, we approximate (2.18) by

$$\begin{aligned} L_{\mathrm{NW}}(\sigma^2, h_{\mathrm{S}}(u)) = \sum_{n=1}^{N} \Big[-\Big\{ \sum_{j=1}^{J} W_{nj} \exp h_{\mathrm{S}}(u_j) \Big\}^2 \Big/ 2\sigma^2 + \sum_{j=1}^{J} \zeta_{nj} h_{\mathrm{S}}(u_j) \Big] \\ - \frac{\rho}{2} \int_{u_{\mathrm{L}}}^{u_{\mathrm{U}}} \Big(\frac{d^2 h_{\mathrm{S}}(u)}{du^2}\Big)^2 du, \end{aligned}$$

where, W_{nj} is the $n \times j$ component of the matrix, and

$$\left|\omega_n \int_{y_n}^{f_n} \exp h_{\mathrm{S}}(u) du\right| \approx \sum_{j=1}^{J} W_{nj} \exp(h_{\mathrm{S}}(u_j)).$$

In addition, ζ_{nj} is chosen so as to

$$h_{\mathrm{S}}(y_n) \approx \sum_{j=1}^{J} \zeta_{nj} h_{\mathrm{S}}(u_j).$$

If we set

$$\begin{aligned} \boldsymbol{h}_{\mathrm{S}} &= (h_{\mathrm{S}}(u_1), ..., h_{\mathrm{S}}(u_J))^{\mathrm{T}}, \\ \frac{d^2 \boldsymbol{h}_s(u)}{du^2} &= (d^2 h_{\mathrm{S}}(u_1)/du^2, ..., d^2 h_{\mathrm{S}}(u_J)/du^2)^{\mathrm{T}}, \end{aligned}$$

the natural cubic spline $h_{\mathrm{S}}(u)$ with knots $u_1, ..., u_J$ can be determined uniquely as $\boldsymbol{h}_{\mathrm{S}}$. So we rewrite (2.18) as

$$L_{\mathrm{NW}}(\boldsymbol{h}_{\mathrm{S}}) = -\frac{1}{2} \boldsymbol{h}_{\mathrm{S}}^{*\mathrm{T}} O \boldsymbol{h}_{\mathrm{S}}^{*} + \boldsymbol{\zeta}^{\mathrm{T}} \boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}} - \frac{\rho}{2} \boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{h}_{\mathrm{S}}, \tag{2.20}$$

where

$$\boldsymbol{h}_{\mathrm{S}}^{*} = (\exp[h_{\mathrm{S}}(u_1)], ..., \exp[h_{\mathrm{S}}(u_J)])^{\mathrm{T}}, \ \boldsymbol{O} = \boldsymbol{W}^{\mathrm{T}} \mathrm{diag}(\boldsymbol{V}) \boldsymbol{W},$$

and $\boldsymbol{W}$ is the matrix with components W_{nj}, $\boldsymbol{V}$ is the $N \times N$ matrix with the diagonal components $V_{nn} = f(\boldsymbol{x}_n; \boldsymbol{\beta})^{2-2\phi}$ with $n = 1, 2, ..., N$ and furthermore other components are 0.

$\boldsymbol{\zeta} = (\zeta_{\mathrm{i}}, ..., \zeta_j)^{\mathrm{T}}$. Then

$$\zeta_j = \sum_{n=1}^{N} \zeta_{nj},$$

and $\boldsymbol{K}$ is the symmetric $J \times J$ matrix obtained by composing $\boldsymbol{h}_{\mathrm{S}}$ and $\dfrac{d^2 \boldsymbol{h}_s(u)}{du^2}$. We differentiate (2.20) partially by $\{h_{\mathrm{S}}(u_j)\}$. Consequently, we obtain

$$\frac{\partial L_{\mathrm{PA}}(h_{\mathrm{S}}(u))}{\partial h_{\mathrm{S}}(u)}\Big|_{u=u_j} = -h_{\mathrm{S}}(u_j)[\boldsymbol{O}\boldsymbol{h}_{\mathrm{S}}^{*}]_j + \zeta_j - \rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0, \ j = 1, 2, ..., J. \tag{2.21}$$

Further, $\{h_{\mathrm{S}}(u_j)\}$ can be obtained as satisfying (2.21). Thus, we can determine the estimation equation of $\hat{h}_{\mathrm{S}}(u)$ uniquely. Finally, for the fixed parameter $\boldsymbol{\beta}$, ϕ, σ^2, the estimation algorithm of $h_{\mathrm{S}}(u)$ is as follows:

(step 1) determine the knots $\{u_j\}$ $(j = 1, 2, ..., J)$.

(step 2) based on the fixed parameter $\boldsymbol{\beta}$ and ϕ, compute $\boldsymbol{V}$ and $\boldsymbol{W}$.

(step 3) compute $\boldsymbol{O}$ and $\boldsymbol{\zeta}$.

(step 4) set $h_{\mathrm{S}0} = 0$.

(step 5) based on $h_{\mathrm{S}0}$,$\boldsymbol{O}$,$\boldsymbol{\zeta}$, compute $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_J)^{\mathrm{T}}$ and the weight $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_J)^{\mathrm{T}}$.

(step 6) based on $h_{\mathrm{S}0}$ and the weight $\{\eta_j\}$, $\{u_j, Z_j\}$, estimate the cubic spline $h_{\mathrm{S}1}$.

(step 7) set $h_{\mathrm{S}0} = \hat{h}_{\mathrm{S}1}$.

Step 1, step 2, step 3 and step 4 are initialization. For these values, we iterate step 5, step 6 and step 7 till we finish converging as $h_{\mathrm{S}0} = \hat{h}_{\mathrm{S}1}$. As well, in step 6, we have the weight $\{\eta_j\}$, the pairs of knots and working response $\{u_j, Z_j\}$ by initial value $h_{\mathrm{S}0}$. These values depend on O_{jj} and D_j and the process of calculation is as follows:

(pattern 1) $O_{jj} = 0$: If $O_{jj} = 0$, set $Z_j = D_j + h_{\mathrm{S}0}(u_j)$ and $\eta_j = 1$.

(pattern 2) $O_{jj} > 0$ and $D_j = 0$: (2.20) can be written by

$$\exp(2h_{\mathrm{S}}(u_j))O_{jj} + \exp h_{\mathrm{S}}(u_j) \sum_{j \neq j_0}^{J} O_{jj_0} \exp h_{\mathrm{S}0}(u_{j_0}) + \rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0,$$

where $j = 1, 2, ..., J$. First, at the second term, we conduct the diagonalization by updating $h_{\mathrm{S}0}(u_{j_0})$ as $S_0 \to S$. Here, the Taylor-expansion can be used about $h_{\mathrm{S}0}(u_j)$ in this linearization. Namely, for the first term in above equation, it can be written by

$$\exp(2h_{\mathrm{S}}(u_j)) \approx \exp(2h_{\mathrm{S}0}(u_j))\{1 + 2(h_{\mathrm{S}}(u_j) - h_{\mathrm{S}0}(u_j))\}.$$

For the second term, we alternate $\exp h_{\mathrm{S}}(u_j)$ by $\exp h_{\mathrm{S}0}(u_j)$. Hence we get the approximate expression

$$\exp(2h_{\mathrm{S}0}(u_j))\{1 + 2(h_{\mathrm{S}}(u_j) - h_{\mathrm{S}0}(u_j))\}O_{jj} + \exp h_{\mathrm{S}0}(u_j) \sum_{j \neq j_0}^{J} O_{jj_0} \exp h_{\mathrm{S}0}(u_j)$$
$$+\rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0.$$

If we set $h_{\mathrm{S}}^*(u_j) \equiv \exp h_{\mathrm{S}}(u_j)$, we have

$$-2h_{\mathrm{S}0}^*(u_j)^2 O_{jj}\Big(-\frac{1}{2h_{\mathrm{S}0}^*(u_j)O_{jj}}[\boldsymbol{O}\boldsymbol{h}_{\mathrm{S}0}(u_j)]_j + h_{\mathrm{S}0}^*(u_j) - h_{\mathrm{S}}^*(u_j)\Big) + \rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0.$$

Namely, we have

$$-2\eta_j(Z_j - h_{\mathrm{S}}(u_j)) + \rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0.$$

(pattern 3) $O_{jj} > 0$ and $D_j > 0$: (2.20) can be written by

$$\exp(2h_{\mathrm{S}}(u_j))\Big\{O_{jj} + \exp(-h_{\mathrm{S}}(u_j))\sum_{j\neq j_0}^{N} O_{jj_0}\exp h_{\mathrm{S}}(u_j) - \exp(-2h_{\mathrm{S}}(u_j))D_j\Big\}$$
$$+\rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0.$$

In a similar way to (pattern 2), we conduct the linearization

$$-2D_j\Big(\frac{-h_{\mathrm{S}0}(u_j)[\boldsymbol{O}\boldsymbol{h}_{\mathrm{S}0}]_j}{2D_j + 1/2 + h_{\mathrm{S}0}(u_j)} - h_{\mathrm{S}}(u_j)\Big) - \rho[\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]_j = 0.$$

As the above formula, we set $\eta_j = D_j$ and the first term in the parenthesis on Z_j.

As well, $H_{\mathrm{S}}(u)$ can be expressed by indefinite integral $\exp[h_{\mathrm{S}}(u)]$, so in this paper, we calculate this integral by trapezoid approximation.

2.3.4 Algorithms for identification of the weighted cubic smoothing spline function

In the above section, we suggested the estimation algorithm of $h_{\mathrm{S}}(u)$. This optimization need to identify the weighted cubic smoothing spline function. Here we provide this algorithm in detail suggested by Green and Silverman (1994). Define

$$h_{\mathrm{S}j} = h_{\mathrm{S}}(u_j), \quad \gamma_j = \frac{d^2 h_{\mathrm{S}}(u)}{du^2}\Big|_{u=u_j}, \quad j = 1, ..., J.$$

By the definition of a NCS (Natural Cubic Spline), $\gamma_1 = \gamma_N = 0$. Also we set

$$\boldsymbol{h}_{\mathrm{S}} = (h_{\mathrm{S}1}, ..., h_{\mathrm{S}J})^{\mathrm{T}},$$
$$\boldsymbol{\gamma} = (\gamma_2, ..., \gamma_{J-1})^{\mathrm{T}}.$$

The vectors $\boldsymbol{h}_{\mathrm{S}}$ and $\boldsymbol{\gamma}$ specify the curve h_{S} completely, and it is possible to give explicit formulae in terms of $\boldsymbol{h}_{\mathrm{S}}$ and $\boldsymbol{\gamma}$ for the value and derivatives of $\boldsymbol{h}_{\mathrm{S}}$ at any point.

The condition depends on two band matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ which we now define. Let $v_j = u_{j+1} - u_j$ for $j = 1, ..., J-1$. Let $\boldsymbol{Q}$ be the $J \times (J-2)$ matrix with entries q_{lj}, for $j = 1, ..., J-1$ and $l = 2, ..., J-1$, given by

$$q_{l-1,l} = v_{l-1}^{-1}, \quad q_{ll} = -v_{l-1}^{-1} - v_l^{-1}, \quad q_{l+1,l} = v_l^{-1}$$

for $l = 2, ..., j-1$, and $q_{jl} = 0$ for $|j-l| \geq 2$. The columns of $\boldsymbol{Q}$ are numbered in the same non-standard way as the entries of $\boldsymbol{\gamma}$, starting at $l = 2$, so that the top left element of $\boldsymbol{Q}$ is q_{12}. The symmetric matrix $\boldsymbol{R}$ is $(J-2) \times (J-2)$ with elements r_{jl}, for j and l running from 2 to $(j-1)$, given by

$$\begin{aligned} r_{jj} &= \frac{1}{3}(v_{j-1} + v_j), \quad j = 2, ..., J-1 \\ r_{j,j+1} = r_{j+1,j} &= \frac{1}{6}v_j, \quad j = 2, ..., J-2 \\ r_{jl} &= 0, \qquad\quad |j-l| \geq 2. \end{aligned}$$

The matrix $\boldsymbol{R}$ is strictly diagonal dominant. Standard arguments in numerical linear algebra show that $\boldsymbol{R}$ is strictly positive-definite. Also, we define the matrix $\boldsymbol{\Upsilon}$ to be the diagonal matrix with diagonal elements η_j. We can therefore define a matrix $\boldsymbol{K}$ by

$$\boldsymbol{K} = \boldsymbol{\Upsilon}^{-1}\boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}.$$

The vectors $\boldsymbol{h}_{\mathrm{S}}$ and $\boldsymbol{\gamma}$ specify a natural cubic spline h_{S} if and only if the condition

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{h}_{\mathrm{S}} = \boldsymbol{R}\boldsymbol{\gamma} \tag{2.22}$$

is satisfied. If (2.22) is satisfied then the roughness penalty will satisfy

$$\int_a^b \{\frac{d^2 h_{\mathrm{S}}(u)}{du^2}\}^2 du = \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{\gamma} = \boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}. \tag{2.23}$$

$\rho\boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}$ is expressed as the roughness penalty. Also, for the observational vectors $\boldsymbol{Z} = (Z_1, ..., Z_J)^{\mathrm{T}}$, the residual sum of squares about $\boldsymbol{h}_{\mathrm{S}}$ can be written

$$\sum \eta_j\{Z_j - h(u_j)\}^2 = (\boldsymbol{Z} - \boldsymbol{h}_{\mathrm{S}})^{\mathrm{T}}\boldsymbol{\Upsilon}(\boldsymbol{Z} - \boldsymbol{h}_{\mathrm{S}}).$$

Therefore we can rewrite (2.19) as

$$\mathscr{S}(h_{\mathrm{S}}) \;=\; (\boldsymbol{Z}-\boldsymbol{h}_{\mathrm{S}})^{\mathrm{T}}\boldsymbol{\Upsilon}(\boldsymbol{Z}-\boldsymbol{h}_{\mathrm{S}})+\rho\boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}. \tag{2.24}$$

Since $\rho\boldsymbol{K}$ is non-negative definite, the matrix $\boldsymbol{\Upsilon}+\rho\boldsymbol{K}$ is strictly positive definite. It therefore follows that (2.24) has a unique minimum, obtained by setting

$$\boldsymbol{h}_{\mathrm{S}} = (\boldsymbol{\Upsilon}+\rho\boldsymbol{K})^{-1}\boldsymbol{\Upsilon}\boldsymbol{Z}. \tag{2.25}$$

To estimate $\boldsymbol{h}_{\mathrm{S}}$ efficiently, we use the algorithm proposed by Reinsch (1967). From (2.25), we have

$$\boldsymbol{h}_{\mathrm{S}} = \boldsymbol{Z}-\rho\boldsymbol{\Upsilon}^{-1}\boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{h}_{\mathrm{S}}, \tag{2.26}$$

and hence

$$\boldsymbol{h}_{\mathrm{S}} = \boldsymbol{Z}-\rho\boldsymbol{\Upsilon}\boldsymbol{Q}\boldsymbol{\gamma}. \tag{2.27}$$

As before, substituting $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{h}_{\mathrm{S}} = \boldsymbol{R}\boldsymbol{\gamma}$ we obtain, after some manipulation,

$$(\boldsymbol{R}+\rho\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}\boldsymbol{Q})\boldsymbol{\gamma} = \boldsymbol{Q}\boldsymbol{Y}. \tag{2.28}$$

Because $\boldsymbol{\Upsilon}$ is a strictly positive definite diagonal matrix, the matrix $(\boldsymbol{R}+\rho\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}\boldsymbol{Q})$ is a band matrix with $j=2$ and has a Cholesky decomposition$\boldsymbol{L}\boldsymbol{E}\boldsymbol{L}^{\mathrm{T}}$where, as before, $\boldsymbol{L}$ is a lower diagonal band matrix with unit diagonal and $\boldsymbol{E}$ is a strictly positive diagonal matrix. The resulting algorithm, all of whose steps can be performed in $O(n)$ algebraic operations, can now be set out.

Step 1. For $j=2,...,J-1$, evaluate the vector $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Z}$. Where we can use

$$(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{h}_{\mathrm{S}})_k = \frac{h_{j+1}-h_j}{v_j} - \frac{h_j-h_{j-1}}{v_{j-1}}.$$

Step 2. Find the non-zero diagonals of $\boldsymbol{R}+\rho\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}\boldsymbol{Q}$, and its Cholesky decomposition factors $\boldsymbol{L}$ and $\boldsymbol{E}$.

Step 3. Write (2.28) as $\boldsymbol{L}\boldsymbol{E}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\gamma} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Z}$ and solve this equation for $\boldsymbol{\gamma}$ by forward and back substitution.

Step 4. From (2.26), use

$$\boldsymbol{h}_{\mathrm{S}} = \boldsymbol{Z}-\rho\boldsymbol{Q}\boldsymbol{\gamma}$$

to find $\boldsymbol{h}_{\mathrm{S}}$.

2.3.5 Selection of the smoothing parameter

Most nonparametric estimates of functions have a free parameter that controls the flexibility of the resulting curve estimate. In this case, it is ρ, the relative weight given to the roughness penalty over the log-likelihood. One practical issue is to choose an appropriate value for this parameter when no a priori information is available about the smoothness of the transformation. Because this is a non-linear estimate, previous work on smoothing parameter or bandwidth selection does not apply. One computationally intensive strategy is to estimate ρ on the basis of cross-validation. For a fixed value of ρ each data point is omitted from the likelihood and $\boldsymbol{h}_{\mathrm{S}}, \boldsymbol{\beta}\ and\ \phi$ are estimated on the basis of the remaining $n-1$ data points. The log-likelihood based on these estimates is now evaluated at the omitted data point. On the other hand, we hope that a good estimate of $\boldsymbol{h}_{\mathrm{S}}$ will yield transformed residuals that are independent with constant variance. Thus any statistic used to test for departures from these assumptions can be used to judge the suitability of a particular value for ρ. However, the idea of cross-validation is to optimize on predicting responses, which does not match to a primary objective of the nonparametric regression, that is, to explore nonlinear regression structure. Moreover, a distance between response variables might be difficult to take information on complicated regression structure into account (Sakamoto, 2007). Nychka and Ruppert (1995) examined the heteroscedasticity in the transformed residuals based on the Spearman rank correlation of the absolute residuals with the predicted values. Relatively small values of this correlation may suggest good choices for the smoothing parameter.

For (2.20), it can be also explained in the Bayesian context. The Bayesian justification of penalized maximum likelihood is to place a prior density proportional to

$$\exp\Big[-\frac{\rho}{2}\int_{u_{\mathrm{L}}}^{u_{\mathrm{U}}}\Big(\frac{d^2 h_{\mathrm{S}}(u)}{du^2}\Big)^2 du\Big]$$

over the space of all smooth functions. The larger the value of ρ, the more weight is put on functions with smaller roughness. With this prior, the posterior log density of the function h_{S} is then, in the regression context, equal to L_{PA} as defined in (2.20) above, and so the spline smoother $\hat{h}_{\mathrm{S}}$ is the posterior mode given the data (Green and Silverman, 1994). Suppose that the prior density of $\boldsymbol{h}_{\mathrm{S}}$ is $p(\boldsymbol{h}_{\mathrm{S}};\rho) \propto \exp[-\frac{\rho}{2}\boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{h}_{\mathrm{S}}]$. Let the conditional density of $\boldsymbol{Z}$ for given $\boldsymbol{\beta}$ and $\boldsymbol{h}_{\mathrm{S}}$ be denoted by $p(\boldsymbol{Z}|\boldsymbol{\beta},\boldsymbol{h}_{\mathrm{S}},\phi;\sigma^2) =$

$(2\pi\sigma^2)^{-n/2}\exp[-\frac{\sigma^2}{2}(\boldsymbol{Z}-\boldsymbol{h}_{\mathrm{S}})^{\mathrm{T}}\boldsymbol{\Upsilon}(\boldsymbol{Z}-\boldsymbol{h}_{\mathrm{S}})]$. By using the Bayes theorem, the joint posterior density of $\boldsymbol{\beta}, \phi, \boldsymbol{h}_{\mathrm{S}}$ is proportional to the joint density of $\boldsymbol{Z}$, $\boldsymbol{\beta}, \phi$ and $\boldsymbol{h}_{\mathrm{S}}$:

$$p(\boldsymbol{\beta}, \boldsymbol{h}_{\mathrm{S}}, \phi|\boldsymbol{Z}, \rho; \sigma^2) \propto p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{h}_{\mathrm{S}}; \phi, \sigma^2)p(\boldsymbol{\beta})p(\boldsymbol{h}_{\mathrm{S}}, \rho). \tag{2.29}$$

Hence obtaining the mode of the joint posterior density of $\boldsymbol{\beta}, \boldsymbol{h}_{\mathrm{S}}$ is equivalent to maximizing the penalized log-likelihood.

Another procedure of selecting the smoothing parameters is to maximize the marginal likelihood. The smoothing parameters, which govern global nonlinear regression structure, are estimated with the maximum marginal likelihood estimation, or the empirical Bayes method (Sakamoto, 2007). In the NTBWLS approach, We calculate $\boldsymbol{h}_{\mathrm{S}}$, $\boldsymbol{\beta}$, ϕ and ρ by maximizing the marginal likelihood. Then the marginal density of $\boldsymbol{Z}$ becomes

$$\begin{aligned} p(\boldsymbol{Z}; \phi, \rho, \sigma^2) &= \int \cdots \int_D p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{h}_{\mathrm{S}}; \phi, \sigma^2)p(\boldsymbol{\beta})p(\boldsymbol{h}_{\mathrm{S}}; \rho)d\boldsymbol{\beta}d\boldsymbol{h}_{\mathrm{S}} \\ &\propto \int \cdots \int_D \exp L_{\mathrm{NW}}(\boldsymbol{\beta}, \boldsymbol{h}_{\mathrm{S}}; \boldsymbol{Z})d\boldsymbol{\beta}d\boldsymbol{h}_{\mathrm{S}}, \end{aligned} \tag{2.30}$$

where $D = \boldsymbol{R}^{p+1} \times D_1$. We can consider (2.30) as a function of (ϕ, ρ, σ^2), the marginal likelihood of $\boldsymbol{Z}$, and maximize it with respect to these parameters. Some approaches of computing the marginal density of $\boldsymbol{Z}$ approximately have been discussed. In this paper, we use a Laplace approximation approach which derive approximation forms without using the integral because of its easy computation (Tierney and Kadane, 1986; Davison, 1986; Sakamoto, 2007).

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{h}_{\mathrm{S}}^{\mathrm{T}})^{\mathrm{T}}$ for simplicity of notation. We consider Taylor expansion of the penalized log-likelihood $L_{\mathrm{PA}}(\boldsymbol{\theta}; \boldsymbol{Z})$ around its maximum point, that is, the MPLEs $\hat{\boldsymbol{\theta}}$. Then we obtain the Laplace approximation

$$L_{\mathrm{PA}}(\boldsymbol{\theta}; \boldsymbol{Z}) \approx L_{\mathrm{PA}}(\hat{\boldsymbol{\theta}}; \boldsymbol{Z}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}}H(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{2.31}$$

where $H(\hat{\boldsymbol{\theta}})$ is the negative Hessian of the penalized log-likelihood

$$H(\hat{\boldsymbol{\theta}}) = \Big(-\frac{\partial^2 L_{\mathrm{PA}}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\Big)_{\hat{\boldsymbol{\theta}}},$$

By substituting (2.31) into (2.30), an approximated marginal density of $\boldsymbol{Z}$ becomes

$$\begin{aligned} p(\boldsymbol{Z}; \phi, \rho, \sigma^2) &\approx \exp L_{\mathrm{PA}}(\hat{\boldsymbol{\theta}}; \boldsymbol{Z}) \int \cdots \int_D \exp\Big[\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}}H(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\Big]d\boldsymbol{\theta} \\ &\propto |H(\hat{\boldsymbol{\theta}})|_+^{-1/2} \exp L_{\mathrm{PA}}(\hat{\boldsymbol{\theta}}; \boldsymbol{Z}), \end{aligned}$$

where $|H(\hat{\boldsymbol{\theta}})|_+$ is the product of non-zero eigenvalues of $H(\hat{\boldsymbol{\theta}})$. Hence, we obtain an approximated marginal log-likelihood

$$L_{\mathrm{M}}(\phi, \rho, \sigma^2; \boldsymbol{Z}) = L_{\mathrm{PA}}(\hat{\boldsymbol{\theta}}; \boldsymbol{Z}) - \frac{1}{2}\log|H(\hat{\boldsymbol{\theta}})|_+ + const., \tag{2.32}$$

and we maximize (2.32) with respect to (ϕ, ρ, σ^2) to obtain marginal maximum likelihood estimates.

3. Case studies and numerical investigation

3.1 Case studies

3.1.1 Conical model

The girth and to a lesser extent the height, are easily measured, but it is the volume of usable timber that determines the value of a tree. The aim is therefore to find a formula for predicting volume from the other two measurements. Table 1.1 contains 70 observations on the volume in cubic feet of shortleaf pine, from Bruce and Schumacher (1935) together with x_1, the girth of each tree, that is, the diameter at breast height, in inches and x_2, the height of the tree in feet. Atkinson and Rinani (2000) suggests a conical model

$$f(\boldsymbol{x}; \beta_1) = \beta_1 x_1^2 x_2. \tag{3.1}$$

The trees are arranged in the table from small to large, so that one indication of a systematic failure of a model would be the presence of anomalies relating to the smallest or the largest observations. Atkinson and Riani (2000) uses a PTB approach with six kinds of λ to investigate transformations for these data. Finally, they conclude that the log transformation is supported by all the data. Table 1 shows the estimates of $\boldsymbol{\beta}$, λ, ϕ, ρ and L for each approach. Also, to evaluate skewness and heteroscedasticity of residuals of predicted values, we calculated mean of absolute values of skewness for the error and mean of absolute values of Spearman rank correlation between residuals and predicted values in Table 1. The result of the estimates $\hat{\beta}_1$ of PTB, PWT, PTBWLS and NTBWLS were almost the same excluding the estimates of OLS. From the result of the estimates $\hat{\lambda}$ of PTB, it was near 0 and hence log transformation model was suggested as a transform-both-sides model as well as the results of Atkinson and Riani (2000). For the results

Table 3.1.1.1: Results for the estimates of parameters and the skewness and heteroscedasticity of residuals of predicted values for each approach

Parameters	OLS	PTB	PWT	PTBWLS	NTBWLS
$\hat{\beta}_1$	0.00298	0.00306	0.00306	0.00307	0.00306
$\widehat{SE}(\hat{\beta}_1)$	0.000014	0.000024	0.000024	0.000024	0.000024
$\hat{\lambda}$	$-$	0.049	$-$	-0.184	$-$
$\widehat{SE}(\hat{\lambda})$	$-$	0.023	$-$	0.018	$-$
$\hat{\phi}$	$-$	$-$	0.145	1.724	1.015
$\widehat{SE}(\hat{\phi})$	$-$	$-$	0.021	0.093	0.076
log-likelihood	-105.41	-64.75	-66.04	-64.43	$-$
$\hat{\rho}$	$-$	$-$	$-$	$-$	41.86
SRC	0.677	0.043	0.017	0.049	0.001
skewness	1.456	0.333	0.648	0.346	0.256

of log-likelihood estimates, the estimate of OLS was smaller than the estimates of other approaches. SRC shows a Spearman rank correlation between residuals and predicted values and skewness shows a degree of skew for the error distribution. That is, SRC and skewness can be considered as the indicator for a heteroscedasticity and a normality of the error. From the results of SRC, it is considered to have heteroscedasticity for the error as there is the correlation of 0.677 in the result of OLS. The SRC of PWT, PTBWLS and NTBWLS were smaller than the results of OLS, especially the results of NTBWLS was near 0.

3.1.2 Ricker model and Beverton & Holt model

When managing a fishery, one must model the relationship between the size of the annual spawning stock and its production of new catchable-sized fish, called recruits or returns. There are several theoretical models relating recruits and spawners. These are derived from simple assumptions about factors influencing the survival of juvenile fish. All spawner-recruit models known to us are deterministic, i.e., the response Y is nonrandom given $\boldsymbol{X}$, though Y itself can depend on stochastic variables. If the biological and physical factors affecting fish survival were constant from year to year, then a deterministic model would be realistic since abundance of fish makes the law of large numbers applicable. However, for most fish stocks these factors are far from constant. There has

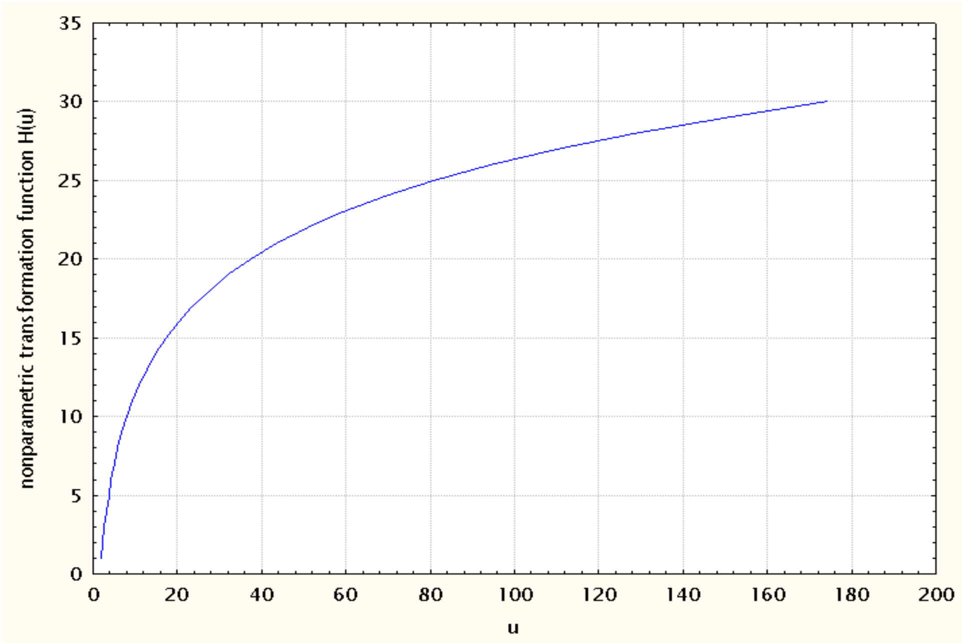

Figure 3.1.1.1: Estimated nonparametric transformation function in NTBWLS

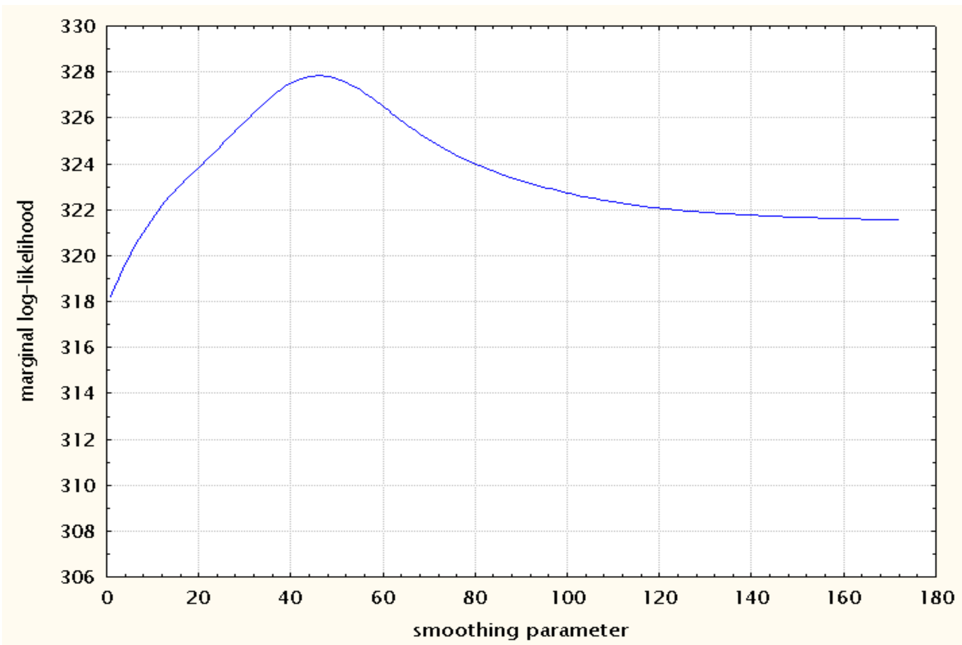

Figure 3.1.1.2: Marginal log-likelihood estimates for each ρ value in NTBWLS

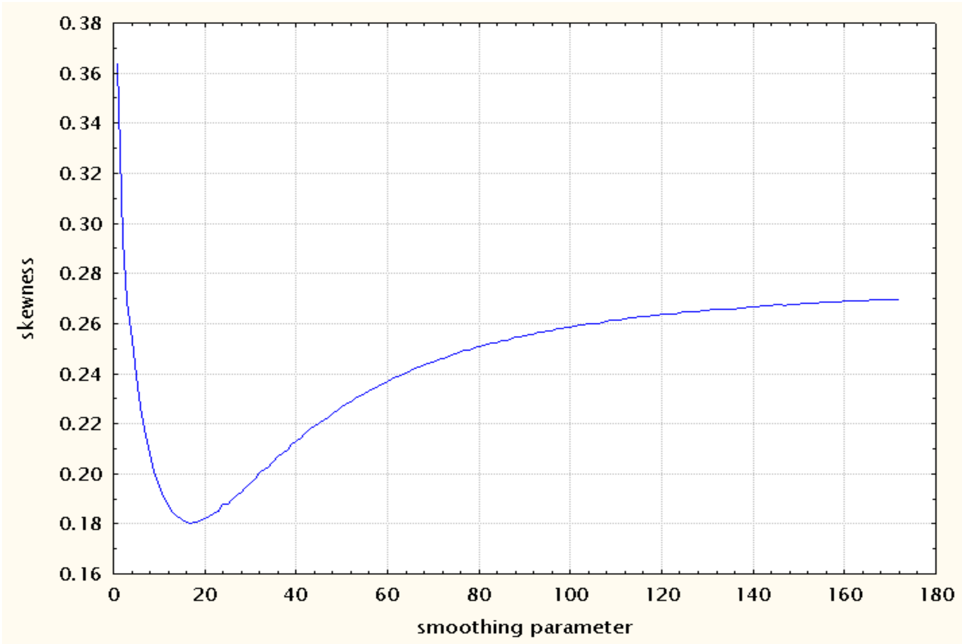

Figure 3.1.1.3: Skewness for the error distribution of each ρ value in NTBWLS

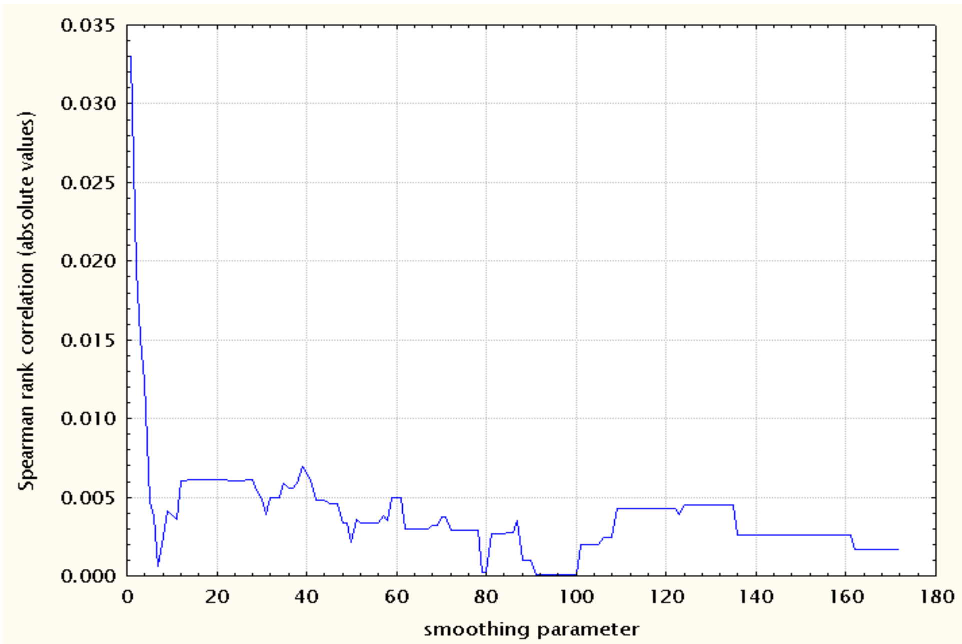

Figure 3.1.1.4: Spearman rank correlation of each ρ value in NTBWLS

been little work on stochastic models for recruitment , probably because the mechanisms causing survival rates to vary are not well understood. It is common practice to take a deterministic model relating Y and $\boldsymbol{X}$ and to assume multiplicative lognormal errors. The transform-both-sides approach allows us to test this assumption, and to model the errors empirically when the assumption seems unwarranted. Ricker (1954) derived the theoretical deterministic model

$$f(x;\boldsymbol{\beta}) = \beta_1 x \exp(-\beta_2 x) \tag{3.2}$$

In this model $f(x;\boldsymbol{\beta})$ tends to 0 as x goes to 0, as would be expected in any realistic model. Moreover, $f(x;\boldsymbol{\beta})$ has a maximum at β_2^{-1}, provided β_2 is strictly positive, and $f(x;\boldsymbol{\beta})$ tends to 0 as x goes to ∞. The biological interpretation of this behavior is that as the number of juveniles increases, increased competition and predation affect the survival rate so drastically that the absolute number of juveniles reaching maturity decreases.
A second model was derived by Beverton and Holt (1957), namely

$$f(x;\boldsymbol{\beta}) = \frac{1}{\beta_1 + \beta_2/x}, \quad \beta_1 \geq 0, \quad \beta_2 \geq 0. \tag{3.3}$$

The Beverton-Holt model also has the characteristic that Y tends to 0 as x tend to 0, but Y increases asymptotically to $1/\beta_1$ as Y tends to ∞. It is natural to think of $1/\beta_1$ as the carrying capacity of the environment, the maximum number of recruits that the available space, food and other resources can support. When fit to the same data set, the Ricker and Beverton-Holt models are often similar over the range of spawner values in the data, despite qualitatively different behavior as the number of spawners increases to infinity.

Ricker and Smith (1975) give numbers of spawners and recruits from 1940 until 1967 for the Skeena River sockeye salmon stock. The objectives here are the following two; 1) to compare the results of OLS, PTB, PWT, PTBWLS and NTBWLS and to estimate the performance of the model, 2) to confirm NTB corresponds PTB approximately when the smoothing parameter ρ is set with large value. Table 3.1.2.1 shows the results of each parameter. For $\hat{\beta}_1$ and $\hat{\beta}_2$, the standard error of PTB, PWT, PTBWLS and NTBWLS were smaller than that of OLS. The both $\hat{\beta}_1$ and $\hat{\beta}_2$ of standard error of NTBWLS were the smallest in all approaches. $\hat{\phi}$ in PWT and PTBWLS were estimated near 0. It can be thought that these models have heteroscedasticity for the error and the variance

Table 3.1.2.1: Results for the estimates of parameters and the skewness and heteroscedasticity of residuals of predicted values for each approach

Parameters	OLS	PTB	PWT	PTBWLS	NTBWLS
$\hat{\beta}_1$	3.79	3.29	3.24	3.19	2.76
$\widehat{SE}(\hat{\beta}_1)$	1.25	0.80	0.60	0.68	0.52
$\hat{\beta}_2$	0.00080	0.00070	0.00057	0.00058	0.00047
$\widehat{SE}(\hat{\beta}_2)$	0.00041	0.00033	0.00030	0.00033	0.00030
$\hat{\lambda}$	–	0.314	–	0.735	–
$\widehat{SE}(\hat{\lambda})$	–	0.021	–	0.107	–
$\hat{\phi}$	–	–	−0.041	0.019	1.167
$\widehat{SE}(\hat{\phi})$	–	–	0.022	0.028	0.121
log-likelihood	−190.27	−186.47	−185.88	−185.66	–
$\hat{\rho}$	–	–	–	–	0.00156
SRC	0.545	0.308	0.150	0.172	0.102
skewness	0.407	−0.328	−0.057	−0.068	−0.383

function distributed exponential function. From the results of SRC, it is considered to have heteroscedasticity for the error as there is the correlation of about 0.5 in the result of OLS. The SRC of PWT, PTBWLS and NTBWLS were smaller than the results of OLS and PTB. This results shows that the heteroscedasticity for the error was improved by weighted transformation with parameter ϕ in PWT, PTBWLS and NTBWLS. On the other hand, From the results of skewness, the skewness of PWT and PTBWLS were smaller than the results of OLS, PTB and NTBWLS. In NTBWLS, $\hat{\rho}$ was estimated 0.00156, therefore the necessity of the transform-both-sides was suggested because it was considerably small. Figure 3.1.2.1-3.1.2.3 show the results of estimated nonparametric transformation function $\hat{H}(u)$ and the marginal log-likelihood estimates and skewness for the error distribution when the value of ρ was gradually moved. The maximum marginal log-likelihood estimate was 0.00156 and the minimum absolute value of skewness was about 0.001, therefore it can be considered that to optimize a marginal likelihood corresponds to optimize a symmetry of the error distribution. Figure 3.1.2.4 shows the fitting plot of Ricker model by each approach. OLS, PTB and PTBWLS showed the saturation of Y in case over $X = 1,200$, but NTBWLS did not show such the saturations.

Carroll and Ruppert (1988) analyzed the skeena data with exception of one data. A rockslide occurred in 1951 and severely reduced the number of recruits. So, we conduct

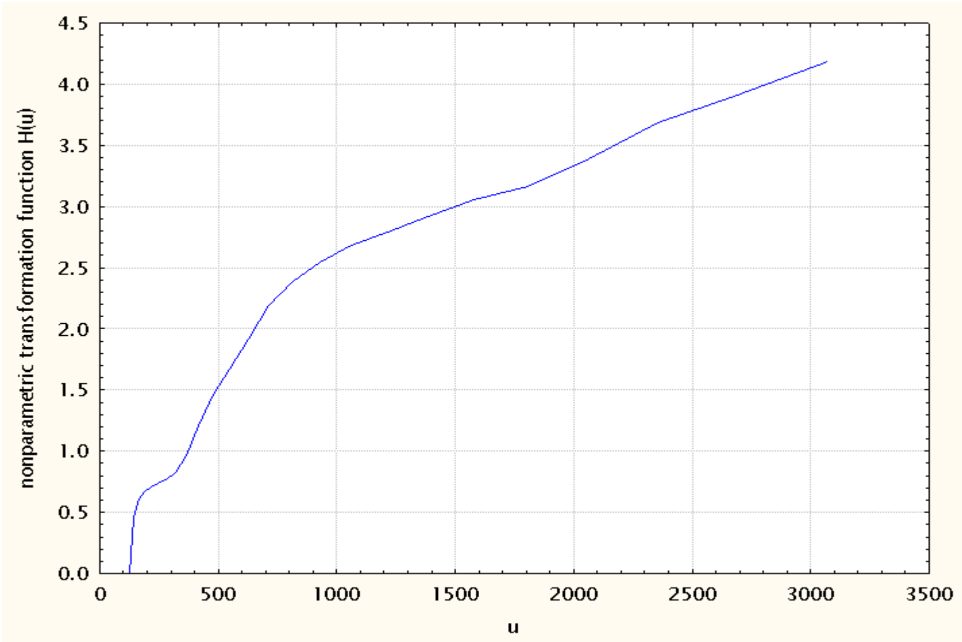

Figure 3.1.2.1: Estimated nonparametric transformation function in NTBWLS

the outlier analysis by excepting an observation in 1951 as well and investigate the robustness of each estimator for the model parameters. Table 3.1.2.2 shows the results of each parameter when we do not use an observation of year 1951. There were the decent differences between full data and exception data for the parametric transformation approaches. The difference of NTBWLS was smallest, so we can consider that NTBWLS gave the most robust estimates for the model parameters. Figure 3.1.2.4-3.1.2.6 show the results of estimated nonparametric transformation function $\hat{H}(u)$ and the marginal log-likelihood estimates and skewness for the error distribution when the value of ρ was gradually moved. Figure 3.1.2.7 shows the fitting plot of Ricker model by each approach with exception of an observation of year 1951. OLS, PTB and PTBWLS showed the saturation of Y in case over $X = 800$ to $X = 1,000$, but NTBWLS did not show such the saturations.

3.2 Numerical investigation for 1-compartment model

Channer and Roberts (1985) studied the effect of delayed esophageal transit on the absorption of acetaminophen. Patients awaiting cardiac catheterization took a single 500-

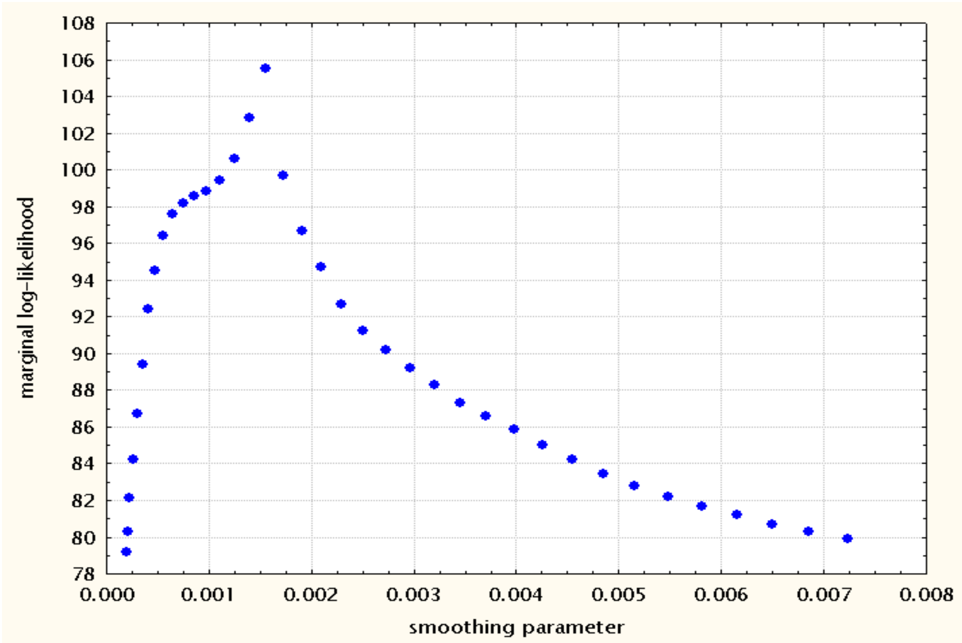

Figure 3.1.2.2: Marginal log-likelihood estimates for each ρ value in NTBWLS

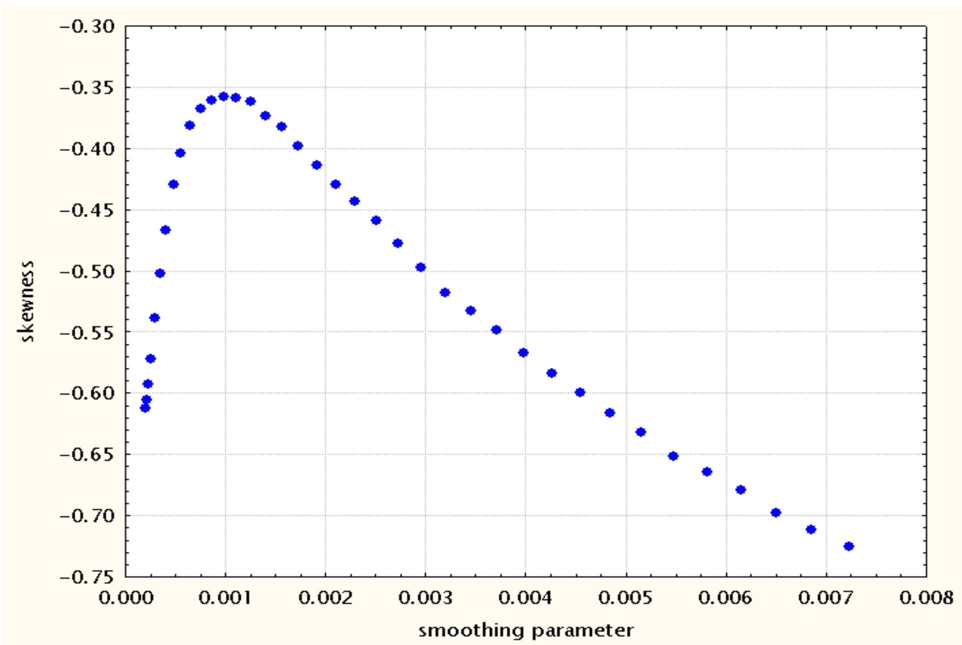

Figure 3.1.2.3: Skewness for the error distribution for each ρ value in NTBWLS

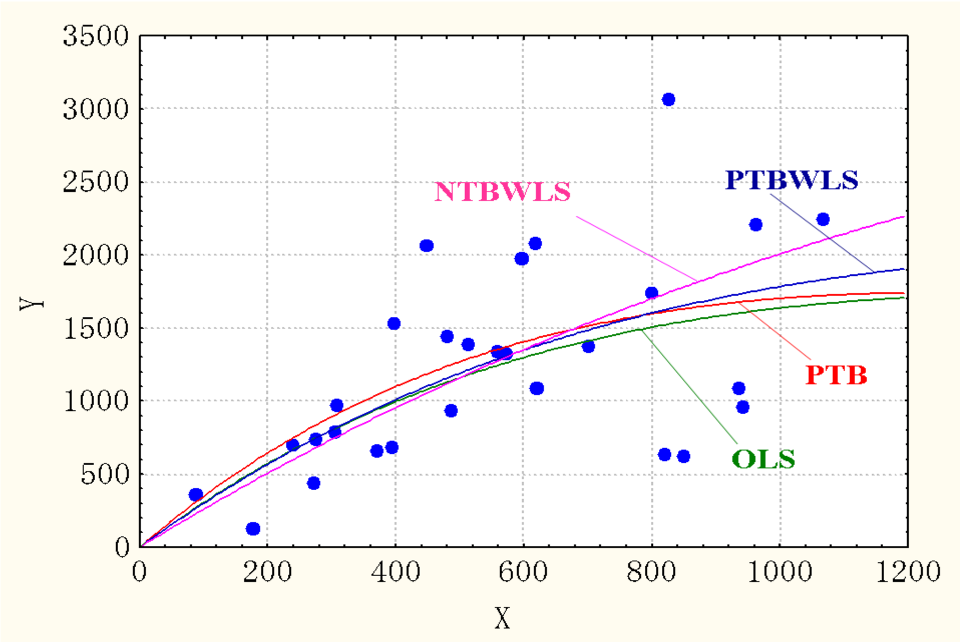

Figure 3.1.2.4: Fitting plot of Ricker model for each approach

Table 3.1.2.2: Results for the estimates of parameters and the skewness and heteroscedasticity of residuals of predicted values for each approach (do not use an observation of year 1951)

Parameters	OLS	PTB	PWT	PTBWLS	NTBWLS
$\hat{\beta}_1$	3.92	3.78	4.16	4.09	2.89
$\widehat{SE}(\hat{\beta}_1)$	1.33	0.80	0.39	0.45	0.72
$\hat{\beta}_2$	0.00085	0.00095	0.00096	0.00099	0.00062
$\widehat{SE}(\hat{\beta}_2)$	0.00042	0.00031	0.00017	0.00021	0.00035
$\hat{\lambda}$	–	-0.203	–	0.425	–
$\widehat{SE}(\hat{\lambda})$	–	0.023	–	0.021	–
$\hat{\phi}$	–	–	-1.023	-2.015	1.007
$\widehat{SE}(\hat{\phi})$	–	–	0.023	0.045	0.574
log-likelihood	-183.61	-178.49	-177.52	-176.83	–
$\hat{\rho}$	–	–	–	–	0.00767
SRC	0.584	0.281	0.065	0.070	0.026
skewness	0.429	-0.467	0.202	-0.168	-0.868

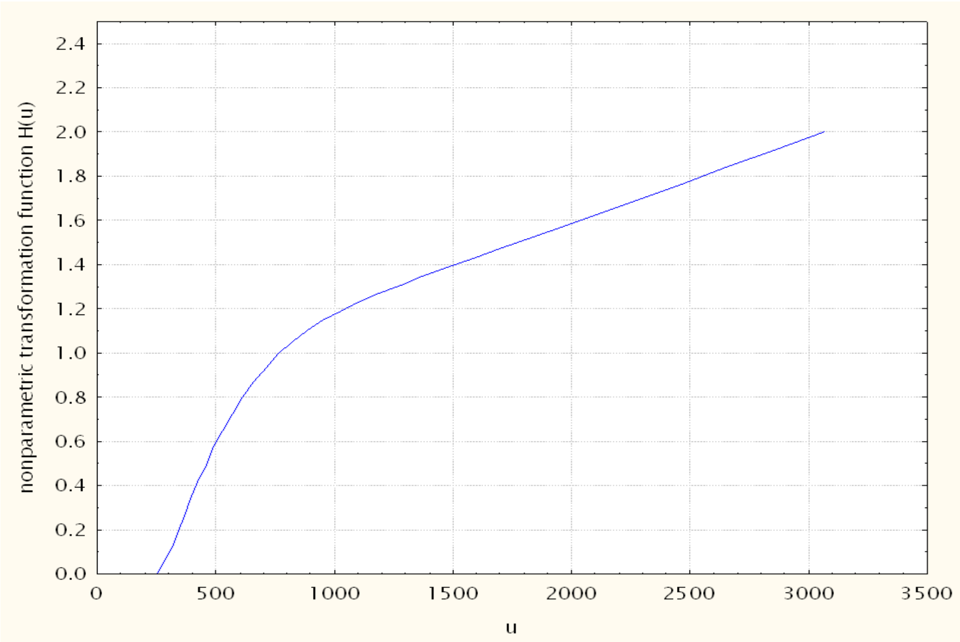

Figure 3.1.2.5: Marginal log-likelihood estimates for each ρ value in NTBWLS (do not use an observation of year 1951)

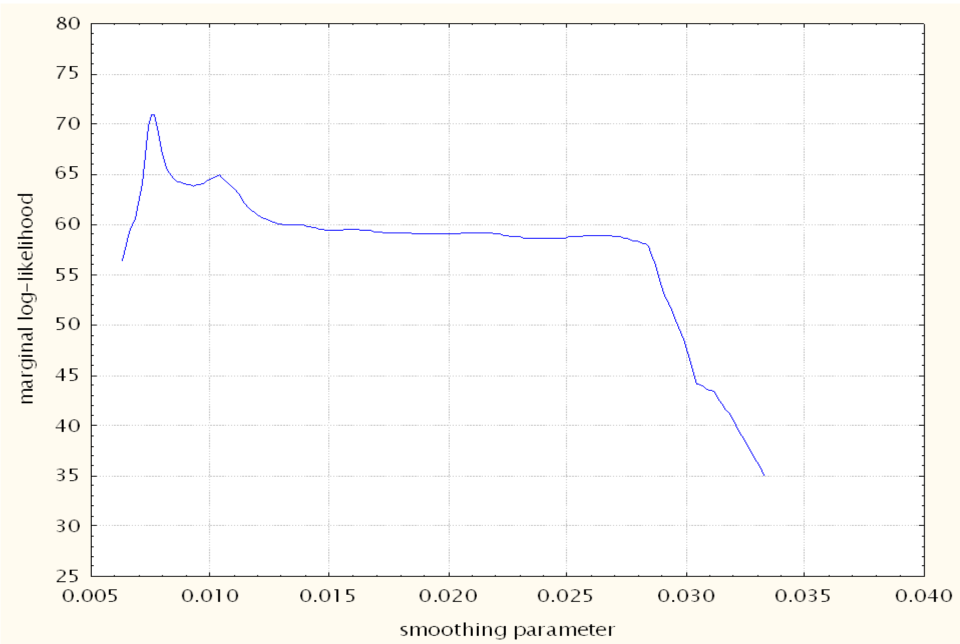

Figure 3.1.2.6: Skewness for the error distribution for each ρ value in NTBWLS (do not use an observation of year 1951)

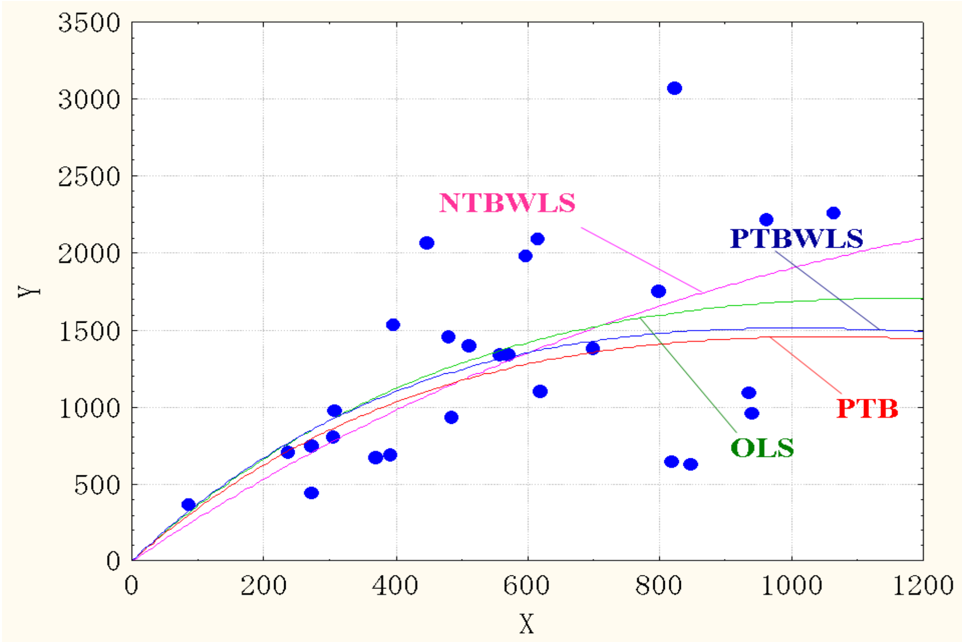

Figure 3.1.2.7: Fitting plot of Ricker model for each approach (do not use an observation of year 1951)

milligram tablet containing acetaminophen and barium sulfate. The blood drug concentration in the systematic circulation compartment (non-linear predictive function) is

$$f(t;\boldsymbol{\beta}) = \frac{500K_{12}}{\mathscr{V}_1(K_{12}-K_{20})}\left\{\exp(-K_{20}t)-\exp(-K_{12}t)\right\}, \tag{3.4}$$

where t is the time following administration, $\mathscr{V}_1$ is the volume of distribution, K_{12} is the first-order absorption rate constant, K_{20} is the first-order elimination rate constant and $\boldsymbol{\beta} = (\mathscr{V}_1, K_{12}, K_{20})^{\mathrm{T}}$. We generate random numbers for the parameter estimation of the 1-compartment model in the example data. The goals are to assess how much each method can improve non-normality and heteroscedasticity. For the example data, the estimates of ordinary least square (OLS) were $\mathscr{V}_1 = 69.48, K_{12} = 0.0686$, and $K_{20} = 0.0084$, then we set these estimates as the true value in this numerical study. The number of time points are set to 13 points like the example. In this situation, in order to generate data with heteroscedasticity, in consideration of large variance of the blood drug concentration near the time to attain maximum concentration (Tmax), we obtained simulated data as follows. For $t = 10, 20, 30, 40, 50$ and 60, we generated 100 sets of random numbers to distribute independent normally with mean "true value" and variance "0.4 or 0.6" about each variance. For $t = 90, 120, 150, 180, 210, 240$ and 360, we generated 100 sets of random numbers to distribute independent normally with mean "true value" and variance "0.1". Namely, the variance of blood drug concentration for $t = 10, 20, 30, 40, 50$ and 60 is 4 times or 6 times of the variance of $t = 90, 120, 150, 180, 210, 240$ and 360. Data generated by the above approach were fitted to the 1-compartment model by use of OLS, PTB, PWT, PTBWLS and NTBWLS, and the approaches are assessed by the estimates of mean square error. For the selection of smoothing parameter in NTBWLS, we selected based on skewness and heteroscedasticity of residuals of predicted values. That is, we set $\rho = 0.001$ to $\rho = 10^5$ at decuple intervals and calculated mean of absolute values of skewness for the error and mean of absolute values of Spearman rank correlation between residuals and predicted values for each smoothing parameter, and we selected the smoothing parameter whose skewness and correlation were the smallest. In the result, both of these statistics were the smallest when the smoothing parameter was 0.01, so we selected it. As well, the estimates of regression parameters did not converge for $\rho < 0.001$. Figure 1 shows the results of mean square errors by use of each approach for each parameter of the 1-compartment model in the case of $\sigma_o^2 = 0.4$ and $\sigma_o^2 = 0.6$. Table 1-3 show the squared bias and the variance of the estimators for each mean square error, and Table 4 shows the mean absolute values of skewness for the error and Spearman rank correlation between residuals and predicted values for each approach.

From the results of $\mathscr{V}_1$, the MSE of NTBWLS was the smallest. In the case of $\sigma_o^2 = 0.6$, the MSE of PWT, PTBWLS and NTBWLS was particularly smaller than the MSE of OLS and PTB. This suggested that performance of the power weighted transformation was high. Next, from results of the first-order absorption rate constant K_{12}, the results of all methods were similar all in the case of $\sigma_o^2 = 0.4$, but the MSE of PTB, PWT, PTBWLS and NTBWLS were smaller than the MSE of OLS in the case of $\sigma_o^2 = 0.6$. From the results of the first-order elimination rate constant K_{20}, the MSE was decreasing in order of NTBWLS, PTBWLS, PWT, PTB, OLS. Also, from the results of the squared bias and variance of the estimators in Table 1-3, the variance of all estimators were decreasing in order of NTBWLS, PTBWLS, PTB, OLS. Finally, from the results of the skewness for the error and the Spearman rank correlation between residuals and predicted values in Table 4, these statistics were improved in order of NTBWLS, PTBWLS, PTB, OLS.

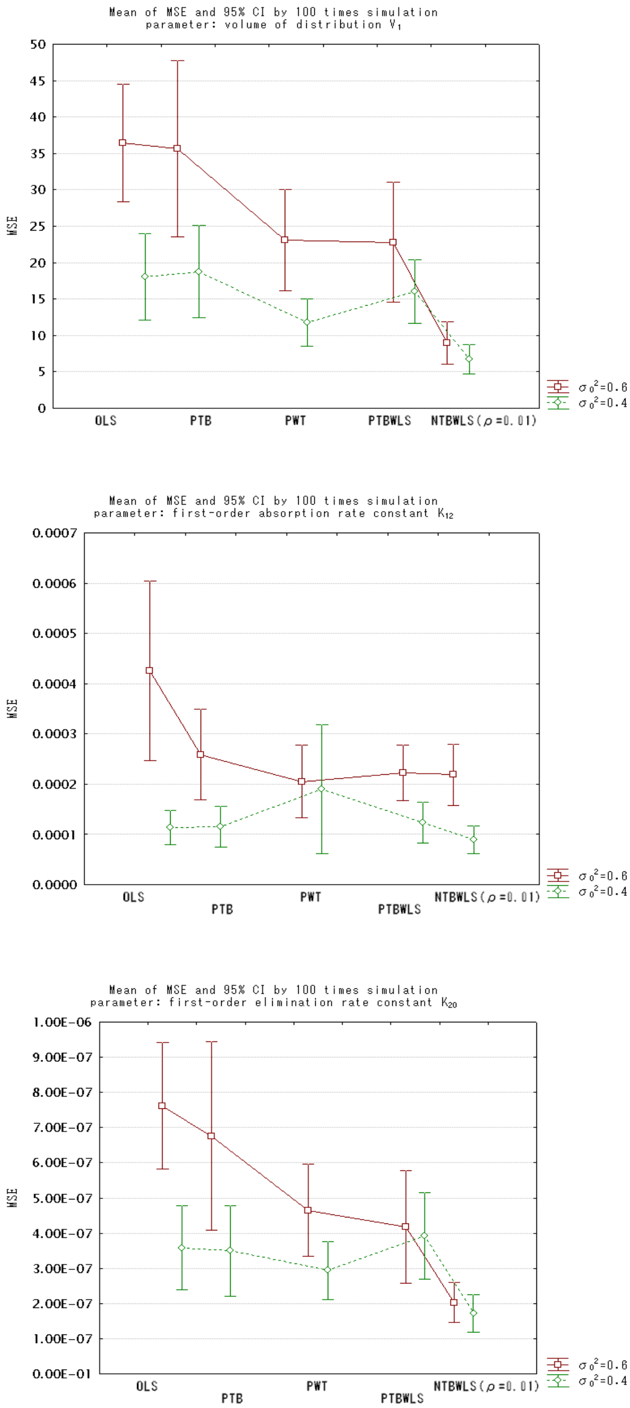

Figure 3.2.1: Results of MSE for each parameter of 1-compartment model based on 100 times of simulation study

Table 3.2.1: Results of the squared bias and variance of the estimators for the volume of distribution ($\mathrm{E}[\mathscr{V}_1]$)

approach	$\sigma_0^2 = 0.4$		$\sigma_0^2 = 0.6$	
	bias	variance	bias	variance
OLS	0.001	18.04	0.38	36.09
PTB	0.02	18.72	0.63	35.02
PTBWLS	0.09	15.97	0.94	21.87
NTBWLS($\rho = \infty$)	0.04	11.67	0.06	21.16
NTBWLS($\rho = 1$)	0.11	13.51	0.42	11.18
NTBWLS($\rho = 0.01$)	0.31	6.44	0.14	8.81

Table 3.2.2: Results of the squared bias and variance of the estimators for the first-order absorption rate constant ($\mathrm{E}[K_{12}]$)

approach	$\sigma_0^2 = 0.4$		$\sigma_0^2 = 0.6$	
	bias	variance	bias	variance
OLS	4.20×10^{-7}	1.13×10^{-4}	8.64×10^{-5}	3.67×10^{-3}
PTB	3.79×10^{-7}	1.15×10^{-4}	7.40×10^{-8}	2.59×10^{-4}
PTBWLS	7.74×10^{-9}	1.21×10^{-4}	9.29×10^{-6}	2.13×10^{-4}
NTBWLS($\rho = \infty$)	7.11×10^{-8}	9.98×10^{-5}	9.49×10^{-6}	3.49×10^{-4}
NTBWLS($\rho = 1$)	1.26×10^{-6}	1.24×10^{-4}	3.36×10^{-5}	2.08×10^{-4}
NTBWLS($\rho = 0.01$)	4.50×10^{-7}	8.86×10^{-5}	2.54×10^{-5}	1.93×10^{-4}

Table 3.2.3: Results of the squared bias and variance of the estimators for the first-order elimination rate constant ($\mathrm{E}[K_{20}]$)

approach	$\sigma_0^2 = 0.4$		$\sigma_0^2 = 0.6$	
	bias	variance	bias	variance
OLS	1.44×10^{-10}	3.58×10^{-7}	1.25×10^{-10}	7.61×10^{-7}
PTB	2.31×10^{-10}	3.50×10^{-7}	2.93×10^{-8}	6.47×10^{-7}
PTBWLS	2.56×10^{-10}	3.92×10^{-7}	9.53×10^{-9}	4.08×10^{-7}
NTBWLS($\rho = \infty$)	2.18×10^{-9}	4.28×10^{-7}	1.60×10^{-9}	4.65×10^{-7}
NTBWLS($\rho = 1$)	1.67×10^{-8}	3.13×10^{-7}	5.27×10^{-8}	2.40×10^{-7}
NTBWLS($\rho = 0.01$)	4.98×10^{-9}	1.67×10^{-7}	3.69×10^{-8}	1.66×10^{-7}

Table 3.2.4: Results of the mean absolute values of skewness for the error and Spearman rank correlation between residuals and predicted values for each approach

approach	$\sigma_0^2 = 0.4$		$\sigma_0^2 = 0.6$	
	skewness	rank correlation	skewness	rank correlation
OLS	8.39×10^{-2}	0.264	2.23×10^{-2}	0.504
PTB	1.04×10^{-2}	0.053	9.36×10^{-2}	0.037
PTBWLS	1.41×10^{-2}	0.019	1.39×10^{-2}	0.048
NTBWLS($\rho = \infty$)	2.64×10^{-2}	0.022	2.09×10^{-2}	0.035
NTBWLS($\rho = 1$)	5.56×10^{-3}	0.015	1.27×10^{-2}	0.023
NTBWLS($\rho = 0.01$)	1.04×10^{-2}	0.032	1.01×10^{-2}	0.029

3.3 Discussions

In the investigation result of the conical model, the transformation approaches (PTB, PWT, PTBWLS and NTBWLS) much improved the heteroscedasticity and the non-normality of the error in a post-transformation compared with that of OLS. It suggested that these transformation approaches are effective to fit a simple non-linear model for heteroscedastic data. The estimate of power-parameter of PTB was near 0. It suggested that log-transformation was selected as the optimal transformation function and it corresponded with the conclusion of Atkinson & Riani (2000). The result of SRC for NTBWLS was near 0 and it showed that NTBWLS was best transformation as the variance stabilization. However, the estimates of model parameter were nearly same between these transformation approaches. It suggested that PTB or PWT were enough to improve heteroscedasticity of the error in this data.

From the results of Ricker model and Beverton & Holt model, the transformation approaches (PTB, PWT, PTBWLS and NTBWLS) much improved the heteroscedasticity and the non-normality of the error in a post-transformation compared with that of OLS. Also, PWT, PTBWLS and NTBWLS improved them more than PTB. It probably means that the power-weighted transformation was more appropriate than the both-sides transformation for the Skeena salmon data. The approach in which the absolute value of SRC was the smallest was NTBWLS, but the absolute values of skewness of the error distribution in PWT and PTBWLS were smaller than that of NTBWLS. We should discuss which approaches are superior in the performance of parameter estimation in the simulation experiments in Section 4.

In the numerical investigation based on 1-compartment model, the results suggested that performance of NTBWLS was high. From the results of the variance of all estimators were decreasing in order of NTBWLS, PTBWLS, PTB, OLS. Therefore, we concluded that NTBWLS is superior to the other method in the situation of hardy heteroscedasticity and provides a robust estimator to make the smoothing parameter smaller because it reduces the effect of the intensity of heteroscedasticity.

4. Simulation

4.1 Motivations and Objectives

The following hypotheses can be established from the results of case studies:

Hypothesis 1: In the case the error for a true non-linear model is distributed non-normal with constant variance, the transformation approaches are superior to usual least square approach in the performance for the model-parameter estimation.

Hypothesis 2: Based on the Hypothesis 1, in the case the error for a true non-linear model is distributed non-normal with heteroscedasticity, PTBWLS and NTBWLS are superior to PTB and PWT in the performance for the model-parameter estimation.

Hypothesis 3: Based on the Hypothesis 2, in the case the true transformation is not included in a power function family and the error is distributed non-normal with heteroscedasticity, a optimal smoothing parameter estimate in NTBWLS does not diverge but give certain value and NTBWLS improves the performance of estimation for model parameters more than PTB, PWT and PTBWLS.

To confirm Hypothesis 1, we build the model which is distributed normal with constant variance in the case the both sides of model is log-transformed. That is, a log-transformation is assumed as the true variance stabilization transformation. To confirm Hypothesis 2, we generate a non-constant variance by using the variance function that is proportional to the predictor in the model of Hypothesis 1. In the investigation for Hypothesis 3, we use a complicated function as a true transformation function in place of a log-transformation in case of Hypothesis 2.

4.2 Designs

4.2.1 Simulation design for Hypothesis 1

As the simulation model, we set

$$f(x; \boldsymbol{\beta}) = x\beta_1 \exp(-\beta_2 x), \tag{4.1}$$

and consider about the nonparametric transform-both-sides model

$$H_{\mathrm{N}}(y) = H_{\mathrm{N}}\{f(x; \boldsymbol{\beta})\} + \varepsilon_{\mathrm{N}}. \tag{4.2}$$

Where, ε_{N} is distributed $\mathrm{N}(0, \sigma^2)$. For the model (4.2), we give a transformation function

$$H_1(u) = \log(u). \tag{4.3}$$

The true transformation function in the model (4.2) is a log-transformation and it is included in a power-transformation family. That is, the true value for a smoothing parameter is $\rho \to \infty$ in NTB. The simulation model obtained from (4.2) and (4.3) is as follows:

$$Y = [\beta_1 x \exp(-\beta_2 x)] \exp(\varepsilon_{\mathrm{N}}). \tag{4.4}$$

We set $\beta_1 = 3,\ \beta_2 = 0.0008$ as the true values for the model parameters β_1 and β_2. This is based on the presumption result of case study for the Ricker model in Section 3.2.2. The observable range for predictor variable x is $0 < x < 1,000$. We provide the sample size and the error variance as simulation factors with three kinds of levels that influence the results.

4.2.2 Simulation design for Hypothesis 2

The model (4.1) is assumed to be a potential model as well as Hypothesis 1. However, we provide the heteroscedastic variance for the error. We set that ε_{N} is distributed $\mathrm{N}(0, \sigma_n^2)$ as the error, where σ_n^2 is proportional to the predictor variable x. In this situation, there are not only the heteroscesasticity of the error variance, but the non-normality of the error distribution. We set $\beta_1 = 3$, $\beta_2 = 0.0008$ as the true values for the model parameters β_1 and β_2 and the acceptable observation range for predictor variable x is $0 < x < 1,000$ as well as the case of Hypothesis 1.

4.2.3 Simulation design for Hypothesis 3

In setting Hypothesis 3, (4.1) is assumed to be a potential model as well as Hypothesis 2. In order to have the model that the true transformation function is not included in a power-transformation family, we set the true transformation function as follows:

$$H_2(u) = \log(u/\sqrt{X}). \tag{4.5}$$

Where, ε_{N} is distributed $\mathrm{N}(0, \sigma_n^2)$. In the context of NTBWLS approach, there is a certain ρ_0 and $H_{\mathrm{N}}(u; \rho_0) = H_2(u)$. From (4.1) and (4.5), the simulation model is obtained as follows:

$$Y = [\beta_1 x \exp(-\beta_2 x)] \exp(\sqrt{x}\varepsilon_{\mathrm{N}}). \tag{4.6}$$

Other settings are same as the Hypothesis 1 and Hypothesis 2.

4.3 Generating simulation data

For the all combinations of simulation factors that provide in Hypothesis 1, we generated the N uniform random numbers defined on $[0,\ 1,000]$ and the N normal random numbers distributed $\mathrm{N}(0, \sigma^2)$. In case that Hypothesis 2 and Hypothesis 3, we generated the N uniform random numbers defined on $[0,\ 1,000]$ and the N normal random numbers distributed $\mathrm{N}(0, \sigma^2 x)$. It was replicated 1,000 times. We conducted the Bartlett test for a homoscedasticity to set the meaningful sample size to target heteroscedasticity of the error variance. That is, in the situation of Hypothesis 1, we divided into two datasets as group A ($0 < X < 500$) and group B ($500 < X < 1,000$) and for the variance σ_A^2 and σ_B^2 of group A and group B, in case that we test the null hypothesis $\mathrm{H}_0 : \sigma_A^2 = \sigma_B^2$ against the alternative hypothesis $\mathrm{H}_1 : \sigma_A^2 \neq \sigma_B^2$ with a 0.05 two-sided significance level and the error variance $\sigma^2 = 0.02$, then the sample size was $N = 29$ with the power 0.80, $N = 41$ with the power 0.90 and $N = 63$ with the power 0.95. Therefore, the sample size was set as $N = 29, 41, 63$ and the error variance was set as $\sigma^2 = 0.01, 0.02, 0.03$ in the simulation to conduct the significant simulation experiments. The sample size in the investigation of Hypothesis 2 and Hypothesis 3 were same as Hypothesis 1 in view of a comparability.

The OLS, PTB, PWT, PTBWLS and NTBWLS were applied for the inference on a simulation model to the data generated based on above setting, and mean square error of $\hat{\beta}_1$ [MSE($\hat{\beta}_1$)], $\hat{\beta}_2$ [MSE($\hat{\beta}_2$)] and the results of resolving these the variance and the bias [$\mathrm{VAR}(\hat{\beta}_1), \mathrm{BIAS}(\hat{\beta}_1)$ and $\mathrm{VAR}(\hat{\beta}_2), \mathrm{BIAS}(\hat{\beta}_2)$] were calculated for each approach. And, to assess the normality and homoscedasticity of the error after the transformation, the mean absolute values of skewness for the error (Skewness) and Spearman rank correlation between residuals and predicted values (SRC) were calculated. For the Hypothesis 1, we included the true simulation model as the contrast of each approach.

4.4 Results and Interpretations

4.4.1 Result and Interpretation from Simulation of Hypothesis 1

Table 4.4.1.1 to Table 4.4.1.3 show the results of the simulation for Hypothesis 1 in case that sample size was $n = 29$ and for $\sigma^2 = 0.01$, $\sigma^2 = 0.02$ and $\sigma^2 = 0.03$, respectively. From the results in case that the sample size is small (Table 4.4.1.1 to Table 4.4.1.3), All transformation approaches were superior to OLS in terms of the mean square errors. In addition, in the situation of the larger variance, the greater those differences. It may be shown that the transformation approaches caught the structure of the true model (that is, log-transformational model) as compare to OLS approach. In a view of the error distribution, the transformation approaches improved SRC better than that of OLS, but the improving of the skewness was not showed in any transformation approaches. It seems the natural result because we did not give the skewness for the error distribution of the true model but give the heteroscedasticity of that intentionally. In particular, in case of $\sigma^2 = 0.03$, we can find that the MSE of PWT and that of PTBWLS were smaller than OLS and PTB. It can be thought that the power-weighted transformation approach improves a hardy heteroscedasticity as compare to the transformation-both-sides approach. For the situation of the larger sample size $n = 41$ and $n = 63$ (Table 4.4.1.4 to Table 4.4.1.9), it seems that we can give the same interpretations as Table 4.4.1.1 to Table 4.4.1.3.

Table 4.4.1.1: Simulation results for the Hypothesis 1: $n = 29$, $\sigma^2 = 0.01$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0044	0.0037	0.0036	0.0037	0.0031
BIAS($\hat{\beta}_1$)	2.62×10^{-8}	2.03×10^{-6}	6.12×10^{-6}	6.92×10^{-5}	1.18×10^{-5}
VAR($\hat{\beta}_1$)	0.0044	0.0037	0.0036	0.0037	0.0030
MSE($\hat{\beta}_2$)	8.78×10^{-10}	8.98×10^{-10}	8.28×10^{-10}	7.95×10^{-10}	7.39×10^{-10}
BIAS($\hat{\beta}_2$)	2.07×10^{-11}	5.56×10^{-12}	1.01×10^{-12}	4.63×10^{-12}	1.85×10^{-12}
VAR($\hat{\beta}_2$)	8.57×10^{-10}	8.93×10^{-10}	8.27×10^{-10}	7.91×10^{-10}	7.37×10^{-10}
SRC	0.180	0.192	0.092	0.121	0.373
Skewness	0.288	0.302	0.311	0.369	0.141

Table 4.4.1.2: Simulation results for the Hypothesis 1: $n = 29$, $\sigma^2 = 0.02$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0166	0.0109	0.0166	0.0127	0.0099
BIAS($\hat{\beta}_1$)	3.67×10^{-4}	1.28×10^{-5}	3.33×10^{-5}	4.91×10^{-6}	3.73×10^{-6}
VAR($\hat{\beta}_1$)	0.0162	0.0109	0.0160	0.0127	0.0099
MSE($\hat{\beta}_2$)	4.32×10^{-9}	2.68×10^{-9}	3.66×10^{-9}	2.75×10^{-9}	2.70×10^{-9}
BIAS($\hat{\beta}_2$)	1.67×10^{-11}	3.19×10^{-13}	5.01×10^{-11}	4.00×10^{-13}	3.74×10^{-12}
VAR($\hat{\beta}_2$)	4.30×10^{-9}	2.68×10^{-9}	3.66×10^{-9}	2.75×10^{-9}	2.70×10^{-9}
SRC	0.178	0.149	0.110	0.101	0.386
Skewness	0.337	0.359	0.338	0.286	0.159

Table 4.4.1.3: Simulation results for the Hypothesis 1: $n = 29$, $\sigma^2 = 0.03$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0321	0.0318	0.0263	0.0274	0.0252
BIAS($\hat{\beta}_1$)	6.18×10^{-4}	1.05×10^{-4}	4.56×10^{-4}	2.83×10^{-4}	3.00×10^{-4}
VAR($\hat{\beta}_1$)	0.0315	0.0317	0.0258	0.0272	0.0249
MSE($\hat{\beta}_2$)	8.08×10^{-9}	7.31×10^{-9}	6.37×10^{-9}	5.34×10^{-9}	5.47×10^{-9}
BIAS($\hat{\beta}_2$)	2.01×10^{-10}	2.72×10^{-11}	6.02×10^{-11}	1.66×10^{-11}	7.37×10^{-11}
VAR($\hat{\beta}_2$)	7.88×10^{-9}	7.28×10^{-9}	6.31×10^{-9}	5.32×10^{-9}	5.39×10^{-9}
SRC	0.169	0.155	0.089	0.301	
Skewness	0.345	0.421	0.301	0.364	0.130

Table 4.4.1.4: Simulation results for the Hypothesis 1: $n = 41$, $\sigma^2 = 0.01$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0020	0.0021	0.0016	0.0019	0.0019
BIAS($\hat{\beta}_1$)	7.39×10^{-6}	6.03×10^{-5}	5.31×10^{-5}	7.18×10^{-5}	3.87×10^{-5}
VAR($\hat{\beta}_1$)	0.0020	0.0021	0.0015	0.0019	0.0018
MSE($\hat{\beta}_2$)	5.57×10^{-10}	5.64×10^{-10}	4.00×10^{-10}	5.40×10^{-10}	4.70×10^{-10}
BIAS($\hat{\beta}_2$)	9.72×10^{-12}	1.77×10^{-11}	1.45×10^{-11}	7.84×10^{-12}	7.95×10^{-12}
VAR($\hat{\beta}_2$)	5.47×10^{-10}	5.46×10^{-10}	3.85×10^{-10}	5.32×10^{-10}	4.62×10^{-10}
SRC	0.189	0.156	0.087	0.114	0.298
Skewness	0.261	0.284	0.279	0.289	0.128

Table 4.4.1.5: Simulation results for the Hypothesis 1: $n = 41$, $\sigma^2 = 0.02$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0090	0.0076	0.0076	0.0078	0.0076
BIAS($\hat{\beta}_1$)	9.79×10^{-5}	6.64×10^{-6}	8.14×10^{-5}	4.11×10^{-5}	3.29×10^{-5}
VAR($\hat{\beta}_1$)	0.0089	0.0076	0.0075	0.0077	0.0075
MSE($\hat{\beta}_2$)	2.30×10^{-9}	1.75×10^{-9}	1.92×10^{-9}	1.82×10^{-9}	1.51×10^{-9}
BIAS($\hat{\beta}_2$)	1.64×10^{-11}	1.57×10^{-11}	6.04×10^{-12}	1.91×10^{-11}	3.83×10^{-12}
VAR($\hat{\beta}_2$)	2.28×10^{-9}	1.74×10^{-9}	1.91×10^{-9}	1.80×10^{-9}	1.50×10^{-9}
SRC	0.155	0.136	0.088	0.098	0.167
Skewness	0.339	0.295	0.297	0.342	0.130

Table 4.4.1.6: Simulation results for the Hypothesis 1: $n = 41$, $\sigma^2 = 0.03$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0225	0.0216	0.0192	0.0164	0.0112
BIAS($\hat{\beta}_1$)	2.34×10^{-3}	6.96×10^{-7}	1.20×10^{-4}	8.31×10^{-4}	1.31×10^{-4}
VAR($\hat{\beta}_1$)	0.0201	0.0216	0.0191	0.0156	0.0111
MSE($\hat{\beta}_2$)	5.22×10^{-9}	4.97×10^{-9}	5.18×10^{-9}	4.01×10^{-9}	3.70×10^{-9}
BIAS($\hat{\beta}_2$)	3.62×10^{-10}	2.05×10^{-13}	3.18×10^{-13}	1.65×10^{-10}	1.21×10^{-13}
VAR($\hat{\beta}_2$)	4.86×10^{-9}	4.97×10^{-9}	5.18×10^{-9}	3.84×10^{-9}	3.69×10^{-9}
SRC	0.158	0.142	0.079	0.092	0.147
Skewness	0.317	0.304	0.324	0.292	0.123

Table 4.4.1.7: Simulation results for the Hypothesis 1: $n = 63$, $\sigma^2 = 0.01$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0016	0.0013	0.0009	0.0013	0.0012
BIAS($\hat{\beta}_1$)	1.96×10^{-5}	4.69×10^{-10}	1.30×10^{-5}	3.72×10^{-5}	1.14×10^{-7}
VAR($\hat{\beta}_1$)	0.0016	0.0013	0.0009	0.0012	0.0012
MSE($\hat{\beta}_2$)	3.36×10^{-10}	3.34×10^{-10}	2.57×10^{-10}	3.19×10^{-10}	2.88×10^{-10}
BIAS($\hat{\beta}_2$)	7.02×10^{-12}	2.93×10^{-13}	2.95×10^{-12}	1.08×10^{-11}	9.00×10^{-14}
VAR($\hat{\beta}_2$)	3.29×10^{-10}	3.34×10^{-10}	2.54×10^{-10}	3.08×10^{-10}	2.88×10^{-10}
SRC	0.146	0.154	0.065	0.100	0.242
Skewness	0.228	0.247	0.243	0.239	0.097

Table 4.4.1.8: Simulation results for the Hypothesis 1: $n = 63$, $\sigma^2 = 0.02$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0059	0.0046	0.0052	0.0054	0.0053
BIAS($\hat{\beta}_1$)	8.32×10^{-6}	3.30×10^{-6}	1.96×10^{-5}	5.86×10^{-4}	3.35×10^{-5}
VAR($\hat{\beta}_1$)	0.0059	0.0046	0.0052	0.0048	0.0052
MSE($\hat{\beta}_2$)	1.88×10^{-9}	1.15×10^{-9}	1.32×10^{-9}	1.27×10^{-9}	1.26×10^{-9}
BIAS($\hat{\beta}_2$)	2.69×10^{-11}	6.28×10^{-12}	7.74×10^{-12}	7.69×10^{-11}	1.05×10^{-12}
VAR($\hat{\beta}_2$)	1.85×10^{-9}	1.14×10^{-9}	1.31×10^{-9}	1.20×10^{-9}	1.25×10^{-9}
SRC	0.141	0.136	0.066	0.070	0.227
Skewness	0.256	0.233	0.210	0.227	0.112

Table 4.4.1.9: Simulation results for the Hypothesis 1: $n = 63$, $\sigma^2 = 0.03$

Approach	OLS	PTB	PWT	PTBWLS	TRUE
MSE($\hat{\beta}_1$)	0.0208	0.0136	0.0118	0.0178	0.0097
BIAS($\hat{\beta}_1$)	4.60×10^{-4}	5.41×10^{-4}	1.05×10^{-4}	5.40×10^{-4}	2.35×10^{-4}
VAR($\hat{\beta}_1$)	0.0203	0.0131	0.0118	0.0173	0.0097
MSE($\hat{\beta}_2$)	3.46×10^{-9}	3.07×10^{-9}	3.00×10^{-9}	3.08×10^{-9}	2.31×10^{-9}
BIAS($\hat{\beta}_2$)	1.73×10^{-12}	8.33×10^{-11}	2.85×10^{-14}	2.32×10^{-11}	6.73×10^{-11}
VAR($\hat{\beta}_2$)	3.46×10^{-9}	2.99×10^{-9}	3.00×10^{-9}	3.05×10^{-9}	2.24×10^{-9}
SRC	0.149	0.127	0.063	0.079	0.230
Skewness	0.249	0.259	0.234	0.242	0.092

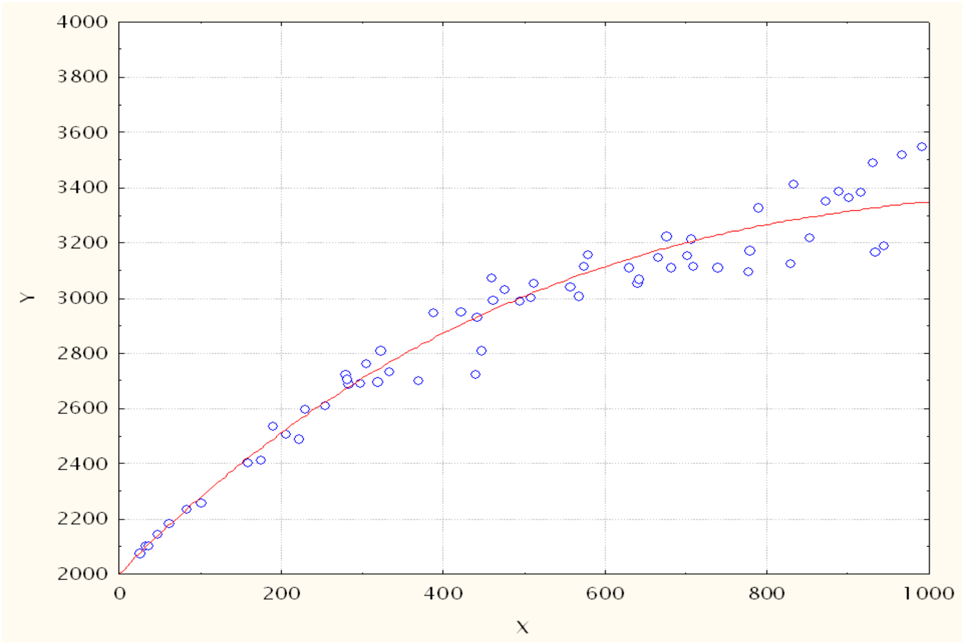

Figure 4.4.2.1: Example of simulation data ($\sigma^2 = 0.01, n = 63$)

4.4.2 Result and Interpretation from Simulation of Hypothesis 2

Table 4.4.2.1 to Table 4.4.2.3 show the results of the simulation for Hypothesis 2 in case that sample size was $n = 29$ and for $\sigma^2 = 0.01$, $\sigma^2 = 0.02$ and $\sigma^2 = 0.03$, respectively. From the results in case that the sample size is small (Table 4.4.1.1 to Table 4.4.1.3), the mean square errors were basically improved in order of NTBWLS, PTBWLS, PWT, PTB, OLS. In particular, the difference between OLS, PTB and PWT, PTBWLS, NTBWLS were remarkable. SRC improved same as the results of MSE. It can be thought that the power-weighted transformation approach improved better than the ordinary least squares approach and the transformation-both-sides approach in the situation of the non-normal and heteroscedastic error distributions. Also, in the situation of the larger variance (Table 4.4.2.3), the improvement of NTBWLS for the MSE was remarkable. For the situation of the larger sample size (Table 4.4.2.4 to Table 4.4.2.9), it seems that we can give the same interpretations as Table 4.4.2.1 to Table 4.4.2.3.

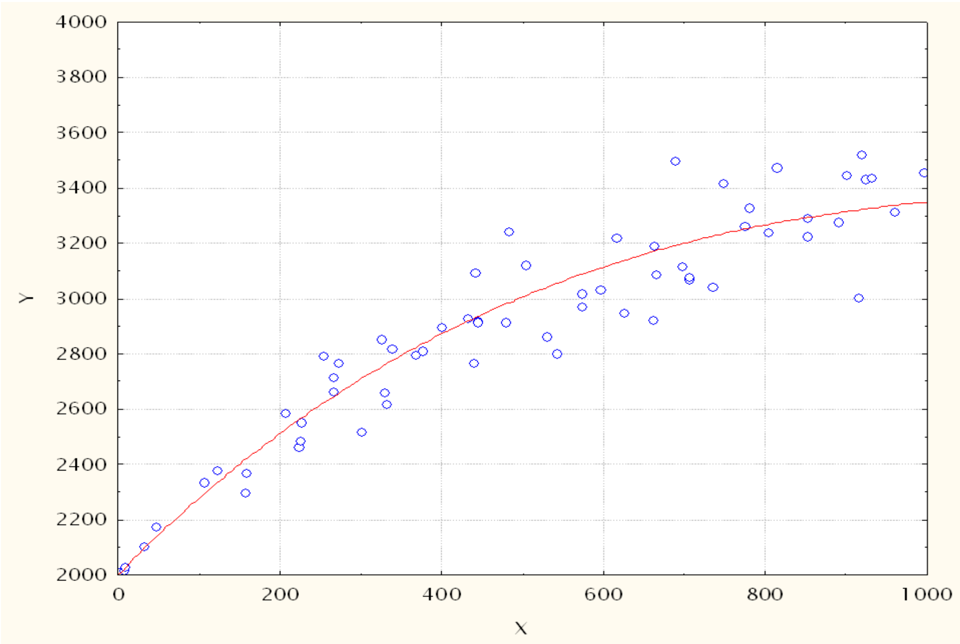

Figure 4.4.2.2: Example of simulation data ($\sigma^2 = 0.02, n = 63$)

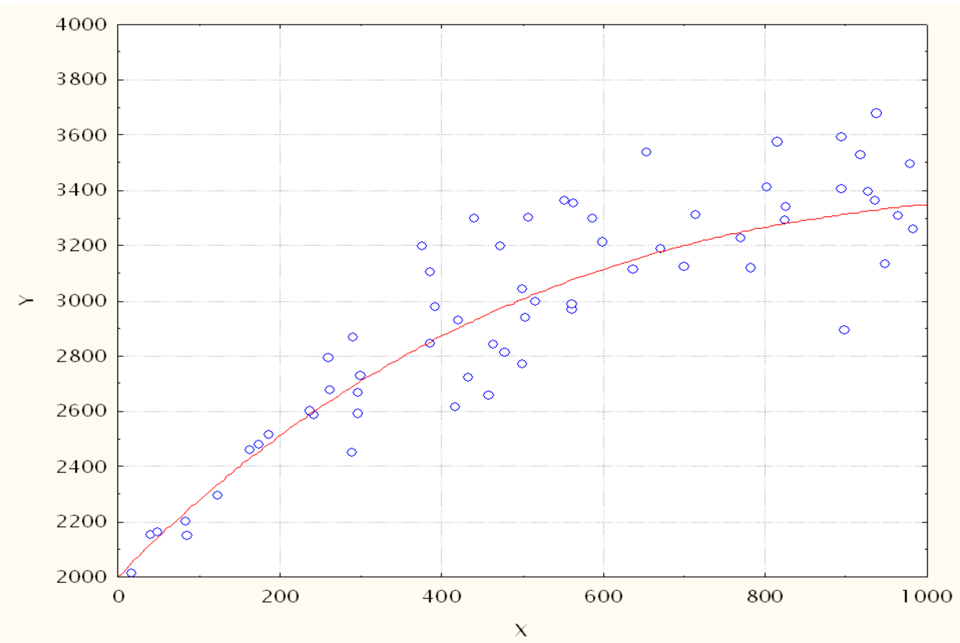

Figure 4.4.2.3: Example of simulation data ($\sigma^2 = 0.03, n = 63$)

Table 4.4.2.1: Simulation results for the Hypothesis 2: $n = 29$, $\sigma^2 = 0.01$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0156	0.0112	0.0086	0.0082	0.0067
BIAS($\hat{\beta}_1$)	7.52×10^{-5}	9.94×10^{-5}	1.22×10^{-4}	7.77×10^{-5}	3.56×10^{-5}
VAR($\hat{\beta}_1$)	0.0155	0.0111	0.0084	0.0081	0.0067
MSE($\hat{\beta}_2$)	4.59×10^{-9}	3.13×10^{-9}	2.99×10^{-9}	2.92×10^{-9}	2.18×10^{-9}
BIAS($\hat{\beta}_2$)	7.19×10^{-12}	1.15×10^{-11}	4.70×10^{-12}	3.42×10^{-13}	3.78×10^{-14}
VAR($\hat{\beta}_2$)	4.59×10^{-9}	3.12×10^{-9}	2.99×10^{-9}	2.92×10^{-9}	2.18×10^{-9}
SRC	0.460	0.419	0.087	0.107	–
Skewness	0.393	0.412	0.322	0.326	–

Table 4.4.2.2: Simulation results for the Hypothesis 2: $n = 29$, $\sigma^2 = 0.02$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0440	0.0631	0.0439	0.0281	0.0345
BIAS($\hat{\beta}_1$)	5.12×10^{-4}	6.13×10^{-4}	4.34×10^{-4}	2.08×10^{-3}	1.46×10^{-3}
VAR($\hat{\beta}_1$)	0.0435	0.0625	0.0437	0.0279	0.0331
MSE($\hat{\beta}_2$)	2.23×10^{-8}	1.89×10^{-8}	1.33×10^{-8}	7.61×10^{-9}	1.41×10^{-8}
BIAS($\hat{\beta}_2$)	2.28×10^{-11}	4.48×10^{-12}	3.95×10^{-11}	3.39×10^{-10}	4.39×10^{-10}
VAR($\hat{\beta}_2$)	1.23×10^{-8}	1.89×10^{-8}	1.33×10^{-8}	7.27×10^{-9}	1.37×10^{-8}
SRC	0.432	0.392	0.085	0.105	–
Skewness	0.407	0.395	0.290	0.301	–

4.4.3 Result and Interpretation from Simulation of Hypothesis 3

Table 4.4.3.1 to Table 4.4.3.3 show the results of the simulation for Hypothesis 2 in case that sample size was $n = 29$ and for $\sigma^2 = 0.01$, $\sigma^2 = 0.02$ and $\sigma^2 = 0.03$, respectively. Figure 4.4.3.1 and Figure 4.4.3.4 show the results of MSE($\hat{\beta}_1$) and MSE($\hat{\beta}_2$) for the Hypothesis 3 in case of $n = 29$. In Figure 4.4.3.1 and Figure 4.4.3.4, MSE was clearly improved in order of NTBWLS, PTBWLS, PWT, OLS in the situation of $\sigma^2 = 0.03$. There were no difference between each approach in the situation of $\sigma^2 = 0.01$. In contrast to above results, in case that the sample size was large (Figure 4.4.3.2, Figure 4.4.3.3, Figure 4.4.3.5 and Figure 4.4.3.6), PWT, PTBWLS and NTBWLS were clearly improved better than OLS for MSE, but there were not so much difference between each transformation approach. Therefore, it can be thought that NTBWLS is superior to the other approaches for MSE in the situation of small sample size and larger variance with non-normal and heteroscedastic error. In Table 4.4.3.1 to 4.4.3.9, it seems that we can provide

Table 4.4.2.3: Simulation results for the Hypothesis 2: $n = 29$, $\sigma^2 = 0.03$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.165	0.146	0.085	0.103	0.049
BIAS($\hat{\beta}_1$)	5.48×10^{-4}	4.95×10^{-4}	2.08×10^{-3}	4.45×10^{-3}	5.60×10^{-3}
VAR($\hat{\beta}_1$)	0.1639	0.1458	0.0831	0.0990	0.0437
MSE($\hat{\beta}_2$)	4.32×10^{-8}	3.89×10^{-8}	2.22×10^{-8}	3.54×10^{-8}	2.14×10^{-8}
BIAS($\hat{\beta}_2$)	1.53×10^{-10}	1.48×10^{-11}	1.36×10^{-11}	3.48×10^{-11}	9.80×10^{-10}
VAR($\hat{\beta}_2$)	4.30×10^{-8}	3.87×10^{-8}	2.22×10^{-8}	3.54×10^{-8}	2.04×10^{-8}
SRC	0.402	0.292	0.089	0.100	–
Skewness	0.442	0.395	0.326	0.302	–

Table 4.4.2.4: Simulation results for the Hypothesis 2: $n = 41$, $\sigma^2 = 0.01$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0112	0.0097	0.0064	0.0069	0.0068
BIAS($\hat{\beta}_1$)	3.86×10^{-4}	2.18×10^{-8}	4.93×10^{-6}	3.33×10^{-7}	2.82×10^{-4}
VAR($\hat{\beta}_1$)	0.0109	0.0097	0.0064	0.0069	0.0066
MSE($\hat{\beta}_2$)	3.02×10^{-9}	2.74×10^{-9}	2.38×10^{-9}	1.89×10^{-9}	2.33×10^{-9}
BIAS($\hat{\beta}_2$)	7.70×10^{-11}	6.08×10^{-13}	1.54×10^{-14}	6.50×10^{-12}	8.12×10^{-11}
VAR($\hat{\beta}_2$)	2.94×10^{-9}	2.74×10^{-9}	2.38×10^{-9}	1.89×10^{-9}	2.25×10^{-9}
SRC	0.446	0.426	0.081	0.089	–
Skewness	0.341	0.329	0.296	0.268	–

Table 4.4.2.5: Simulation results for the Hypothesis 2: $n = 41$, $\sigma^2 = 0.02$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0333	0.0297	0.0271	0.0236	0.0230
BIAS($\hat{\beta}_1$)	1.01×10^{-4}	2.82×10^{-4}	1.44×10^{-5}	1.76×10^{-5}	3.37×10^{-4}
VAR($\hat{\beta}_1$)	0.0332	0.0296	0.0271	0.0235	0.0226
MSE($\hat{\beta}_2$)	1.04×10^{-8}	0.96×10^{-8}	9.27×10^{-9}	7.42×10^{-9}	7.40×10^{-9}
BIAS($\hat{\beta}_2$)	1.17×10^{-10}	2.56×10^{-11}	1.17×10^{-10}	2.92×10^{-11}	6.40×10^{-11}
VAR($\hat{\beta}_2$)	1.03×10^{-8}	0.96×10^{-8}	9.27×10^{-9}	7.42×10^{-9}	7.33×10^{-9}
SRC	0.449	0.428	0.078	0.083	–
Skewness	0.397	0.342	0.303	0.252	–

Table 4.4.2.6: Simulation results for the Hypothesis 2: $n = 41$, $\sigma^2 = 0.03$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0865	0.0797	0.0565	0.0499	0.0522
BIAS($\hat{\beta}_1$)	6.02×10^{-4}	3.98×10^{-4}	4.01×10^{-4}	7.02×10^{-4}	4.35×10^{-3}
VAR($\hat{\beta}_1$)	0.0859	0.0797	0.0565	0.0499	0.0478
MSE($\hat{\beta}_2$)	2.48×10^{-8}	2.06×10^{-8}	1.99×10^{-8}	1.54×10^{-8}	1.66×10^{-8}
BIAS($\hat{\beta}_2$)	2.24×10^{-11}	2.56×10^{-11}	1.17×10^{-10}	2.92×10^{-11}	2.93×10^{-10}
VAR($\hat{\beta}_2$)	2.48×10^{-8}	2.06×10^{-8}	1.99×10^{-8}	1.53×10^{-8}	1.63×10^{-8}
SRC	0.457	0.232	0.076	0.092	–
Skewness	0.329	0.191	0.272	0.316	–

Table 4.4.2.7: Simulation results for the Hypothesis 2: $n = 63$, $\sigma^2 = 0.01$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.00526	0.00653	0.00348	0.00343	0.00366
BIAS($\hat{\beta}_1$)	4.62×10^{-5}	1.92×10^{-4}	1.88×10^{-5}	4.21×10^{-5}	2.70×10^{-5}
VAR($\hat{\beta}_1$)	0.00522	0.00634	0.00346	0.00339	0.0036
MSE($\hat{\beta}_2$)	1.52×10^{-9}	1.81×10^{-9}	1.09×10^{-9}	1.20×10^{-9}	1.39×10^{-9}
BIAS($\hat{\beta}_2$)	4.87×10^{-12}	8.17×10^{-11}	2.86×10^{-14}	4.04×10^{-12}	7.10×10^{-13}
VAR($\hat{\beta}_2$)	1.51×10^{-9}	1.89×10^{-9}	1.09×10^{-9}	1.20×10^{-9}	1.39×10^{-9}
SRC	0.453	0.385	0.071	0.075	–
Skewness	0.337	0.301	0.215	0.254	–

Table 4.4.2.8: Simulation results for the Hypothesis 2: $n = 63$, $\sigma^2 = 0.02$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0277	0.0201	0.0161	0.0185	0.0121
BIAS($\hat{\beta}_1$)	7.15×10^{-5}	5.92×10^{-5}	7.41×10^{-5}	6.21×10^{-4}	5.94×10^{-4}
VAR($\hat{\beta}_1$)	0.0276	0.0201	0.0161	0.0179	0.0115
MSE($\hat{\beta}_2$)	7.77×10^{-9}	7.78×10^{-9}	5.04×10^{-9}	4.83×10^{-9}	4.31×10^{-9}
BIAS($\hat{\beta}_2$)	8.82×10^{-11}	3.50×10^{-11}	8.33×10^{-11}	1.28×10^{-11}	3.68×10^{-11}
VAR($\hat{\beta}_2$)	7.68×10^{-9}	7.78×10^{-9}	4.96×10^{-9}	4.82×10^{-9}	4.27×10^{-9}
SRC	0.449	0.384	0.071	0.085	–
Skewness	0.364	0.337	0.252	0.269	–

Table 4.4.2.9: Simulation results for the Hypothesis 2: $n = 63$, $\sigma^2 = 0.03$

Approach	OLS	PTB	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0670	0.0648	0.0313	0.0275	0.0318
BIAS($\hat{\beta}_1$)	5.19×10^{-5}	4.92×10^{-5}	3.82×10^{-4}	3.48×10^{-4}	3.34×10^{-3}
VAR($\hat{\beta}_1$)	0.0669	0.0648	0.0309	0.0271	0.0285
MSE($\hat{\beta}_2$)	2.12×10^{-8}	1.78×10^{-8}	1.00×10^{-8}	6.76×10^{-9}	9.67×10^{-9}
BIAS($\hat{\beta}_2$)	4.00×10^{-11}	3.44×10^{-11}	3.56×10^{-11}	2.43×10^{-11}	3.16×10^{-10}
VAR($\hat{\beta}_2$)	2.12×10^{-8}	1.78×10^{-8}	9.99×10^{-9}	6.74×10^{-9}	9.36×10^{-9}
SRC	0.465	0.444	0.073	0.063	–
Skewness	0.354	0.321	0.257	0.259	–

the similar interpretations as the results of the simulation for Hypothesis 2. However, we should focus on the results of skewness. There were trend toward that the skewness were decreased in order of NTBWLS, PTBWLS, PWT, OLS. We can consider that the simulation model for Hypothesis 3 had the intentional skewness for the error distribution and NTBWLS improved the skewness of the error as compare to OLS and PWT.

Table 4.4.3.1: Simulation results for the Hypothesis 3: $n = 29$, $\sigma^2 = 0.01$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0125	0.0056	0.0050	0.0039
BIAS($\hat{\beta}_1$)	1.11×10^{-5}	2.39×10^{-4}	4.37×10^{-5}	5.56×10^{-5}
VAR($\hat{\beta}_1$)	0.0124	0.0054	0.0050	0.0039
MSE($\hat{\beta}_2$)	3.97×10^{-9}	2.17×10^{-9}	2.27×10^{-9}	1.71×10^{-9}
BIAS($\hat{\beta}_2$)	1.79×10^{-11}	1.94×10^{-11}	3.97×10^{-12}	1.84×10^{-11}
VAR($\hat{\beta}_2$)	3.95×10^{-9}	2.17×10^{-9}	2.27×10^{-9}	1.70×10^{-9}
SRC	0.489	0.082	0.081	–
Skewness	0.447	0.293	0.263	–

Table 4.4.3.2: Simulation results for the Hypothesis 3: $n = 41$, $\sigma^2 = 0.01$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0105	0.0039	0.0042	0.0032
BIAS($\hat{\beta}_1$)	9.37×10^{-5}	1.16×10^{-4}	6.69×10^{-5}	2.80×10^{-6}
VAR($\hat{\beta}_1$)	0.0104	0.0038	0.0041	0.0032
MSE($\hat{\beta}_2$)	3.36×10^{-9}	1.49×10^{-9}	1.67×10^{-9}	1.51×10^{-9}
BIAS($\hat{\beta}_2$)	5.60×10^{-11}	2.27×10^{-14}	2.91×10^{-11}	8.74×10^{-12}
VAR($\hat{\beta}_2$)	3.30×10^{-9}	1.49×10^{-9}	1.67×10^{-9}	1.51×10^{-9}
SRC	0.510	0.077	0.082	–
Skewness	0.388	0.251	0.282	–

4.5 Discussions

In this section, we discuss the simulation results for three Hypotheses established in Section 4.1.

From the results of the simulation for Hypothesis 1, MSE($\hat{\boldsymbol{\beta}}$) of PTB, PWT and PTBWLS were smaller than that of OLS for each situation, therefore it was confirmed that these parametric transformation approaches improved performance of model parameters estimation. In particular, it seemed that the smaller the sample size, the larger the improvements. In the same way, it seemed that the larger the error variance, the larger the improvements. From the results of the error distribution at post-transformation, there were no difference for the skewness between the approaches but SRCs were improved in PWT and PTBWLS. it could be thought that the power-weighted transformation improved heteroscedasticity of the error. In view of these results, it was showed that the transformation approaches were superior to usual least square approach in the perfor-

Table 4.4.3.3: Simulation results for the Hypothesis 3: $n = 63$, $\sigma^2 = 0.01$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0065	0.0027	0.0031	0.0026
BIAS($\hat{\beta}_1$)	4.70×10^{-5}	1.07×10^{-6}	4.32×10^{-5}	2.70×10^{-5}
VAR($\hat{\beta}_1$)	0.0064	0.0027	0.0030	0.0026
MSE($\hat{\beta}_2$)	1.80×10^{-9}	1.10×10^{-9}	1.25×10^{-9}	1.04×10^{-9}
BIAS($\hat{\beta}_2$)	7.27×10^{-11}	4.99×10^{-12}	2.87×10^{-11}	1.31×10^{-12}
VAR($\hat{\beta}_2$)	1.73×10^{-9}	1.10×10^{-9}	1.25×10^{-9}	1.04×10^{-9}
SRC	0.521	0.075	0.064	–
Skewness	0.438	0.190	0.217	–

Table 4.4.3.4: Simulation results for the Hypothesis 3: $n = 29$, $\sigma^2 = 0.02$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0524	0.0206	0.0159	0.0135
BIAS($\hat{\beta}_1$)	4.26×10^{-4}	2.73×10^{-5}	2.18×10^{-4}	6.35×10^{-6}
VAR($\hat{\beta}_1$)	0.0520	0.0206	0.0158	0.0135
MSE($\hat{\beta}_2$)	1.50×10^{-8}	9.56×10^{-9}	7.66×10^{-9}	7.07×10^{-9}
BIAS($\hat{\beta}_2$)	2.62×10^{-12}	1.81×10^{-11}	1.00×10^{-11}	2.93×10^{-11}
VAR($\hat{\beta}_2$)	1.50×10^{-8}	9.54×10^{-9}	7.65×10^{-9}	7.04×10^{-9}
SRC	0.474	0.099	0.084	–
Skewness	0.436	0.327	0.299	–

Table 4.4.3.5: Simulation results for the Hypothesis 3: $n = 41$, $\sigma^2 = 0.02$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0413	0.0179	0.0138	0.0124
BIAS($\hat{\beta}_1$)	1.27×10^{-4}	5.66×10^{-6}	2.27×10^{-7}	8.53×10^{-5}
VAR($\hat{\beta}_1$)	0.0411	0.0179	0.0138	0.0123
MSE($\hat{\beta}_2$)	1.32×10^{-8}	6.40×10^{-9}	5.50×10^{-9}	5.38×10^{-9}
BIAS($\hat{\beta}_2$)	8.61×10^{-11}	1.78×10^{-10}	1.95×10^{-11}	1.51×10^{-11}
VAR($\hat{\beta}_2$)	1.31×10^{-8}	6.22×10^{-9}	5.48×10^{-9}	5.36×10^{-9}
SRC	0.522	0.081	0.084	–
Skewness	0.449	0.276	0.275	–

Table 4.4.3.6: Simulation results for the Hypothesis 3: $n = 63$, $\sigma^2 = 0.02$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0251	0.0089	0.0118	0.0077
BIAS($\hat{\beta}_1$)	6.37×10^{-4}	3.85×10^{-5}	2.11×10^{-4}	1.53×10^{-4}
VAR($\hat{\beta}_1$)	0.0244	0.0088	0.0116	0.0076
MSE($\hat{\beta}_2$)	8.09×10^{-9}	3.52×10^{-9}	4.74×10^{-9}	3.80×10^{-9}
BIAS($\hat{\beta}_2$)	4.68×10^{-10}	2.81×10^{-13}	2.02×10^{-13}	2.48×10^{-14}
VAR($\hat{\beta}_2$)	7.63×10^{-9}	3.52×10^{-9}	4.74×10^{-9}	3.80×10^{-9}
SRC	0.529	0.068	0.071	–
Skewness	0.417	0.187	0.222	–

Table 4.4.3.7: Simulation results for the Hypothesis 3: $n = 29$, $\sigma^2 = 0.03$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.1281	0.0516	0.0420	0.0225
BIAS($\hat{\beta}_1$)	2.83×10^{-3}	7.93×10^{-5}	4.41×10^{-4}	6.35×10^{-6}
VAR($\hat{\beta}_1$)	0.1252	0.0515	0.0416	0.0225
MSE($\hat{\beta}_2$)	4.74×10^{-8}	1.98×10^{-8}	1.95×10^{-8}	1.68×10^{-8}
BIAS($\hat{\beta}_2$)	2.10×10^{-9}	4.55×10^{-10}	6.43×10^{-14}	6.95×10^{-12}
VAR($\hat{\beta}_2$)	4.53×10^{-8}	1.93×10^{-8}	1.95×10^{-8}	1.68×10^{-8}
SRC	0.492	0.095	0.103	–
Skewness	0.416	0.330	0.244	–

Table 4.4.3.8: Simulation results for the Hypothesis 3: $n = 41$, $\sigma^2 = 0.03$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0802	0.0343	0.0297	0.0236
BIAS($\hat{\beta}_1$)	2.10×10^{-4}	2.71×10^{-4}	5.50×10^{-5}	4.44×10^{-4}
VAR($\hat{\beta}_1$)	0.0800	0.0342	0.0297	0.0232
MSE($\hat{\beta}_2$)	2.47×10^{-8}	1.47×10^{-8}	1.43×10^{-8}	1.20×10^{-8}
BIAS($\hat{\beta}_2$)	4.78×10^{-10}	2.59×10^{-10}	3.43×10^{-11}	1.79×10^{-10}
VAR($\hat{\beta}_2$)	2.43×10^{-8}	1.44×10^{-8}	1.43×10^{-8}	1.19×10^{-8}
SRC	0.518	0.083	0.083	–
Skewness	0.469	0.310	0.299	–

Table 4.4.3.9: Simulation results for the Hypothesis 3: $n = 63$, $\sigma^2 = 0.03$

Approach	OLS	PWT	PTBWLS	NTBWLS
MSE($\hat{\beta}_1$)	0.0573	0.0232	0.0230	0.0234
BIAS($\hat{\beta}_1$)	1.44×10^{-4}	2.04×10^{-6}	1.21×10^{-6}	7.68×10^{-4}
VAR($\hat{\beta}_1$)	0.0572	0.0232	0.0230	0.0226
MSE($\hat{\beta}_2$)	1.87×10^{-8}	8.75×10^{-9}	9.44×10^{-9}	8.75×10^{-9}
BIAS($\hat{\beta}_2$)	4.02×10^{-10}	6.33×10^{-10}	1.29×10^{-10}	2.41×10^{-11}
VAR($\hat{\beta}_2$)	1.83×10^{-8}	8.12×10^{-9}	9.31×10^{-9}	8.73×10^{-9}
SRC	0.518	0.072	0.063	–
Skewness	0.489	0.082	0.081	–

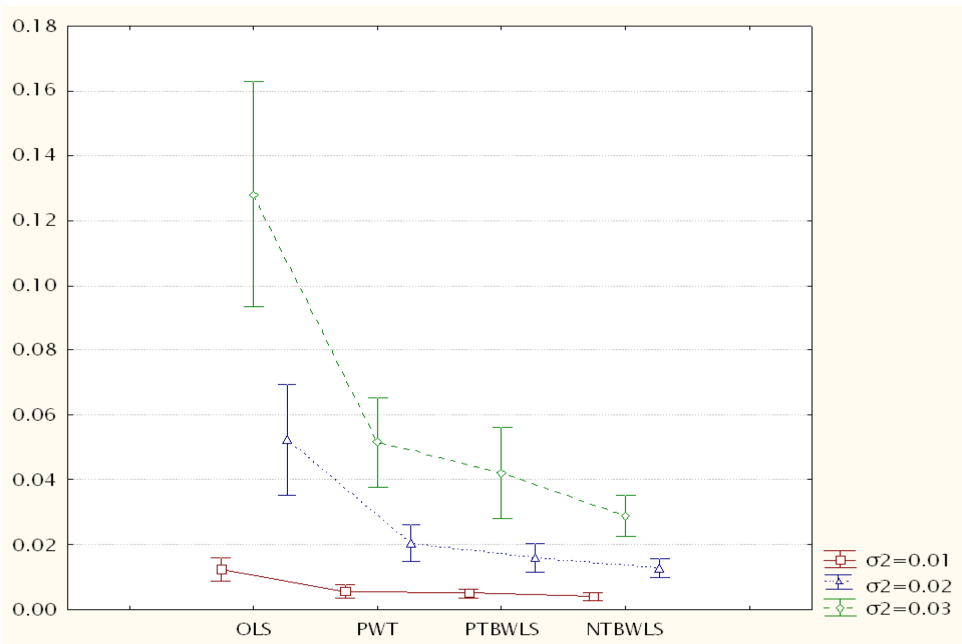

Figure 4.4.3.1: Results of MSE($\hat{\beta}_1$) for the Hypothesis 3 ($n = 29$)

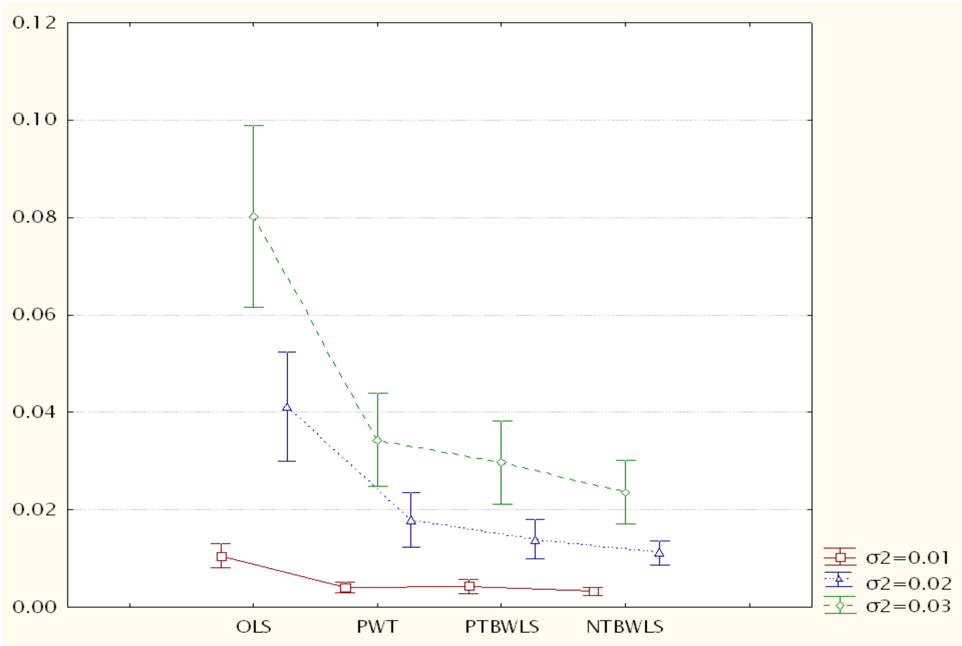

Figure 4.4.3.2: Results of MSE($\hat{\beta}_1$) for the Hypothesis 3 ($n = 41$)

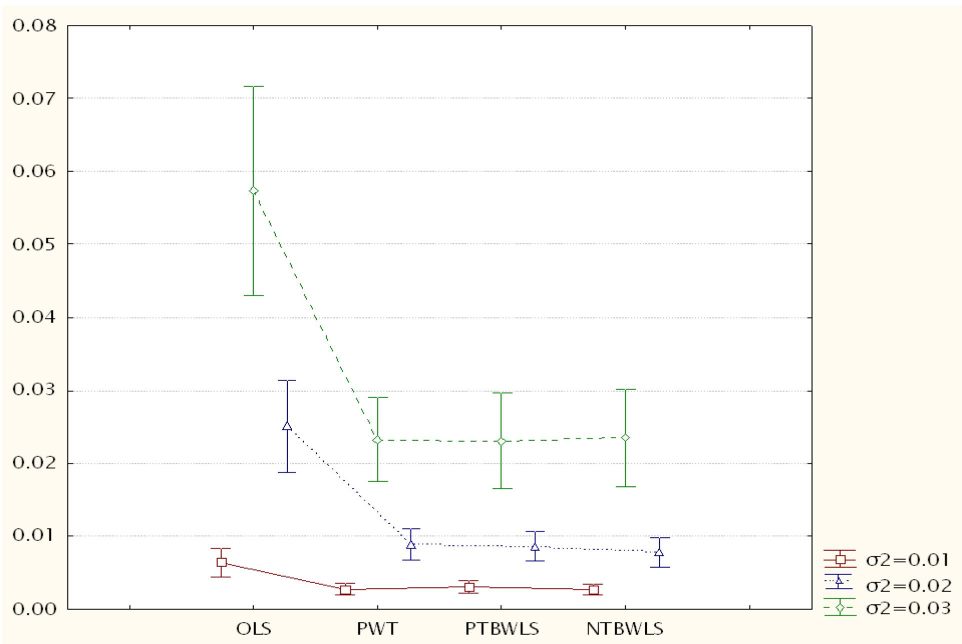

Figure 4.4.3.3: Results of MSE($\hat{\beta}_1$) for the Hypothesis 3 ($n = 63$)

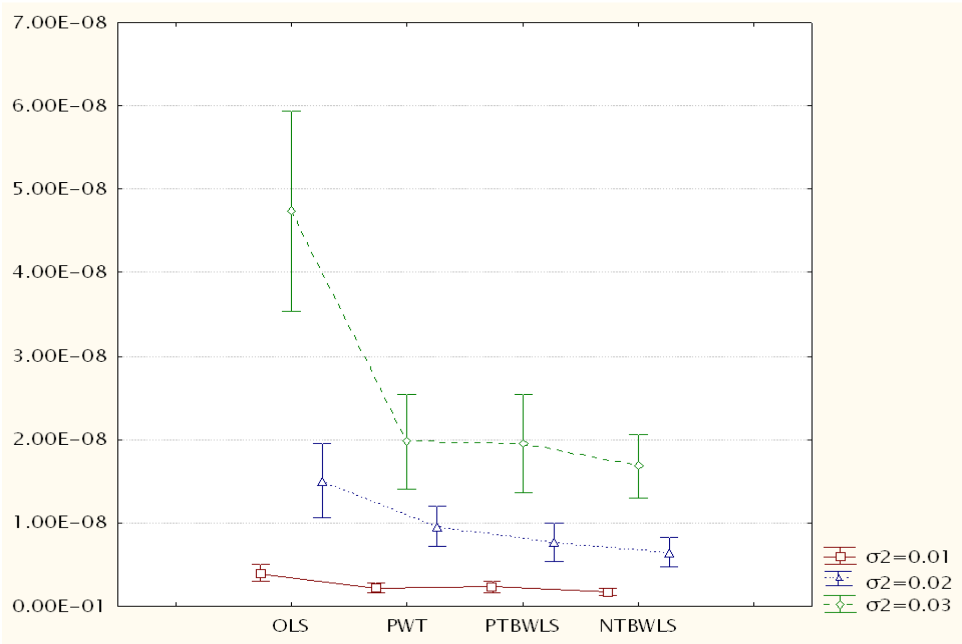

Figure 4.4.3.4: Results of MSE($\hat{\beta}_2$) for the Hypothesis 3 ($n = 29$)

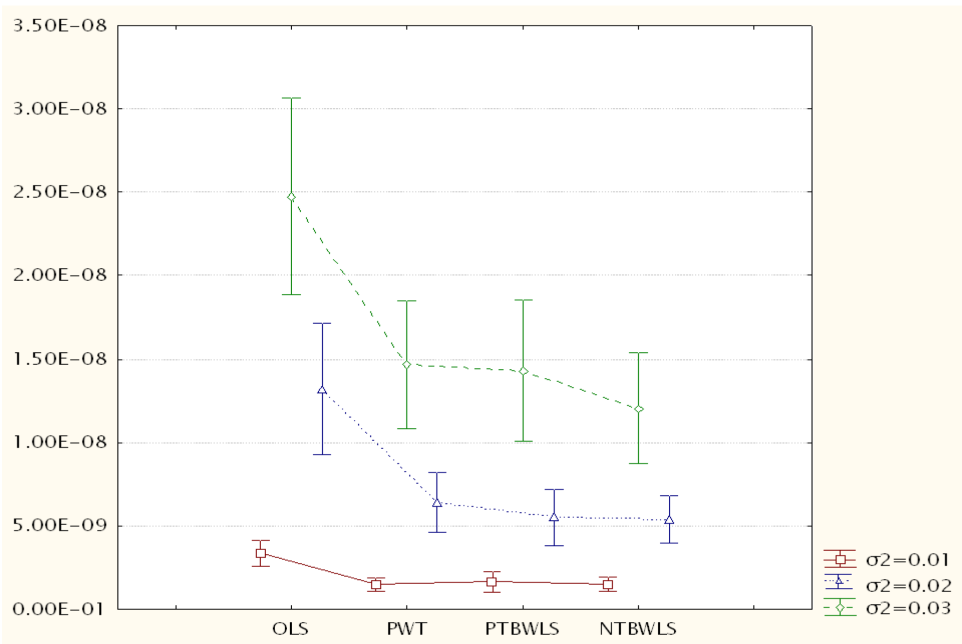

Figure 4.4.3.5: Results of MSE($\hat{\beta}_2$) for the Hypothesis 3 ($n = 41$)

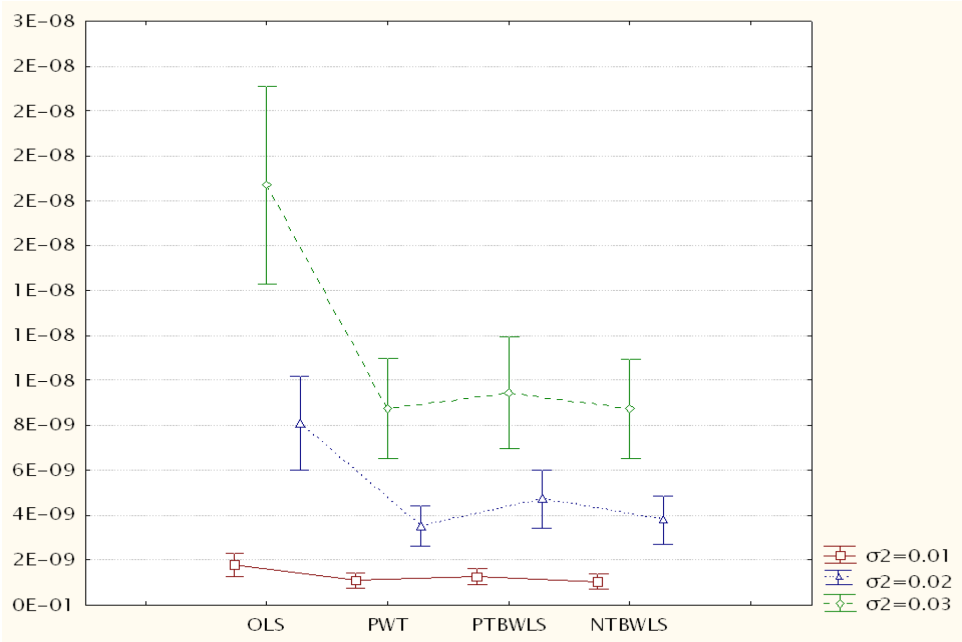

Figure 4.4.3.6: Results of MSE($\hat{\beta}_2$) for the Hypothesis 3 ($n = 63$)

mance for the model-parameter estimation in the case the error for a true non-linear model was distributed non-normal with constant variance.

From the results of the simulation for Hypothesis 2, MSE($\hat{\boldsymbol{\beta}}$) of PTB, PWT, PTBWLS and NTBWLS were smaller than that of OLS for each situation. In addition, MSE($\hat{\boldsymbol{\beta}}$) of PTBWLS and NTBWLS were almost smaller than that of PTB and PWT. Also, the results of PWT were superior to that of PTB. From the results of the error distribution at post-transformation, the skewness and the SRCs were improved in PWT and PTBWLS. There was a difference from the results of Hypothesis 1, because there were no difference for the skewness between the approaches in the simulation of Hypothesis 1. it could be thought that the power transform-both-sides and the power-weighted transformation improved heteroscedasticity and non-normality of the error. For NTBWLS, in particular, MSE($\hat{\boldsymbol{\beta}}$) of NTBWLS was the smallest in the situation of small sample size and large variance. Interestingly, however, the BIAS($\hat{\boldsymbol{\beta}}$) of NTBWLS was larger than that of any other parametric approaches but the VAR($\hat{\boldsymbol{\beta}}$) of NTBWLS was smaller that of any other approaches. It could be thought that NTBWLS provided a kind of "Biased Estimator", therefore it could provide a good estimation efficiency for the model parameters. In view of these results, it was showed that PTBWLS and NTBWLS were superior to PTB and

PWT in the performance for the model-parameter estimation in the case the error for a true non-linear model was distributed non-normal with heteroscedasticity.

From the results of the simulation for Hypothesis 3, MSE($\hat{\boldsymbol{\beta}}$) of PWT, PTBWLS and NTBWLS were smaller than that of OLS for each sample size situation. For the setting of the error variance of the simulation model, the larger the error variance, the larger the improvements of PWT, PTBWLS and NTBWLS compared to that of OLS. In comparison between the transformation approaches, MSE($\hat{\boldsymbol{\beta}}$) were smaller in order of NTBWLS, PTBWLS and PWT, in particular, it was clear in the situation of large variance. From the results of the error distribution at post-transformation, the skewness and the SRCs were improved in PWT and PTBWLS. These improvements were markedly larger than the results of simulation of Hypothesis 1 and Hypothesis 2. In view of these results, NTBWLS improved the performance of estimation for model parameters more than PWT and PTBWLS in the case the true transformation was not included in a power function family and the error was distributed non-normal with heteroscedasticity.

5. Conclusions and Further Developments

In this paper, we attempted to examine some difficult points in the statistical inference on a theoretical model. Especially we focused the statistical error that shows the gap between data and the model, we introduced and suggested some transformation approaches to design the error of the theoretical model statistically. As a conventional parametric approach, we introduced the power transformation-both-sides approach (PTB) and power-weighted transformation approach (PWT). PTB has an objective to improve heteroscedasticity and normality, and PWT improves heteroscedasticity for the error. We could examine how minimizing the sum of squares in PTB. We confirmed that this minimizing was related to not only the heteroscedasticity of the error, but the skewness of the error distribution. For the sum of squares in PTB, the first term corresponded to the sum of squares in the PWT. In accordance with the result of those considerations, we suggested the power transform-both-sides and weighted least square approach (PTBWLS). PTBWLS was the extension of PTB model and involved two separate transformation parameters: one was the parameter to induce the normality of the error and the other was to estimate an appropriate weight for stabilizing the error variance. A Taylor expansion for the response around the predictive function in the second order gave us the outcome for the first term corresponds to the sum of squares in the PWT and the second term stands for the third moment corresponding to the skewness of the error distribution.

Since the choice of transformation was largely empirical it is important to consider the sensitivity of the model parameters to the power transformation function. One problem with using parametric transformation was the difficulty in extending the parametric transformation with only one parameter. Thus, it is not easy to assess the effect of more flexible transformations on the regression parameters or on prediction intervals in PTB.

Rather than to create more complicated transformations based on parametric expressions, we believed that it is more efficient to consider a nonparametric method of determining the transformation-both-sides function. Therefore, as an alternative to PTB, we proposed a Nonparametric Transform-Both-sides (NTB) approach to express function transformation as a cubic spline curve. Further, as an estimation method which combines PTBWLS with NTB together, we proposed a Nonparametric Transform-Both-sides and Weighted Least Squares (NTBWLS) approach. NTBWLS was designed to implement both nonparametric estimation of a transformation function and parametric estimation of a power weighted transformation function.

In the investigation result of the conical model, the transformation approaches much improved the heteroscedasticity and the non-normality of the error in a post-transformation compared with that of OLS. It suggested that these transformation approaches are effective to fit a simple non-linear model for heteroscedastic data. From the results of Ricker model and Beverton & Holt model, the transformation approaches much improved the heteroscedasticity and the non-normality of the error distribution in a post-transformation compared with that of OLS. Also, PWT, PTBWLS and NTBWLS improved them more than PTB. The approach in which the absolute value of SRC was the smallest was NTBWLS, but the absolute values of skewness of the error distribution in PWT and PTBWLS were smaller than that of NTBWLS. Furthermore, we conducted the outlier analysis by excepting an observation and investigate the robustness of each estimator for the model parameters. As a result, the difference of NTBWLS was smallest, so we were able to consider that NTBWLS gave the most robust estimates for the model parameters. Next, based on our case studies and numerical investigation of an example which include data generated from a 1-compartment model, we concluded that NTBWLS was superior to the other method in the situation of hardy heteroscedasticity and non-normality. NTBWLS provided a robust estimator to make the smoothing parameter smaller because it reduced the effect of the intensity of heteroscedasticity. In addition, we conducted the simulation experiments to confirm a superiority of NTBWLS to other approaches.

In the result, it was showed that 1) the transformation approaches were superior to usual least square approach in the performance for the model-parameter estimation in the case the error for a true non-linear model was distributed non-normal with constant variance, 2) PTBWLS and NTBWLS were superior to PTB and PWT in the performance for the model-parameter estimation in the case the error for a true non-linear model was distributed non-normal with heteroscedasticity, 3) NTBWLS improved the performance of estimation for model parameters more than PWT and PTBWLS in the case the true transformation was not included in a power function family and the error was distributed non-normal with heteroscedasticity.

The remaining problems for the future are 1) clarification of the roles played by transform-both-sides and weighted transformation, 2) development of "Double Nonparametric transformation", which implements nonparametric estimation for the weighted-transformation function, 3) to apply these transformation approaches to the empirical models.

Appendix: Consistency of parameter estimates for transformation both sides model

We briefly summarize the consistency of parameter estimates for transformation-both-sides model using the example of PTB by reference to Hernandez and Johnson (1980). As we discussed in section 2.1, based on the assumption of the error ε and ε_{P} are distributed as $\mathrm{N}(0,\sigma_n^2)$ and $\mathrm{N}(0,\sigma^2)$ respectively, log-likelihood equation is

$$\begin{aligned} L_{\mathrm{P}}(\boldsymbol{\beta},\sigma^2,\lambda) &= \sum_{n=1}^{N}\Big(-\frac{1}{2}[H_{\mathrm{P}}(y_n;\lambda) - H_{\mathrm{P}}\{f(\boldsymbol{x}_n;\boldsymbol{\beta}),\lambda\}]^2/\sigma^2 + \log\frac{d}{dt}H_{\mathrm{P}}(y_n;\lambda) \\ &\quad -\frac{1}{2}\log\sigma^2\Big) + C_0 \end{aligned} \tag{A.1}$$

for the observations $\{(\boldsymbol{x}_n, y_n),\ \ n = 1,2,...,N\}$, where C_0 is a constant including the coefficient of the probability density function. (A.1) is derived by supposing that there exists a value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0' = (\boldsymbol{\beta}_0,\sigma_0,\lambda_0)$ for which the distribution of $H_P(Y;\lambda_0)$ is normal with mean $H_P[f(\boldsymbol{X};\boldsymbol{\beta}_0),\lambda_0]$ and standard deviation σ_0. Except for the log-normal case, $H_P(Y;\lambda_0)$ cannot be normal for positive random variables. We now show the consequence of maximizing the wrong log-likelihood function (A.1). Draper and Cox (1969) tried to derive properties of $\hat{\lambda}$, but Hinkley (1975) found errors in their derivations that invalidate some of their results. Moreover, Hinkley stated, under rather loose conditions, a theorem giving the asymptotic normal distribution of $\hat{\boldsymbol{\theta}}_n$. Theorem 1 gives uniformity conditions under which $\hat{\boldsymbol{\theta}}_n$ is strongly consistent and has an asymptotic normal distribution. We first record some properties of the transformation $H_P(Y;\lambda)$.

Lemma 1: Define $\upsilon:\ (0,\infty)\times(-\infty,\infty)\to(-\infty,\infty)$ as

$$\upsilon(x;\lambda) \;=\; \begin{cases} (x^\lambda - 1)/\lambda & \lambda \neq 0\ , \\ \log x & \lambda = 0\ . \end{cases}$$

Then (a) $\upsilon(x, \lambda) > 0$ if $x > 1$, and $\upsilon(x, \lambda) \leq 0$ if $0 < x \leq 1$; (b) $\upsilon(\cdot, \cdot)$ is increasing in both variables; (c) $\upsilon(\cdot, \cdot)$ is convex in λ for $x \geq 1$ and concave in λ for $x \leq 1$; (d) $(\partial^r / \partial \lambda^r)\upsilon(x, \lambda)$ is continuous in x and λ, $r \geq 1$.

Let $L_{\mathrm{P}}(\boldsymbol{\theta}|\boldsymbol{X}_n)$ be given by (A.1) and $L_{\mathrm{P}}(\boldsymbol{\theta}|X) = L_{\mathrm{P}}(\boldsymbol{\theta}|X_1)$.

Theorem 1: Suppose the parameter space Θ, the true pdf $h(\cdot)$, and the log-likelihood function (A.1) satisfy the following conditions:

(i) The parameter space Θ is a compact set defined as

$$\Theta = \Big\{\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, \lambda)' \Big| |\boldsymbol{\beta}| \leq M, \quad c \leq \sigma \leq d, \quad a \leq \lambda \leq b \\ \text{with} \quad \infty < a < 0 < b, d, c, M < \infty\Big\}. \tag{A.2}$$

(ii) The true pdf $h(\cdot)$ is concentrated on $(0, \infty)$, and the moments $\mathrm{E}_h(X^{2a})$ and $\mathrm{E}_h(X^{2b})$ are finite.

(iii) $\mathrm{E}_h[L_{\mathrm{P}}(\boldsymbol{\theta}|X)]$ has a unique global maximum at $\boldsymbol{\theta}_0$.

Then the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ is a strongly consistent estimator of $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \sigma_0, \lambda_0)'$. Furthermore, if

(iv) $\boldsymbol{\theta}_0$ is an interior point of Θ,

(v) $\mathrm{E}_h[X^a \log(X)]^2$ and $\mathrm{E}_h[X^b \log(X)]^2$ are finite,

(vi) $\mathrm{E}_h[\nabla L_{\mathrm{P}}(\boldsymbol{\theta}_0|X)] = 0$, where the column vector

$$\nabla L_{\mathrm{P}}(\boldsymbol{\theta}_0|X) = \Big(\frac{\partial L_{\mathrm{P}}(\boldsymbol{\theta}|X)}{\partial \theta_i}\Big|_{\theta=\theta_0}\Big)$$

is the gradient of the log-likelihood function for

$$\boldsymbol{\theta}' = (\theta_1, \theta_2, \theta_3) = (\boldsymbol{\beta}, \sigma, \lambda),$$

(vii) $\mathrm{E}_h[\nabla^2 L_{\mathrm{P}}(\boldsymbol{\theta}_0|X)]$ is nonsingular, where

$$\nabla^2 L_{\mathrm{P}}(\boldsymbol{\theta}_0|X) = \Big(\frac{\partial^2 L_{\mathrm{P}}(\boldsymbol{\theta}|X)}{\partial \theta_i \partial \theta_j}\Big|_{\theta=\theta_0}\Big)$$

is the Hessian of the log-likelihood function, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}_3(\mathbf{0}, VWV'),$$

where $V = \{\mathrm{E}_h[\nabla^2 L_{\mathrm{P}}(\boldsymbol{\theta}_0|X)]\}^{-1}$ and $W = \mathrm{E}_h\{\nabla L_{\mathrm{P}}(\boldsymbol{\theta}_0|X)[\nabla L_{\mathrm{P}}(\boldsymbol{\theta}_0|X)]'\}$.

We indicate the method of proof and refer the reader to Hernandez and Johnson (1979) or Hernandez (1978) for details.

References

1. Atkinson, A. C. (1985). *Plots, Transformation, and Regression.* Oxford University Press.

2. Atkinson, A. and Riani, M. (2000). *Robust Diagnostic Regression Analysis.* Springer.

3. Bartlett, M. S. (1947). The use of transformations. *Biometrics* **3**, 39–52.

4. Beverton, R. J. H. and Holt, S. J. (1957). *On the dynamics of Exploited fish populations.* Fishery Investigations Series II, Volume XIX, Ministry of Agriculture, Fisheries and Food.

5. Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *Journal of the American Statistical Association* **80**, 580–619.

6. Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26**, 211–252.

7. Box, G. E. P. and Hill, W. J. (1974). Correcting inhomogeneticy of variance power-transformation weighting. *Technometrics* **16**, 385–389.

8. Bruce, D. and Schumacher, F. X. (1935). *Forest Mensuration.* New York: McGraw-Hill.

9. Carroll, R. J. and Ruppert, D. (1984). Power transformation when fitting theoretical models to data. *Journal of the American Statistical Association* **79**, 321–328.

10. Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* London: Chapman & Hall.

11. Channer, K. S. and Roberts, C. J. (1985). Effect of delayed esophageal transit on acetaminophen absorption. *Clinical Pharmacology and Therapeutics* **17**, 72–76.

12. Davison, A. C. (1986). Approximate predictive likelihood. *Biometrika***73**, 323–332.

13. Draper, N. R. and Cox, D. R. (1969). On Distributions and Their Transformation to Normality. *Journal of the Royal Statistical Society*, Ser. B, **31**, 472–476.

14. Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall.

15. Goto, M. (1974). Curve fitting. *Quality*, **4(1)**, 14–24.

16. Goto, M., Inoue, T., and Tsuchiya, Y. (1987). Double power-transformation and its performances: An extensive version of Box-Cox transformation. *Journal of the Japan Statistical Society* **17**, 149–163.

17. Goto, M., Yamamoto, S. and Inoue, T. (1991). Estimation of parameters of the power-normal distribution: Asymptotic behaviors of estimators of parameters. *Bulletin of the Computational Statistics of Japan*, **4(1)**, 45-60.

18. Goto, M. (1992). Extensive views of power transformation: Some recent developments. In *Proceedings of Honolulu Conference on Computational Statistics as a Memorial of the Fifth Anniversary of JSCS*, Honolulu, USA, December 1-5.

19. Goto, M. (1995). Double power-transformations and their applications. In *Proceedings of International Conference on Statistical Computing for Quality and Productivity Improvement, Volume I,* 386-397, Seoul, Korea, August 17-19.

20. Goto, M. and Daimon, T. (2000). Power transformation approaches to fitting of the compartment models. *Proceedings of the XXth of Meeting of the International Biometric Society*, **2**, 97, Berkeley, July 4.

21. Goto, M., Isomura, T., and Hamasaki, T. (2000). Guinea pigs in statistical science. In *Proceedings of the Tenth Japan and Korea Joint Conference of Statistics 2000*, 265–285, Oita, Japan, December 4-5.

22. Goto, M. and Hamasaki, T. (2003). Simultaneous diagnostic methods for the double power-normal transformation model. *Journal of the Japanese Society of Computational Statistics* **16(1)**, 39–52.

23. Hamasaki, T. and Goto, M. (2003). Diagnostic methods in the asymmetric power-transformation model for ordered categorical data. *Bulletin of Informatics and Cybernetics* **35(1–2)**, 1–18.

24. Hamasaki, T. and Goto, M. (2005). Power-transformation model to additivity in regression. *Beheviormetrika* **32**, 111–125.

25. Hernandez, F. (1978). The Large Sample Behavior of Transformations to Normal or Expornential Distributions. *unpublished PhD dissertation*, University of Wisconsin, Madison.

26. Hernandez, F. and Johnson, R. A. (1979). The Large Sample Behavior of Transformations to Normality. *Technical Report No. 545*, University of Wisconsin, Madison, Dept. of Statistics.

27. Hernandez, F. and Johnson, R. A. (1980). The Large-Sample Behavior of Transformations to Normality. *Journal of the American Statistical Association* **75**, 855–861.

28. Hinkley, D. V. (1975). On Power Transformations to Symmetry. *Biometrika*, **62**, 101–111.

29. Howard, R. D. (1963). System analysis of linear models. *Multistage Inventory Models and Techniques*, ed. by Scart, H. E., Gilford, D. M. and Shelly, M. N., 143–185, Stanford University Press.

30. Ito, M., Daimon, T., Sugimoto, T. and Goto, M. (2001). Diagnosis by nonparametric approach for power-transformation (in Japanese). *Proceedings of the 15th Symposium of Japanese Society of Computational Statistics*, 107–110, Osaka, Japan.

31. Ito, M. and Goto, M. (2001). Various types of nonparametric-transformation and it's diagnosis (in Japanese). *Proceedings of the 24th Meetings of Statistical Workshop in Oita*, Japan.

32. Ito, M. (2002). Various types of nonparametric transformation and its diagnosis. *Unpublished Master thesis*, Osaka University.

33. Ito, M. and Goto, M. (2002). Inference and evaluation of nonparametric-transformation for non-linear models (in Japanese). *Proceedings of the 16th conference of Japanese Society of Computational Statistics*, 126–129, Kochi, Japan.

34. Ito, M. and Kawai, N. (2003). Exploration and evaluation for logistic model by nonparametric-transformation approach (in Japanese). *Proceedings of the Scientific Research Fund Symposium 2003*, Hiroshima, Japan.

35. Ito, M. (2003). Nonparametric-transformation and it's applications (in Japanese). *Proceedings of the 28th Meetings of Statistical Workshop in Oita*, Japan.

36. Ito, M. and Goto, M. (2004). Various types of nonparametric transformation and its diagnosis. *Japanese Journal of Applied Statistics* **33**, 3–26.

37. Ito, M. and Goto, M. (2005). Nonparametric-transformation models for double power-transformation (in Japanese). *Proceedings of the Cooperative Research Symposium 2004 by the Institute of Statistical Mathematics Projects*, 27–32, Tokyo, Japan.

38. Ito, M. and Goto, M. (2005). Inference for theoretical model based on transformation (in Japanese). *Proceedings of the 4th Seminar of Biostatistical Research Association*, Osaka, Japan.

39. Ito, M. and Goto, M. (2005). Nonparametric transformation-both-sides and weighted least squares: principle of inference and it's performance (in Japanese). *Proceedings of the 19th Symposium of Japanese Society of Computational Statistics*, 67–70, Kagoshima, Japan.

40. Ito, M. and Goto, M. (2005). Nonparametric transformation-both-sides with weighted error. *Proceedings of the 5th IASC Asian Conference on Statistical Computing*, 49–53, Hong Kong University.

41. Ito, M. and Goto, M. (2006). Semiparametric transformation for theoretical model. *Proceedings of the XXIIIrd International Biometric Conference, McGill university*, Montreal, Canada.

42. Ito, M. and Goto, M. (2006). Semiparametric Transformation for Non-linear Regression Model. *Journal of the Japanese Society of Computational Statistics* **19**, 57–68.

43. Ito, M. (2007). Various types of flexible design (in Japanese). *Proceedings of the 35th Meetings of Statistical Workshop in Oita*, Japan.

44. Ito, M. (2007). Flexible design and it's performance (in Japanese). *Proceedings of the Symposium of Adaptive Planning and Restricted Inference*, Osaka, Japan.

45. Ito, M. (2007). Various types of flexible design and it's points (in Japanese). *Proceedings of the Symposium "Possibility of Application for Flexible Design in Clinical Trial and It's Points" of Biostatistical Research Association*, Tokyo, Japan.

46. Ito, M. (2007). Estimation in flexible design: maximum likelihood and median unbiased (in Japanese). *Proceedings of the 36th Meetings of Statistical Workshop in Oita*, Japan.

47. Ito, M., Kuroishi, K. and Hirooka, H. (2008). Various types of flexible design: overall testing and overall estimate in 2-stage design (in Japanese). *Proceedings of the 22th conference of Japanese Society of Computational Statistics, 55-69*, Akita, Japan.

48. Ito, M. (2008). Semiparametric transformation for non-linear regression model. *Proceedings of the Joint Meeting of 4th World Conference of the IASC and 6th Conference of the Asian Regional Section of the IASC on Computational Statistics and Data Analysis*, Yokohama, Japan.

49. Jimura, Y. and Goto, M. (1997). Inference on non-linear model based on transformation. Master thesis, Graduate School of Engineering Science, Osaka University.

50. Kume, H. (1971). System identification. *Quality Control*, **22(2)**, 35-41.

51. Nychka, D. and Ruppert, D. (1995). Nonparametric transformations for both sides of a regression model. *Journal of the Royal Statistical Society, Series B* **57**, 519–532.

52. O'Sullivan, F., Yandell, B. and Raynor, W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association* **81**, 96–104.

53. Ohta, H., Mori, K., Kase, S., Goto, M., Fujimoto, H. and Asano, Ch. (1968). Some considerations in system analysis (1): Investigations by Mason's rule (in Japanese). *Standardization and Quality Control*, 8, 68–73.

54. Goto, M., Fujimoto, H., Asano, Ch., Ohta, H., Mori, K. and Kato, S. (1968). Some considerations in system analysis (2): On GERT and multiple input and output system (in Japanese). *Standardization and Quality Control*, 10, 68–73.

55. Goto, M., Fujimoto, H., Asano, Ch., Ohta, H., Mori, K. and Kato, S. (1968). Some considerations in system analysis (3): On system analysis of semi-Markov process (in Japanese). *Standardization and Quality Control*, 11, 68–73.

56. Reinsch, C. H. (1967). Smoothing by spline functions. *Numerical Mathmatics* **10**, 177-183.

57. Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Research Board of Canada* **32**, 559–623.

58. Ricker, W. E. and Smith, H. D. (1975). A revised interpretation density function by maximum penalized method. *Annals of Statistics* **10**, 795-810.

59. Sakamoto, W. (2007). Extracting non-linear additive regression structure with power-additive smoothing splines. *Journal of the Japanese Society of Computational Statistics.* **20(1)**, 83–108.

60. Sakamoto, W. Inference on non-linear structure with penalized splines (Submitted).

61. Tasaki, T. and Goto, M. (1990). On transformation in generalized linear models. *Bulletin of the Computational Statistics of Japan* **3(1)**, 45–53.

62. Tibshirani, R. (1988). Estimating transformations for regression via additivity and variance stabilization. *Journal of the American Statistical Association* **83**, 394–405.

63. Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86.

64. Wang, N. and Ruppert, D. (1995). Nonparametric estimation of the transformation in the transform-both-sides regression model. *Journal of the American Statistical Association* **90**, 522–534.

List of Publications

Ito, M. (2002). Various types of nonparametric transformation and its diagnosis. *Unpublished Master thesis*, Osaka University.

Ito, M. and Goto, M. (2004). Various types of nonparametric transfromation and its diagnosis (in Japanese). *Japanese Journal of Applied Statistics* **33**, 3–26.

Ito, M. and Goto, M. (2005). Nonparametric transformation-both-sides with weighted error. *Proceedings of the 5th IASC Asian Conference on Statistical Computing*, 49–53, Hong Kong University.

Ito, M. and Goto, M. (2006). Semiparametric transformation for theoretical model. *Proceedings of the XXIIIrd International Biometric Conference, McGill university*, Montreal, Canada.

Ito, M. and Goto, M. (2006). Semiparametric Transformation for Non-linear Regression Model. *Journal of the Japanese Society of Computational Statistics* **19**, 57–68.

Ito, M. (2008). Semiparametric transformation for non-linear regression model. *Proceedings of the Joint Meeting of 4th World Conference of the IASC and 6th Conference of the Asian Regional Section of the IASC on Computational Statistics and Data Analysis*, Yokohama, Japan.