UNIVERSITY OF ALBERTA

INFERENCE FOLLOWING BOX-COX TRANSFORMATION

BY

ZHENLIN YANG ©

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

FALL 1992

ISBN 0-315-77380-4

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Zhenlin Yang

TITLE OF THESIS: Inference Following Box-Cox Transformation

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: FALL 1992

(Signed) Zhenlin Yang

Permanent Address:

Department of Management Engineering
Northeast University of Technology
Shenyang
The People's Republic of China

DATED 20 July 1992

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Inference Following Box-Cox Transformation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Peter Hooper
Supervisor
P. M. Hooper

D. Wiens

Jerome N. Sheahan
J. Sheahan

Edit Gombay
E. Gombay

Gordon Flowerdew
G. Flowerdew

for External Reader
R. Lockhart

Date: June 22, 1992

To my family

ABSTRACT

In many applications a transformation of the dependent variable is required to make a normal-theory linear model plausible. Box and Cox (1964) recommended that, having chosen a suitable transformation, one should estimate and interpret effects on this transformed scale, ignoring the fact that the transformation was estimated from the data. Their methodology was criticized by Bickel and Doksum (1981) and defended by Hinkley and Runger (1984). The dispute centered on the definition of the parameters of interest following the transformation. This thesis presents some results that help clarify this issue. We introduce a definition of data-based parameters of interest associated with the estimated transformation. A first-order asymptotic expansion of the usual normal-theory pivotal quantity demonstrates asymptotic validity of the Box-Cox methodology for inference on these data-based parameters. A second-order expansion shows that in moderate sized experiments validity is related to the linear model for the means. Validity is supported most strongly in structured models, such as regression models, where the transformation parameter is uniquely determined when the error variance is zero. Validity holds to a lesser extent in unstructured models, such as the one-way layout, where the transformation parameter is not estimable when the error variance is zero.

ACKNOWLEDGEMENTS

First and for most, I would like to thank my thesis supervisor, Professor Peter M. Hooper, for introducing me the thesis problem, and for his help and careful guidance. I would also like to thank Professor Douglas Wiens for his helpfulness, and my external examiner, Professor Richard Lockhart, for his critical reading of my thesis.

I wish to express my gratitude to Professor Hooper and the Department of Statistics and Applied Probability of the University of Alberta for the financial support.

Last but not the least, I would like to thank the department secretaries for their kindness and helpfulness.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure | Page

CHAPTER 1
INTRODUCTION

In theoretical statistics one usually adopts a probability model as a starting point and then considers inference about model parameters or prediction of unobserved random variables generated under the model. In applied statistics a model is often selected with the aid of the same data set that is used for further inference. This raises two related issues. First, model selection can be viewed as estimation within the context of a larger underlying model. Should inferences concern parameters defined in the larger model or parameters defined conditionally, given the model selected? Second, does the usual "model-given" analysis need to be adjusted to account for model selection? There are no generally accepted answers to these questions, no clear division between model selection and parametric estimation. In certain applications, however, it seems useful to distinguish between model selection and estimation, and focus attention on parameters defined conditionally, given the model selected. The "model-given" analysis may or may not be appropriate for inference on these data-dependent parameters.

The specification of a linear regression model involves selection of explanatory variables, transformation of explanatory variables, and transformation of the response variable. The issues described above apply to all three components of model selection. In this thesis we consider only data-based transformation of the response, assuming that explanatory variables are determined a priori.

The most commonly used family of transformations is the family of power transformations:

$$h(t, \lambda) = \begin{cases} (t^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0, \\ \log t & \text{if } \lambda = 0, \end{cases} \qquad t > 0. \tag{1.1}$$

Box and Cox (1964) developed methods of inference under the model:

$$h(y, \lambda) = X\beta + \sigma e, \tag{1.2}$$

where $y = (y_1, \ldots, y_n)^T$, the responses y_i are positive, $h(y, \lambda) = (h(y_1, \lambda), \ldots, h(y_n, \lambda))^T$ is the vector of transformed responses, β is a $p\times 1$ vector of regression parameters, σ is the standard deviation of the error term, X is a known $n\times p$ matrix of full rank p, and e is an $n\times 1$ vector of independent and identically distributed random variables. (^{T} denotes transpose.) The error distribution is assumed to be approximately normal. If $\lambda \neq 0$ then exact normality is incompatible with the assumption that responses are positive. The motivation for model (1.2) is that a transformation of the response variable will allow a model with a simple structure for the means, constant error variance, and approximate normality of the error distribution. Box and Cox also considered a two-parameter family of transformations, allowing a shift in location.

A commonly used method of estimation under model (1.2) is maximum likelihood with the error distribution approximated by the normal distribution. The approximate likelihood function is

$$\frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{ -\frac{\|h(y, \lambda) - X\beta\|^2}{2\sigma^2}\right\} J(\lambda, y), \tag{1.3}$$

where $\|\bullet\|$ denotes the Euclidean norm and

$$J(\lambda, y) = \prod_{i=1}^{n} \left| \frac{\partial h(y_i, \lambda)}{\partial y_i} \right| = \prod_{i=1}^{n} y_i^{\lambda-1}.$$

is the Jacobian of the transformation. Box and Cox obtained the estimates in two steps. For given λ, (1.3) is maximized by the usual normal-theory maximum likelihood estimates with responses $h(y_i, \lambda)$; i.e.,

$$\hat{\beta}(\lambda) = (X^T X)^{-1} X^T h(y, \lambda) \text{ and } \hat{\sigma}^2(\lambda) = (1/n) h^T(y, \lambda) Q h(y, \lambda) \tag{1.4}$$

where $Q = I_n - X(X^T X)^{-1} X^T$ and I_n is the $n \times n$ identity matrix. To estimate λ, substitute $\hat{\beta}(\lambda)$ and $\hat{\sigma}^2(\lambda)$ in (1.3) and maximize. We will refer to these estimators as the Box-Cox estimators of β, σ^2, and λ. In practice $\hat{\lambda}$ is often rounded off to a value that is easier to interpret. In modeling textile data, for example, Box and Cox obtained an estimate of $\hat{\lambda} = -.06$ with standard error $.06$. They recommended adopting $\hat{\lambda} = 0$, the log transformation.

Box and Cox recommended that, having chosen a suitable transformation, one should make the usual detailed estimation and interpretation of effects on this transformed scale. In other words, they suggested that inference be carried out as if it were known a priori that the transformation parameter λ equals the value $\hat{\lambda}$.

Bickel and Doksum (1981) criticized this approach. They showed that in some cases the asymptotic variance of the estimate of β is much larger when the transformation parameter λ is estimated than when it is known. As a result confidence intervals are liberal. Box and Cox (1982) argued that this variance inflation phenomenon is not relevant to their analysis because they interpret model effects in terms of a known transformation, hence the parameter-vector of interest is not the vector β related to the unknown λ in (1.1) but rather a vector related to the known value $\hat{\lambda}$. Hinkley and Runger (1984), in defending the Box-Cox approach, defined the parameter-vector of interest to be $E\{\hat{\beta} \mid \hat{\lambda}\}$. They used a Bayesian analysis with improper priors to show that confidence regions for these conditionally-defined parameters, excluding the grand mean, have correct coverage for large n. To obtain this result they replaced model (1.2) with a new model for the scaled responses $h(y,\lambda)/J(\lambda, y)^{1/n}$ and used the fact that the conditionally-defined parameters under the new model are stable with respect to changes in $\hat{\lambda}$. Bickel (1984) pointed out problems in their analysis where they pass from model (1.2) to the new model. He agreed that the Box-Cox methodology is appropriate for inference on $E\{\hat{\beta} \mid \hat{\lambda}\}$ but said he regarded the original parameter-vector β as being of greater interest.

To clarify the issues discussed above, it is helpful to introduce the following notation. Let $\beta(\lambda) = \beta$ and $\sigma(\lambda) = \sigma$ denote the parameters in (1.2) corresponding to the "true" value of λ. Let $\beta(\hat{\lambda})$ and $\sigma(\hat{\lambda})$ denote data-dependent parameters corresponding to the estimate $\hat{\lambda}$. Hinkley and Runger defined $\beta(\hat{\lambda}) = E\{\hat{\beta} \mid \hat{\lambda}\}$. We will introduce a slightly different definition in Chapter 2. A formal definition of $\sigma(\hat{\lambda})$ is not required here since $\sigma(\hat{\lambda})$ is treated as a nuisance parameter, however a definition would be needed for inference on residual variability conditioned on $\hat{\lambda}$. Let $\hat{\beta}(\lambda)$, $\hat{\sigma}(\lambda)$, $\hat{\beta}(\hat{\lambda})$, and $\hat{\sigma}(\hat{\lambda})$ be defined by (1.4); i.e., the maximum likelihood estimates when λ is known and estimated, respectively.

If λ is assumed known then inference for $\beta(\lambda)$ can be based on the pivotal quantity

$$\frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} = \frac{(X^TX)^{-1}X^Te}{\|(I_n - X(X^TX)^{-1}X^T)e\|/\sqrt{n}}. \tag{1.5}$$

The distribution of (1.5) under model (1.2) is closely approximated by its distribution with errors exactly normal. The Box-Cox analysis can be viewed as conditional inference for $\beta(\hat{\lambda})$ based on the pivotal quantity

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}, \tag{1.6}$$

with the conditional distribution of (1.6) given $\hat{\lambda}$ approximated by the distribution of (1.5). Bickel and Doksum's original interpretation of the Box-Cox analysis can be viewed as unconditional inference for $\beta(\lambda)$ based on the pivotal quantity

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta(\lambda)}{\hat{\sigma}(\hat{\lambda})}, \tag{1.7}$$

with the distribution of (1.7) approximated by that of (1.5). Their analysis shows that this latter approximation can be poor.

There are thus two issues to be resolved in evaluating the Box-Cox methodology. The first concerns the definition of the parameter-vector of interest. Our view is that, while some questions of interest may be best expressed in terms of $\beta(\lambda)$, in many instances it is simpler and more meaningful to express questions in terms of a data-dependent quantity $\beta(\hat{\lambda})$. This is discussed in Chapter 2, where a new definition of $\beta(\hat{\lambda})$ is developed. The second issue concerns the validity of approximating the conditional distribution of (1.6) by that of (1.5). First-order asymptotic results supporting this approximation are given in Chapter 3.

Further information about this approximation for moderate-sized experiments is obtained by studying how well the unconditional distribution of (1.6) is approximated by that of (1.5). If the latter approximation is poor then the approximation of the conditional distribution will typically be poor as well. Second-order expansions in Chapters 4 and 5 show that the validity of the approximation for the unconditional distribution depends on the structure of the model for the means and on the relative size of the variance σ^2. We will say that a model for the means is *structured* if λ and β in (1.2) are identifiable (i.e., uniquely determined from the data) when $\sigma = 0$. We will say that a model is *unstructured* if it is equivalent (after reparameterization) to a one-way layout. For a structured model, such as a regression model or a factorial model with additive effects, the unconditional approximation is generally good, and improves as σ decreases. For an unstructured model, such as a one-way layout, the unconditional approximation is good when σ is sufficiently large but can be very poor when σ is small, particularly when the individual means $x_i^T\beta$ are all close together.

A similar but conceptually easier problem where $\hat{\lambda}$ is specified a priori is studied in Chapter 6. In this case $\hat{\lambda}$ is just a fixed number independent of the data y. Here the problem of parameter interpretation is less controversial since, in practice, one would assume that $\hat{\lambda}=\lambda$. We examine the the robustness of the usual inferences about $\beta(\hat{\lambda})$ when $\hat{\lambda}$ differs from λ. Theoretical and simulation results show that coverage probabilities of confidence regions are conservative in structured model with small $\sigma(\lambda)$. In unstructured model with small $\sigma(\lambda)$ we find that misspecification of λ results in heteroscedasticity. Simulation results suggest that this effect is small.

CHAPTER 2

PARAMETERS OF INTEREST AFTER TRANSFORMATION

We introduce issues concerning parameter definition and interpretation in the context of the two simple examples: a two-sample problem and a simple linear regression problem.

Two-sample problem. Consider a completely randomized experiment with two treatments. We wish to compare the distributions of the response variable y under the two treatments. Often it is assumed that the two distributions are approximately normal with the same variance. The distributions can then be compared simply by comparing the two means. If the two distributions have substantially different variances or shapes then comparison is more complicated. In some instances it is possible to find a monotone transformation such that the distributions of the transformed response variable have approximately the same variance and shape. The distributions can then be compared in a simple manner by comparing the means of the transformed variable.

More precisely, let y_{ij} denote the jth replicate for the ith treatment and suppose the y_{ij} are independent with distribution depending on i but not j. Suppose that the distribution of $h(y_{i1},\lambda)$ is approximately normal with the same variance for $i = 1$ and 2. Here h denotes the Box-Cox power transformation (1.1), although the discussion that follows is applicable to other families of transformations as well. If λ is known then the two distributions may be compared by comparing $E\{h(y_{11},\lambda)\}$ with $E\{h(y_{21},\lambda)\}$. A more typical situation is that λ is not known but we have a prior estimate l for λ; i.e., l is a fixed number, chosen independently of the data $\{y_{ij}\}$. If l is close to λ then the distribution of $h(y_{i1},l)$ will be approximately normal with approximately the same variance for $i = 1$ and 2. If the shapes and variances of the two distributions are not too different, a comparison of $E\{h(y_{11},l)\}$ with $E\{h(y_{21},l)\}$

provides a simple comparison of the two distributions that is adequate for most purposes.

The situation described above is well understood and is not controversial. Conceptual difficulties arise when the transformation parameter is not chosen a priori but is estimated from the data $\{y_{ij}\}$. After calculating an estimate $\hat{\lambda}$, it is not clear which two distributions are to be compared. In the Box-Cox approach, we pretend that the number $l \equiv \hat{\lambda}$ was chosen a priori and then we proceed as in the previous paragraph. If we obtain $\hat{\lambda} = 0$, for example, then the distribution of $h(y_{i1}, 0)$ should be approximately normal with approximately the same variance for $i = 1$ and 2. The two distributions may be compared by comparing $E\{h(y_{11}, 0)\}$ with $E\{h(y_{21}, 0)\}$. The interpretation of these means is unclear, however, because $\hat{\lambda}$ and $\{y_{ij}\}$ are not independent. The joint distribution of $\{h(y_{ij}, 0)\}$ is not the same as the joint conditional distribution of $\{h(y_{ij}, \hat{\lambda})\}$ given $\hat{\lambda} = 0$. In particular, $E\{h(y_{ij}, 0)\}$ does not equal $E\{h(y_{ij}, \hat{\lambda}) \mid \hat{\lambda} = 0\}$.

To clarify interpretation of the Box-Cox approach, it is helpful to introduce random variables $\tilde{y}_{ij}$, distributed independently of $\{y_{ij}\}$ and hence of $\hat{\lambda}$, such that the joint distribution $\{\tilde{y}_{ij}\}$ is the same as that of $\{y_{ij}\}$. One can think of the $\tilde{y}_{ij}$ as the results of a future replication of the entire experiment. Now, comparing the distributions of y_{11} and y_{21} is equivalent to comparing the distributions of $\tilde{y}_{11}$ and $\tilde{y}_{21}$. Given $\hat{\lambda}$, the random variables $h(\tilde{y}_{ij}, \hat{\lambda})$ comprise two independent random samples. If $\hat{\lambda}$ is close to λ then the conditional distribution of $h(\tilde{y}_{i1}, \hat{\lambda})$ will be approximately normal with approximately the same variance for $i = 1$ and 2. If the shapes and variances of the two conditional distributions are not too different then a comparison of $E\{h(\tilde{y}_{11}, \hat{\lambda}) \mid \hat{\lambda}\}$ with $E\{h(\tilde{y}_{21}, \hat{\lambda}) \mid \hat{\lambda}\}$ provides a simple comparison of the two distributions that is adequate for most purposes.

It should be emphasized that $E\{h(\tilde{y}_{ij}, \hat{\lambda}) \mid \hat{\lambda}\}$ depends only on the realized value of $\hat{\lambda}$ and not on the definition of the estimator $\hat{\lambda}$. Thus the interpretation of the mean of the transformed response is the same in the Box-Cox approach as the interpretation when the transformation parameter is chosen a priori. This is in contrast to the definition $\mathrm{E}\{h(y_{ij},\hat{\lambda}) \mid \hat{\lambda}\}$ adopted by Hinkley and Runger in Section 3 of their paper, although in the examples of Section 2 they appear to be using the definition given in the previous paragraph; several discussants to the paper made comments to this effect.

The situations where λ is estimated a priori and a posteriori may be distinguished in two respects. The first difference is partly psychological. When we estimate λ from the data we are aware that $\hat{\lambda}$ does not equal λ and hence that the conditional distribution of $h(\tilde{y}_{ij},\hat{\lambda})$ given $\hat{\lambda}$ is only approximated by the usual model assumptions. When we use an a priori estimate l for λ we typically pretend that l equals λ and ignore the fact that the distribution of $h(y_{ij},l)$ is only approximated by the usual model assumptions. In the former situation we are forced to consider issues of parameter interpretation under departures from model assumptions and of robust inference. In the latter, these issues are often ignored.

The second difference is more substantial: the effect of conditioning on $\hat{\lambda}$ on the resulting inferences for $\mathrm{E}\{h(\tilde{y}_{ij},\hat{\lambda}) \mid \hat{\lambda}\}$. If we had access to the data $\{\tilde{y}_{ij}\}$ from the fictitious second experiment, we could apply standard methods of inference to $\{h(\tilde{y}_{ij},\hat{\lambda})\}$ treating $\hat{\lambda}$ as constant. This would be equivalent to applying the same methods to $\{h(y_{ij},l)\}$ with l chosen a priori. The usual robustness issues apply in both situations although, in the latter, robustness is less important if our prior information about λ is sufficiently good. In the Box-Cox analysis, standard methods of inference are applied to $\{h(y_{ij},\hat{\lambda})\}$, not $\{h(\tilde{y}_{ij},\hat{\lambda})\}$, so the effect of dependence between $\{y_{ij}\}$ and $\hat{\lambda}$ must also be considered. Our results in Chapter 3 show that the effect is asymptotically negligible, at least to first order.

Bickel (1984) and Doksum (1984) argued that it is more correct to compare the two treatments in terms of $\mathrm{E}\{h(y_{i1},\lambda)\}$ rather than $\mathrm{E}\{h(\tilde{y}_{i1},\hat{\lambda}) \mid \hat{\lambda}\}$. In contrasting their approach to that of Box and Cox, it is helpful to distinguish between hypothesis testing and estimation.

Consider testing the null hypothesis of no treatment effect. Put $\Delta(\lambda) = E\{h(y_{11},\lambda)\} - E\{h(y_{21},\lambda)\}$ and $\Delta(\hat{\lambda}) = E\{h(\tilde{y}_{11},\hat{\lambda}) \mid \hat{\lambda}\} - E\{h(\tilde{y}_{21},\hat{\lambda}) \mid \hat{\lambda}\}$. In the Bickel-Doksum approach, the null hypothesis is expressed as H_0: $\Delta(\lambda) = 0$ and the test is carried out taking into account the fact that λ is unknown. In the Box-Cox approach, the null hypothesis can be expressed as H_0: $\Delta(\hat{\lambda}) = 0$ and the test is carried out pretending that $\hat{\lambda} = \lambda$. Both approaches yield approximate level α tests; see Doksum and Wong (1983). While the Bickel-Doksum approach seems theoretically more correct, it is not clear that either test has more accurate level or better power than the other. The two tests are asymptotically equivalent to first order. The Box-Cox test is easier to compute using standard software packages. The results of the two tests are equally simple to interpret.

Now consider estimation of the difference between treatment effects. In the Box-Cox approach, one computes a confidence interval for $\Delta(\hat{\lambda})$ pretending that $\hat{\lambda} = \lambda$. The coverage probability of the confidence interval is approximately correct. The parameter $\Delta(\hat{\lambda})$ is easy to interpret because $\hat{\lambda}$ is known. Thus the interval for $\Delta(\hat{\lambda})$, together with the value $\hat{\lambda}$, provides a concise summary of the difference between the two treatment effects. In the Bickel-Doksum approach, one can compute a confidence interval for $\Delta(\lambda)$ but the interpretation of $\Delta(\lambda)$ depends on the unknown λ. Thus the interval, by itself, does not summarize the difference between the two treatment effects. A joint confidence region for λ and $\Delta(\lambda)$ does summarize the difference, but in a less concise manner than an interval for $\Delta(\hat{\lambda})$. The Box-Cox approach separates variability due to estimating the transformation from variability due to estimating the difference on the transformed scale, and focuses on the latter. In the Bickel-Doksum approach, the two kinds of variability cannot be separated.

It is not always appropriate to compare treatments in terms of the means of the transformed response. Rubin (1984) pointed out that the transformation on which comparisons are required is not always the same as the transformation to normality . In some applications it is necessary to compare the locations of the distributions of y_{11} and y_{21}. If the distribution of y_{ij} is asymmetric then the location parameter is not uniquely defined. If location is taken to be the median then the Box-Cox approach yields an approximate solution. The conditional distribution of $h(\tilde{y}_{i1},\hat{\lambda})$ given $\hat{\lambda}$ is approximately symmetric so $h^{-1}(\mathrm{E}\{h(\tilde{y}_{i1},\hat{\lambda}) \mid \hat{\lambda}\},\hat{\lambda})$ approximates the median of the distribution of y_{i1}. Here $h^{-1}(\bullet,\lambda)$ denotes the inverse of the function $h(\bullet,\lambda)$. An estimate for the median is obtained by substituting an estimate for the mean of the transformed response. Carroll and Rupert (1981) showed that, for this estimation problem, there is some cost due to estimating λ but the cost is typically small. If location is taken to be the mean then the estimation problem is more difficult; see Taylor (1986).

Simple linear regression. Consider an experiment in which values of a response variable y are obtained at various levels of an explanatory variable v. We wish to describe how the distribution of y varies with v. In some instances it is possible to find a monotone transformation of the response variable such that the mean of the transformed response is approximated by a linear function of v, the variance is approximately constant, and the distribution is approximately normal. The distribution can then be described by estimating the parameters of the linear model for the transformed response.

More precisely, suppose the explanatory variable takes on values $v_1, ..., v_n$, which are regarded as fixed. If the v_i are in fact random then we implicitly condition on their realized values in all probability statements that follow. Suppose the responses y_i are independent random variables with distribution related to v_i via the model

$$h(y_*,\lambda) = \beta_0 + \beta_1 v_* + \sigma e_*, \tag{2.1}$$

where e_* has mean zero, variance one, and distribution approximately normal. Suppose λ is not known but a prior estimate l for λ is available; i.e., l is a fixed number, independent of the data $\{y_i\}$. If $l \neq \lambda$ then $\mathrm{E}\{h(y_i,l)\}$ is not a linear function of v_i but if l is sufficiently close to λ then the mean is adequately approximated by a linear function. We have

$$h(y_*,l) = \beta_0(l) + \beta_1(l)v_* + \sigma(l)e_*(l), \tag{2.2}$$

where the distribution of $e_*(l)$ varies with v_*, but with mean approximately zero, variance approximately one, and distribution approximately normal. The parameters $\beta_0(l)$ and $\beta_1(l)$ can be formally defined in terms of model (2.1) so that $\beta_0(l) + \beta_1(l)v_*$ is on average close to $\mathrm{E}\{h(y_*,l) \mid v_*\}$ in some sense; e.g., for a specified measure $m(dv_*)$ on the real line, define $(\beta_0(l),\beta_1(l))$ to minimize

$$\int [\mathrm{E}\{h(y_*,l) \mid v_*\} - \beta_0(l) - \beta_1(l)v_*]^2 m(dv_*).$$

It is shown below that if the measure m is taken to be the empirical distribution of $\{v_1, \ldots, v_n\}$ then $\beta_0(l) = \mathrm{E}\{\hat{\beta}_0(l)\}$ and $\beta_1(l) = \mathrm{E}\{\hat{\beta}_1(l)\}$, where $\hat{\beta}_0(l)$ and $\hat{\beta}_1(l)$ are the least square estimators; i.e., putting $z_i = h(y_i,l)$,

$$\hat{\beta}_1(l) = \frac{\sum (v_i - \bar{v})(z_i - \bar{z})}{\sum (v_i - \bar{v})^2} \quad \text{and} \quad \hat{\beta}_0(l) = \bar{z} - \hat{\beta}_1(l)\bar{v}\,. \tag{2.3}$$

Now suppose a prior estimate l is not available but an estimate $\hat{\lambda}$ is obtained from the data. By an argument essentially the same as that in the two-sample problem, the Box-Cox analysis can be interpreted in terms of a fictitious random variable $\tilde{e}_*$, distributed independently of $\{y_i\}$ and hence of $\hat{\lambda}$, with the same distribution as e_* in (2.1). Define $\tilde{y}_*$ by

$$h(\tilde{y}_*,\lambda) = \beta_0 + \beta_1 v_* + \sigma \tilde{e}_*. \tag{2.4}$$

Given $\hat{\lambda}$, parameters $\beta_0(\hat{\lambda})$ and $\beta_1(\hat{\lambda})$ can be formally defined in terms of model (2.1) so that $\beta_0(\hat{\lambda}) + \beta_1(\hat{\lambda})v_*$ is on average close to $E\{h(\tilde{y}_*,\hat{\lambda}) \mid v_*,\hat{\lambda}\}$ in some sense; e.g., for a specified measure $m(dv_*)$ on the real line, define $(\beta_0(\hat{\lambda}),\beta_1(\hat{\lambda}))$ to minimize

$$\int [E\{h(\tilde{y}_*,\hat{\lambda}) \mid v_*,\hat{\lambda}\} - \beta_0(\hat{\lambda}) - \beta_1(\hat{\lambda})v_*]^2 m(dv_*).$$

If the measure m is taken to be the empirical distribution of $\{v_1, \ldots, v_n\}$ then $\beta_0(\hat{\lambda})$ and $\beta_1(\hat{\lambda})$ are given by (2.3) with z_i replaced by $h(\tilde{y}_i,\hat{\lambda})$, where the $\tilde{y}_i$ are fictitious random variables with the same joint distribution as the y_i but distributed independently of $\hat{\lambda}$.

The parameter $\beta_1(\hat{\lambda})$ can be interpreted as the approximate rate of increase in $E\{h(\tilde{y}_*,\hat{\lambda}) \mid v_*,\hat{\lambda}\}$ as v_* varies. This interpretation depends only on the value of the estimate $\hat{\lambda}$ and not on way the estimate is obtained. The Box-Cox and Bickel-Doksum approaches to inference may again be compared for hypothesis testing and estimation problems.

Consider testing the null hypothesis that the distribution of y_* is unrelated to v_*. In the Bickel-Doksum approach, the null hypothesis is expressed as H_0: $\beta_1(\lambda) = 0$ and the test is carried out taking into account the fact that λ is unknown. In the Box-Cox approach, the null hypothesis can be expressed as H_0: $\beta_1(\hat{\lambda}) = 0$ and the test is carried out pretending that $\hat{\lambda} = \lambda$. Both approaches yield approximate level α tests; see Carroll (1982). While the Bickel-Doksum approach seems theoretically more correct, it is not clear that either test has more accurate level or better power than the other. The two tests are asymptotically equivalent to first order. The Box-Cox test is easier to compute using standard software packages. The results of the two tests are equally simple to interpret.

Now consider estimation of the slope. In the Box-Cox approach, one computes a confidence interval for $\beta_1(\hat{\lambda})$ pretend[illegible] $\hat{\lambda} = \lambda$. The coverage probability of the confidence interval is approximately correct. The parameter $\beta_1(\hat{\lambda})$ is easy to interpret because $\hat{\lambda}$ is known. Thus the interval for $\beta_1(\hat{\lambda})$, together with the value $\hat{\lambda}$, provides a concise description of how the distribution of y varies with v. In the Bickel-Doksum approach, one can compute a confidence interval for $\beta_1(\lambda)$ but the interpretation of $\beta_1(\lambda)$ depends on the unknown λ. Thus the interval, by itself, does not describe how the distribution of y varies with v. A joint confidence region for λ and $\beta_1(\lambda)$ does provide such a description, but in a less concise manner than an interval for $\beta_1(\hat{\lambda})$.

The above ideas can be applied in a straightforward manner to the general linear model (1.2); i.e., let the y_i be independent random variables with distribution related to p-vectors x_i via

$$h(y_*,\lambda) = x_*^T\beta + \sigma e_* \,. \tag{2.5}$$

where e_* is approximately normal with mean zero and variance one. If λ is unknown but a prior estimate l is available then (2.5) may be rewritten

$$h(y_*,l) = x_*^T\beta(l) + \sigma(l)e_*(l)$$

where the distribution of $e_*(l)$ varies with x_*, but with mean approximately zero, variance approximately one, and distribution approximately normal. For a specified measure $m(dx_*)$ on the real line, we can define $\beta(l)$ to minimize

$$\int \{\mathrm{E}\{h(y_*,l) \mid x_*\} - x_*^T\beta(l)\}^2 m(dx_*) = \int \mathrm{E}\{h(y_*,l) \mid x_*\}^2 m(dx_*) - 2b^T\beta(l) + \beta(l)^T A\beta(l),$$

where

$$b = \int \mathrm{E}\{h(y_*,l) \mid x_*\} x_* m(dx_*) \text{ and } A = \int x_* x_*^T m(dx_*).$$

We thus have

$$\beta(l) = A^{-1}b.$$

If the measure m is taken to be the empirical distribution of $\{x_1, ..., x_n\}$ then we have $A = n^{-1}X^TX$, $b = n^{-1}X^T\mathrm{E}\{h(y,l)\}$, and

$$\beta(l) = (X^TX)^{-1}X^T\mathrm{E}\{h(y,l)\} = \mathrm{E}\{\hat{\beta}(l)\}. \tag{2.6}$$

If $\hat{\lambda}$ is an estimate of λ based on the data $\{y_i\}$ then we define $\beta(\hat{\lambda})$ as above but with y_* replaced by an independent $\tilde{y}_*$. Thus, letting $\tilde{y}$ denote a random n-vector, distributed independently of y but with the same distribution as y, we replace (2.6) by

$$\beta(\hat{\lambda}) = (X^TX)^{-1}X^T\mathrm{E}\{h(\tilde{y},\hat{\lambda}) \mid \hat{\lambda}\}. \tag{2.7}$$

Again, the interpretation of $\beta(\hat{\lambda})$ depends only on the value of the estimate $\hat{\lambda}$ and not on the choice of estimator.

We will assume that (2.7) defines the parameter-vector of interest when evaluating the least squares estimator $\hat{\beta}(\hat{\lambda}) = (X^TX)^{-1}X^Th(y,\hat{\lambda})$. For evaluating other estimators, such as M estimators, it is convenient to adopt a different definition. Suppose that if λ were known we would estimate β by $\hat{\beta}(\lambda) = f(y,\lambda)$ for a specified function f. Suppose, for λ unknown, we adopt an estimator $\hat{\beta}(\hat{\lambda})$ of the form $f(y,\hat{\lambda})$, where $\hat{\lambda}$ is an estimator of λ and f is the same function as before. Letting $\beta_0 : \mathrm{R} \rightarrow \mathbb{R}^p$ denote the function

$$\beta_u(l) = E\{\hat{\beta}(l)\} = E\{f(y,l)\},\qquad (2.8)$$

we define the parameter-vector of interest, given $\hat{\lambda}$, to be $\beta_u(\hat{\lambda})$. We note that the parameters β and λ are usually estimated jointly and that the function f is not uniquely determined by an estimator $(\hat{\beta}, \hat{\lambda})$. Definition (2.8) is based on the assumption that the estimator of β when λ is unknown is defined in terms of $\hat{\lambda}$ and the estimator of β when λ is known. We note as well that the $\beta_u(\hat{\lambda})$ depends only on the value of the estimate $\hat{\lambda}$ and not on the definition of the estimator $\hat{\lambda}$. We adopt definition (2.8) primarily as a matter of convenience in studying M estimators, to avoid difficulties with asymptotic bias. We feel the definition is well motivated for least squares estimators, where it agrees with (2.7), but the motivation for other estimators is less clear.

It is useful to consider a second definition of parameter-vector of interest:

$$\beta_c(\hat{\lambda}) = E\{\hat{\beta}(\hat{\lambda}) \mid \hat{\lambda}\}.\qquad (2.9)$$

This is the definition adopted by Hinkley and Runger (1984, Section 3). Note that $\beta_c(\hat{\lambda})$ depends on the way $\hat{\lambda}$ is estimated and not just on the value of the estimate. This makes its interpretation more problematic. As Bickel (1984) noted, however, definitions $\beta_u(\hat{\lambda})$ and $\beta_c(\hat{\lambda})$ are asymptotically equivalent to first order. The second definition is easier to work with for some purposes.

CHAPTER 3

ASYMPTOTIC VALIDITY OF USUAL INFERENCES FOLLOWING THE BOX-COX TRANSFORMATION

1. Introduction

As we discussed in Chapter 1, the object of the thesis can be summarized as comparing the λ-unknown pivotal quantities (1.6) and (1.7), in particular (1.6), with the λ-known pivotal quantity (1.5). Our primary interest is to make such comparisons when the estimators involved are the Box-Cox estimators and the transformation applied to the dependent variable is the Box-Cox power transformation, however much of the theory developed here and in Chapter 4 and Chapter 5 is based on general M-estimators and a general strictly increasing transformation. In evaluating the Box-Cox transformation methodology, the most relevant comparison is to compare the conditional distribution of (1.6) given $\hat{\lambda}$ with the unconditional distribution of (1.5). In this chapter we make the comparison for large n. Formal results obtained concern the validity of using the pivotal quantity (1.5) to approximate the pivotal quantity (1.6). Heuristic arguments are employed to discuss the validity of using the distribution of (1.5) to approximate the conditional distribution of (1.6). The consequences of using different definitions for $\beta(\hat{\lambda})$ in (1.6) are investigated. The problems of consistency and asymptotic normality are also discussed.

We first present some heuristic results in Section 3.2. If the transformation estimation does not introduce too much curvature on $g(z,\hat{\lambda},\lambda)=h[h^{-1}(z,\lambda),\hat{\lambda}]=h(y,\hat{\lambda})$, then we expect that (1.6) can be approximated by (1.5). If $\hat{\beta}(\hat{\lambda})$, $\hat{\lambda}$, and $\hat{\sigma}(\hat{\lambda})$ in (1.6) are the joint maximum likelihood estimators (MLE), and $\beta(\hat{\lambda}) = \beta_c(\hat{\lambda})$, the conditional and unconditional distributions of (1.6) can both be approximated by the distribution of (1.5).

Section 3.3 gives a formal first-order asymptotic expansion of (1.6) and (1.7) where the estimators are the general M-estimators and $\beta(\hat{\lambda}) = \beta_u(\hat{\lambda})$. Theorem 3.2 concludes that the first-order term in the expanded (1.6) agrees with (1.5). The result following Theorem 3.2 indicates that the first-order term in the expanded (1.7) does not agree with (1.5) hence the approximation to the distribution of (1.7) by that of (1.5) will generally be poor.

In Section 3.4, we study the problems of consistency and asymptotic normality. The root-n consistency of the joint estimator of $(\beta(\lambda),\lambda,\sigma(\lambda))$ is the key assumption for the results of Section 3.3. Bickel and Doksum (1981) considered this problem for M-estimators where in their model they allowed a simultaneous passage of $\sigma(\lambda)$ to zero as n goes to infinity to reduce technical difficulties. Their results are valid only for small $\sigma(\lambda)$. Hernandez and Johnson (1980) presented some theory concerning the (strong) consistency and asymptotic normality of the Box-Cox estimator in the one-sample case by giving some assumptions on the 'true' probability density function of y. Since the support of the distribution of $h(y_i,\lambda)$ cannot be the entire real line (unless λ=0), we consider a truncated normal model for the transformed variable. The model parameters along with the transformation parameter are jointly estimated by the maximum likelihood method. In most applications this should yield essentially the same result as the Box-Cox estimators obtained by maximizing (1.3). Results of Hoadley (1971) are applied to obtain regularity conditions for consistency and asymptotic normality. This provides support for the heuristic arguments of Section 3.2.

More importantly in Section 3.4 we show that using a definition of parameter of interest that is not (asymptotically) equivalent to $\beta_c(\hat{\lambda})$ inflates the variance of the pivotal quantity (1.6).

In Section 3.5 we present an example where exact distributional results are available to demonstrate the consequences of the Box-Cox analysis and the Bickel-Doksum analysis.

The asymptotic equivalence of $\beta_c(\hat{\lambda})$ and $\beta_u(\hat{\lambda})$ is discussed in Section 3.6. It is shown that, when the estimators are the MLEs defined in Section 3.4, the two definitions are asymptotically equivalent.

3.2. Two Heuristic Arguments

Consider first the Box-Cox estimators. Put $z = h(y,\lambda)$ and, for fixed l, define

$$g(z_i, l, \lambda) = h[h^{-1}(z_i, \lambda), l] = h(y_i, l). \tag{3.1}$$

Letting $g(z, l, \lambda) = \{g(z_i, l, \lambda)\}_{n\times 1}$, we have from (2.6)

$$\beta_u(l) = (X^T X)^{-1} X^T \mathrm{E}\{g(z, l, \lambda)\}. \tag{3.2}$$

For the Box-Cox power transformation, the function $g(\bullet, l, \lambda)$ is strictly increasing, concave if $l < \lambda$, and convex if $l > \lambda$. Theoretical and empirical considerations suggest that, in most applications, $g(\bullet, \hat{\lambda}, \lambda)$ will be approximately linear over the range $\min\{z_i\}$ to $\max\{z_i\}$; i.e.,

$$g(z, \hat{\lambda}, \lambda) \approx c_0 + c_1 z, \tag{3.3}$$

where c_0 and c_1 are functions of λ, $\hat{\lambda}$, and $\beta(\lambda)$. Applying (3.1) and (3.2), we obtain

$$\beta_u(\hat{\lambda}) \approx c_1 \beta(\lambda) + c_0 (X^T X)^{-1} X^T 1_n, \tag{3.4}$$

where 1_n is a column vector of 1's. The estimation of the transformation parameter thus introduces a change of location and a change of scale in the vector z, and hence in

$\beta_u(\hat{\lambda})$. These location and scale changes are the principal cause of the variance inflation phenomenon observed by Bickel and Doksum.

If (3.3) holds and if the column space of X contain 1_n so $(I_n - X(X^TX)^{-1}X^T)1_n = 0$, then the Box-Cox estimators $\hat{\beta}(\hat{\lambda})$ and $\hat{\sigma}^2(\hat{\lambda})$ become

$$\begin{aligned}\hat{\beta}(\hat{\lambda}) &= c_0(X^TX)^{-1}X^T1_n + c_1(X^TX)^{-1}X^Tz \\ &= c_0(X^TX)^{-1}X^T1_n + c_1\hat{\beta}(\lambda),\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2(\hat{\lambda}) &= \frac{1}{n}\|(I_n - X(X^TX)^{-1}X^T)g(z, \hat{\lambda}, \lambda)\|^2 \\ &\approx c_1^2\hat{\sigma}^2(\lambda).\end{aligned}$$

Hence

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})} \approx \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} \tag{3.5}$$

and so the two pivotal quantities have approximately the same distribution.

The preceding argument does not tell us whether the conditional distribution of the pivotal quantity (1.6), given $\hat{\lambda}$, is well approximated by the unconditional distribution of (1.5). This question is addressed in a non-rigorous manner as follows. For the convenience of discussion we put $\theta(\lambda) = (\beta^T(\lambda), \sigma(\lambda))^T$. Suppose that $[\hat{\theta}(\hat{\lambda}), \hat{\lambda}]$ is the joint MLE of $[\theta(\lambda), \lambda]$ and $\hat{\theta}(\lambda)$ is the MLE of $\theta(\lambda)$ calculated under the assumption that λ is known. Suppose that

$$[\hat{\theta}^T(\hat{\lambda}), \hat{\lambda}]^T \rightarrow_{\mathcal{P}} [\theta^T(\lambda), \lambda]^T,$$

and

$$[\hat{\theta}^T(\hat{\lambda}), \hat{\lambda}]^T \sim_{\mathcal{D}} AN\left[(\theta^T(\lambda), \lambda)^T, \ \mathcal{I}_n(\theta(\lambda), \lambda)^{-1}\right],$$

where $\longrightarrow_{\mathcal{P}}$ indicates ' converge in probability ', $\sim_{\mathcal{D}}$ indicates ' distributed as ', $I_n(\theta(\lambda), \lambda)$ is the information matrix, and $X_n \sim_{\mathcal{D}} AN(\mu_n, \Sigma_n)$ indicates that X_n is asymptotically normal, which is equivalent to saying that the quantity $\sqrt{n}(X_n - \mu_n)$ converges in distribution to $N[0, \lim n\Sigma_n]$.

We partition $I_n(\theta(\lambda), \lambda)$ according to $\theta(\lambda)$ and λ, and denote the submatrices by $I_{n11}(\theta(\lambda))$, $I_{n12}(\theta(\lambda), \lambda)$, $I_{n21}(\lambda, \theta(\lambda))$, and $I_{n22}(\lambda)$. When λ is known, the λ-known MLE $\hat{\theta}(\lambda)$ of $\theta(\lambda)$ should also be consistent and asymptotically normal, i.e.,

$$\hat{\theta}(\lambda) \longrightarrow_{\mathcal{P}} \theta^T(\lambda), \text{ and } \hat{\theta}(\lambda) \sim_{\mathcal{D}} AN\left[\theta(\lambda),\ I_{n11}(\theta(\lambda))^{-1}\right].$$

For evaluating the validity of the Box-Cox transformation methodology, we wish to compare the conditional distribution of $[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]/\hat{\sigma}(\hat{\lambda})$ given $\hat{\lambda}$ with the unconditional distribution of $[\hat{\beta}(\lambda) - \beta(\lambda)]/\hat{\sigma}(\lambda)$. We use definition $\beta_c(\hat{\lambda})$ to approximate $\beta_u(\hat{\lambda})$. The asymptotic equivalence of the two will be discussed in Section 3.6. Now for large n we have

$$\hat{\theta}(\lambda) \overset{\text{appr.}}{\sim_{\mathcal{D}}} N\left[\theta(\lambda),\ I_{n11}(\theta(\lambda))^{-1}\right],$$

and

$$[\hat{\theta}^T(\hat{\lambda}), \hat{\lambda}]^T \overset{\text{appr.}}{\sim_{\mathcal{D}}} N\left[\theta^T(\lambda), \lambda)^T,\ I_n(\theta(\lambda), \lambda)^{-1}\right].$$

Writing $I_n(\theta(\lambda), \lambda)^{-1} = \begin{bmatrix} I_{n11} & I_{n12} \\ I_{n21} & I_{n22} \end{bmatrix}^{-1} = \begin{bmatrix} I_n^{11} & I_n^{12} \\ I_n^{21} & I_n^{22} \end{bmatrix}$, we have

$$I_n^{11} = (I_{n11} - I_{n12} I_{n22}^{-1} I_{n21})^{-1}.$$

Similarly we have

$$I_{n11} = (I_n^{11} - I_n^{12}(I_n^{22})^{-1} I_n^{21})^{-1}.$$

Thus

$$\text{VAR}\left[\hat{\theta}(\hat{\lambda})\middle|\hat{\lambda}\right] \approx I_n^{11} - I_n^{12}(I_n^{22})^{-1}I_n^{21} = I_{n11}^{-1} \approx \text{VAR}\left[\hat{\theta}(\lambda)\right].$$

Hence

$$\text{VAR}\left[\hat{\beta}(\hat{\lambda})\middle|\hat{\lambda}\right] \approx \text{VAR}\left[\hat{\beta}(\lambda)\right],$$

and

$$\mathcal{L}\left\{[\hat{\beta}(\hat{\lambda}) - \beta_c(\hat{\lambda})]\middle|\hat{\lambda}\right\} \approx \mathcal{L}\{\hat{\beta}(\lambda) - \beta(\lambda)\}.$$

For large n, $\hat{\sigma}(\hat{\lambda}) \approx \hat{\sigma}(\lambda) \approx \sigma(\lambda)$, since $\hat{\sigma}(\hat{\lambda})$ and $\hat{\sigma}(\lambda)$ are both consistent for $\sigma(\lambda)$, hence

$$\mathcal{L}\left\{\frac{\hat{\beta}(\hat{\lambda}) - \beta_c(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}\middle|\hat{\lambda}\right\} \approx \mathcal{L}\left\{\frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)}\right\},$$

so that

$$\mathcal{L}\left\{\frac{\hat{\beta}(\hat{\lambda}) - \beta_c(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}\right\} \approx \mathcal{L}\left\{\frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)}\right\}.$$

Hence the distribution and the conditional distribution of (1.6) can both be approximated by the distribution of (1.5). Further since $I_{n12}I_{n22}^{-1}I_{n21} \geq 0$, then $\text{VAR}[\hat{\theta}(\hat{\lambda})] \approx (I_{n11} - I_{n12}I_{n22}^{-1}I_{n21})^{-1} \geq \text{VAR}[\hat{\theta}(\lambda)] \approx I_{n11}^{-1}$. It follows that

$$\text{VAR}[\hat{\beta}(\hat{\lambda})] \geq \text{VAR}[\hat{\beta}(\lambda)].$$

This is the variance inflation phenomenon observed by Bickel and Doksum (1981). Rigorous theorems about consistency and asymptotic normality of the MLE are given in Section 3.4. Also the consequences of using a different definition of $\beta(\hat{\lambda})$ are investigated in that section.

3.3. General First-Order Asymptotic Expansions

Let $\xi(\lambda) = [\beta^T(\lambda), \lambda, \sigma(\lambda)]^T$, and let $\hat{\xi}(\hat{\lambda}) = [\hat{\beta}^T(\hat{\lambda}), \hat{\lambda}, \hat{\sigma}(\hat{\lambda})]^T$ be an M-estimator of $\xi(\lambda)$, i.e., for a $p+2$ dimensional vector-valued function $\psi_i(y_i, \xi(\lambda))$, $\hat{\xi}(\hat{\lambda})$ solves the estimating equation

$$\sum_{i=1}^{n} \psi_i\left[y_i, \hat{\xi}(\hat{\lambda})\right] = 0_{(p+2)\times 1}. \tag{3.6}$$

The function ψ_i contains three elements: ψ_{1i}, ψ_{2i}, and ψ_{3i}. The p dimensional subfunction ψ_{1i} is associated with $\beta(\lambda)$ such that if λ and $\sigma(\lambda)$ are known we use ψ_{1i} to estimate $\beta(\lambda)$, and the scalar functions ψ_{2i} and ψ_{3i} are associated with λ and $\sigma(\lambda)$ respectively in a similar way. Hence when λ is known, the λ-known estimators $\hat{\beta}(\lambda)$ and $\hat{\sigma}(\lambda)$ are obtained by solving the equations

$$\begin{cases} \sum_{i=1}^{n} \psi_{1i}\left[y_i, \hat{\beta}(\lambda), \hat{\sigma}(\lambda)\right] = 0_{p\times 1}, \\ \sum_{i=1}^{n} \psi_{3i}\left[y_i, \hat{\beta}(\lambda), \hat{\sigma}(\lambda)\right] = 0_{1\times 1}. \end{cases} \tag{3.7}$$

Example 3.1. The ψ_i function associated with the Box-Cox estimators may be written as

$$\psi_i = \begin{cases} \psi_{1i}(y_i, \xi(\lambda)) = -\dfrac{1}{\sigma^2(\lambda)} x_i[h(y_i, \lambda) - x_i^T \beta(\lambda)], \\ \psi_{2i}(y_i, \xi(\lambda)) = \dfrac{1}{\sigma^2(\lambda)}[h(y_i, \lambda) - x_i^T \beta(\lambda)] h_\lambda(y_i, \lambda) - \dfrac{h_{y\lambda}(y_i, \lambda)}{h_y(y_i, \lambda)}, \\ \psi_{3i}(y_i, \xi(\lambda)) = \dfrac{1}{\sigma(\lambda)} - \dfrac{[h(y_i, \lambda) - x_i^T \beta(\lambda)]^2}{\sigma^3(\lambda)}, \end{cases} \tag{3.8}$$

where $h_\lambda(y_i, \lambda) = \frac{\partial h(y_i, \lambda)}{\partial \lambda}$, $h_y(y_i, \lambda) = \frac{\partial h(y_i, \lambda)}{\partial y_i}$, and $h_{y\lambda}(y_i, \lambda) = \frac{\partial^2}{\partial y_i \partial \lambda} h(y_i, \lambda)$.

Consider the estimators defined by (3.6). If the ψ_i function is smooth enough, and $\hat{\xi}(\hat{\lambda})$ is root-n consistent for $\xi(\lambda)$, then a direct application of first-order Taylor series expansions of $\psi_{1i}(y_i, \hat{\xi}(\hat{\lambda}))$ and $\psi_{3i}(y_i, \hat{\xi}(\hat{\lambda}))$ gives the first-order asymptotic expansions of (1.6) where $\beta(\hat{\lambda})=\beta_u(\hat{\lambda})$. We first introduce some notation applicable to all asymptotic expansions in the thesis.

Notation: Let $\psi_{i\xi}$ denote the first-order partial derivative of ψ_i with respect to $\xi(\lambda)$, i.e.,

$$\psi_{i\xi} = \psi_{i\xi}(y_i, \xi(\lambda)) = \frac{\partial}{\partial \xi^T(\lambda)} \psi_i(y_i, \xi(\lambda)).$$

Put

$$\psi = \frac{1}{n}\sum_{i=1}^{n} \psi_i, \quad \psi_\xi = \frac{1}{n}\sum_{i=1}^{n} \psi_{i\xi}, \quad \text{and } A_\xi = \mathrm{E}\psi_\xi.$$

Write

$$\psi_i = \begin{bmatrix} \psi_{1i} \\ \psi_{2i} \\ \psi_{3i} \end{bmatrix} \begin{matrix} p\times 1 \\ 1\times 1 \\ 1\times 1 \end{matrix} .$$

The $(p+2)\times 1$ vector ψ and the $(p+2)\times(p+2)$ matrices ψ_ξ and A_ξ have the corresponding partitions

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix}, \quad \psi_\xi = \begin{bmatrix} \psi_{1\beta} & \psi_{1\lambda} & \psi_{1\sigma} \\ \psi_{2\beta} & \psi_{2\lambda} & \psi_{2\sigma} \\ \psi_{3\beta} & \psi_{3\lambda} & \psi_{3\sigma} \end{bmatrix}, \text{ and } A_\xi = \begin{bmatrix} A_{1\beta} & A_{1\lambda} & A_{1\sigma} \\ A_{2\beta} & A_{2\lambda} & A_{2\sigma} \\ A_{3\beta} & A_{3\lambda} & A_{3\sigma} \end{bmatrix}.$$

Now we make some assumptions for the asymptotic expansions throughout the thesis:

Assumption **A1.** $\hat{\xi}(\hat{\lambda})$ is root-n consistent;

A2. $\psi_\xi = A_\xi + O_p(n^{-1/2})$;

A3. A_ξ and its inverse are $O(1)$;

A4. $\mathrm{E}\psi(y, \xi) = 0$.

We assume further that the ψ_i function is smooth enough so that the remainder term in Taylor expansion has the order of $[\hat{\xi}(\hat{\lambda}) - \xi(\lambda)]^m$, where m=2 and 3 corresponding to the first- and second-order Taylor expansion respectively. For example, when the first-order Taylor expansion of $\psi_{1i}(y_i, \hat{\xi}(\hat{\lambda}))$ around $\xi(\lambda)$ is carried out, the remainder should be of order $[\hat{\xi}(\hat{\lambda}) - \xi(\lambda)]^2 = O_p(n^{-1})$, which is true if the second-order derivatives of ψ_{1i} evaluated at ξ^* are $O_p(1)$, where ξ^* is an interior point of $L(\xi(\lambda), \hat{\xi}(\hat{\lambda}))$, the line segment joining $\xi(\lambda)$ and $\hat{\xi}(\hat{\lambda})$. We also assume that the random quantities which are bounded in probability have finite expectation.

Now define

$$g_1 = n^{1/2} B_{1\beta} \psi_1, \text{ where } B_{1\beta} = -A_{1\beta}^{-1}, \text{ and}$$

$$g_3 = n^{1/2} B_{3\sigma} \psi_3, \text{ where } B_{3\sigma} = -A_{3\sigma}^{-1}.$$

We have the following theorem.

Theorem 3.1. If g_1 and g_3 are $O_p(1)$ then, as $n \to \infty$, we have

$$\hat{\beta}(\hat{\lambda}) - \beta(\lambda) = n^{-1/2} B g_1 + B B_{1\beta} (A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda)$$

$$+ n^{-1/2} B B_{1\beta} A_{1\sigma} g_3 + O_p(n^{-1}), \tag{3.9}$$

where $B = (I_p - B_{1\beta} A_{1\sigma} B_{3\sigma} A_{3\beta})^{-1}$.

Proof. Since $\hat{\xi}(\hat{\lambda})$ is root-n consistent, by a first-order Taylor series expansion of $\psi_{1i}(y_i, \hat{\xi}(\hat{\lambda}))$ around $\xi(\lambda)$, we have

$$0 = \frac{1}{n}\sum_{i=1}^{n}\{\psi_{1i} + \psi_{1i\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \psi_{1i\lambda}(\hat{\lambda} - \lambda) + \psi_{1i\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))\} + O_p(n^{-1}),$$

which is equivalent to

$$0 = \psi_1 + \psi_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \psi_{1\lambda}(\hat{\lambda} - \lambda) + \psi_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1}).$$

By Assumption A2 we have

$$-A_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda) = \psi_1 + A_{1\lambda}(\hat{\lambda} - \lambda) + A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1}),$$

and pre-multiplying each side by $B_{1\beta}$ gives

$$\hat{\beta}(\hat{\lambda}) - \beta(\lambda) = n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1}). \quad (3.10)$$

Similarly, a first-order Taylor series expansion of $\psi_{3i}(y_i, \hat{\xi}(\hat{\lambda}))$ around $\xi(\lambda)$ gives

$$\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda) = n^{-1/2}g_3 + B_{3\sigma}A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + B_{3\sigma}A_{3\lambda}(\hat{\lambda} - \lambda) + O_p(n^{-1}). \quad (3.11)$$

Substituting (3.11) into (3.10) and collecting the terms of order $O_p(n^{-1})$ yield (3.1).

The Theorem 3.1 is useful in the sense that using it we can easily have a first-order asymptotic expansion for $\beta_u(\hat{\lambda})$, $\hat{\beta}(\lambda)$, and the desired quantity $\dfrac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}$. The results are stated in Corollary 3.1, Corollary 3.2, and Theorem 3.2 respectively.

Corollary 3.1. Under the assumptions of Theorem 3.1, we have

$$\beta_u(\hat{\lambda}) - \beta(\lambda) = BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})(\hat{\lambda} - \lambda) + O_p(n^{-1}). \quad (3.12)$$

Proof. From the definition of $\beta_u(\hat{\lambda})$, we simply take the expectation on both sides of (3.9) pretending $\hat{\lambda}$ is a fixed number.

Corollary 3.2. Under the assumptions of Theorem 3.1, we have the following asymptotic expansion for the λ-known estimator

$$\hat{\beta}(\lambda) - \beta(\lambda) = n^{-1/2}Bg_1 + n^{-1/2}BB_{1\beta}A_{1\sigma}g_3 + O_p(n^{-1}). \tag{3.13}$$

Proof. The proof can be completed by either following the proof of Theorem 3.1 and expanding $\psi_{1i}(y_i, \hat{\beta}(\lambda), \hat{\sigma}(\lambda))$ and $\psi_{3i}(y_i, \hat{\beta}(\lambda), \hat{\sigma}(\lambda))$ around $(\beta(\lambda), \sigma(\lambda))$ or simply equating $\hat{\lambda}$ to λ in (3.9).

Theorem 3.2. Under the assumptions of Theorem 3.1, we have

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})} = \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} + O_p(n^{-1}). \tag{3.14}$$

Proof. From Theorem 3.1 and Corollary 3.1, we have

$$\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda}) = \hat{\beta}(\lambda) - \beta(\lambda) + O_p(n^{-1}). \tag{3.15}$$

By the root-n consistency of $\hat{\sigma}(\hat{\lambda})$ and $\hat{\sigma}(\lambda)$ to $\sigma(\lambda)$, we can write

$$\frac{1}{\hat{\sigma}(\hat{\lambda})} = \frac{1}{\hat{\sigma}(\lambda)} + O_p(n^{-1/2}). \tag{3.16}$$

Multiplying (3.15) and (3.16) side by side, and combining the root-n consistency of $\hat{\beta}(\lambda)$ to $\beta(\lambda)$, we obtain (3.14).

We have obtained an important result (3.14), which says that by an error of order $O_p(n^{-1})$ the λ-unknown pivotal quantity (1.6) can be approximated by the corresponding λ-known quantity (1.5). Hence the overall effect of estimating λ from the data is small for large n. Theorem 3.2 is a theoretical justification of the heuristic argument leading to (3.5). It should be pointed out that any other definition which is within $O_p(n^{-1})$ of $\beta_u(\hat{\lambda})$ will give the same result as Theorem 3.2.

Further conclusions can be drawn from (3.14) if we can show that $(\hat{\beta}(\lambda), \hat{\sigma}(\lambda))$ is asymptotically independent of $\hat{\lambda}$. In this case the conditional distribution of $\{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})\}/\hat{\sigma}(\hat{\lambda}))$ given $\hat{\lambda}$ is asymptotically equivalent to the unconditional distribution of $\{\hat{\beta}(\lambda) - \beta(\lambda)\}/\hat{\sigma}(\lambda))$, hence the Box-Cox analysis is asymptotically justified. In fact this asymptotic independence holds for the MLE; see Bickel (1984).

A direct application of Theorem 3.1 also gives a first-order asymptotic expansion for the other λ-unknown quantity (1.7), i.e.,

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta(\lambda)}{\hat{\sigma}(\hat{\lambda})} = \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} + \frac{BB_{1\beta}}{\sigma(\lambda)}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})(\hat{\lambda} - \lambda) + O_p(n^{-1}). \quad (3.17)$$

From (3.17) we see that the first-order expansion for (1.7) has an extra first-order term. This is consistent with the variance inflation discussed by Bickel and Doksum.

3.4. Asymptotics for the MLE Under a Truncated Normal Model

Note. In this section and only in this section, a different notation is employed to facilitate the asymptotic theory. We use ξ_0 to represent the true model parameters and ξ to represent the parameter variable taking values in a parameter space.

3.4.1. General set up

Define $h : I_1 \times I_2 \longrightarrow I_3(\lambda)$, where I_1 and I_2 are open intervals with fixed end points, and $I_3(\lambda)$ is an open interval with the end points (possibly) depending on the second argument of h. Particular examples of the h function are as follows.

Box-Cox power transformation (Box and Cox, 1964):

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

where $t \in I_1 = (0, \infty)$, $\lambda \in I_2 = (a, b)$, $-\infty < a < b < \infty$,

$$\text{and } h \in I_3(\lambda) = \begin{cases} (-1/\lambda, \infty), & \text{if } \lambda > 0, \\ (-\infty, -1/\lambda), & \text{if } \lambda < 0, \\ (-\infty, \infty), & \text{if } \lambda = 0. \end{cases}$$

Bickel-Doksum transformation (Bickel and Doksum, 1981):

$$h(t, \lambda) = \frac{|t|^{\lambda} \operatorname{sgn}(t) - 1}{\lambda},$$

where $t \in I_1 = (-\infty, \infty)$, $\lambda \in I_2 = (0, b)$, $b < \infty$, and $I_3(\lambda) = (-\infty, \infty)$.

Note that the Box-Cox power transformation is only valid for $t > 0$, hence, unless $\lambda=0$, the distribution of $h(y,\lambda)$ can not be normal since $h(y,\lambda)$ has bounded support. The Bickel-Doksum transformation extends the range of t to the real line for $\lambda > 0$, so that an exact normal model after transformation can theoretically hold. We are only interested in applying the transformation to the positive data, however, so the Box-Cox power transformation seems more appropriate.

To allow for the implementation of the Box-Cox power transformation so that we have a theoretically correct likelihood function, a specific model has to be considered. Let $y_1, y_2, \ldots, y_n$ be a set of independent random variables each taking values in I_1. Write

$$z \equiv h(y, \lambda_0) = X\beta_0 + \sigma_0 e. \tag{3.18}$$

Assume that, for some unknown true parameters $\xi_0 = (\beta_0^T, \lambda_0, \sigma_0)^T$, the random variable $z_i = h(y_i, \lambda_0)$ has a *truncated normal distribution* with density

$$g_i(z_i \mid \xi_0) = \frac{1}{c_i(\xi_0)} \exp\{-\frac{1}{2\sigma_0^2}(z_i - x_i^T\beta_0)^2\}, \ z_i \in I_3(\lambda_0), \tag{3.19}$$

where

$$c_i(\xi_0) = \int_{I_3(\lambda_0)} \exp\{-\frac{1}{2\sigma_0^2}(z_i - x_i^T\beta_0)^2\}\, dz_i \,. \tag{3.20}$$

By a change of variable technique, the density of the original y_i is

$$f_i(y_i \mid \xi_0) = \frac{1}{c_i(\xi_0)} \exp\{-\frac{1}{2\sigma_0^2}(h(y_i, \lambda_0 - x_i^T\beta_0)^2\}\, h_y(y_i, \lambda_0),\ y_i \in I_1, \tag{3.21}$$

where $h_y(y_i, \lambda_0) = \frac{\partial}{\partial y_i} h(y_i, \lambda_0) > 0$ since $h(\bullet, \lambda_0)$ is assumed to be strictly increasing. The constant $c_i(\xi_0)$ becomes, in term of y_i's,

$$c_i(\xi_0) = \int_{I_1} \exp\{-\frac{1}{2\sigma_0^2}(h(y_i, \lambda_0) - x_i^T\beta_0)^2\}\, h_y(y_i, \lambda_0) dy_i \,. \tag{3.22}$$

Let $\Xi = \{\xi = (\beta^T, \lambda, \sigma)^T \mid \beta \in \Omega \subseteq \mathbb{R}^p, \sigma > 0, \lambda \in I_2\}$ and consider the family of distributions $\{ f_i(y_i \mid \xi) : \xi \in \Xi \}$.

Note that when the truncation is negligible, the MLEs defined by the model (3.18) are essentially the same as the Box-Cox estimators.

3.4.2. Calculations

We will calculate the score function, Fisher information matrix and its inverse, and observed Fisher information matrix. The detailed calculations are put in the Appendix A. All results are based on the following assumption of exchangeability of the order of differentiation and integration, i.e.,

Assumption. i) $\frac{\partial}{\partial\xi}\int f_i(y_i\,|\xi)\;dy_i \;=\int\frac{\partial}{\partial\xi}f_i(y_i\,|\xi)\;dy_i,$ (3.23a)

ii) $\frac{\partial^2}{\partial\xi\,\partial\xi^T}\int f_i(y_i\,|\xi)\;dy_i \;=\int\frac{\partial^2}{\partial\xi\,\partial\xi^T}f_i(y_i\,|\xi)\;dy_i.$ (3.23b)

Notation: i) E_ξ, VAR_ξ, and COV_ξ denote the expectation, variance, and covariance operators corresponding to ξ. The usual notations E, VAR, and COV corresponds to ξ_0.

ii) $\varepsilon_i(y_i,\xi) = \frac{h(y_i,\lambda) - x_i^T\beta}{\sigma}$, $\qquad \varepsilon(y,\xi) = \{\varepsilon_i(y_i,\xi)\}_{n\times 1}$,

$v_i(y_i,\xi) = \frac{h_{y\lambda}(y_i,\lambda)}{h_y(y_i,\lambda)} - \frac{1}{\sigma}\varepsilon_i(y_i,\xi)h_\lambda(y_i,\lambda)$, $\qquad v(y,\xi) = \{v_i(y_i,\xi)\}_{n\times 1}$.

and $h_y(y_i,\lambda)$, $h_\lambda(y_i,\lambda)$, $h_{y\lambda}(y_i,\lambda)$, and $h_{y\lambda\lambda}(y_i,\lambda)$ are partial derivatives of $h(y_i,\lambda)$ with respect to y_i and λ.

Now using (3.21) we have the ***Log-likelihood function***:

$$\ell(\xi,y) = -\sum_{i=1}^{n}\log c_i(\xi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[h(y_i,\lambda) - x_i^T\beta]^2 + \sum_{i=1}^{n}\log h_y(y_i,\lambda), \qquad (3.24)$$

and the ***Score function***, $S(\xi,y) = \frac{\partial}{\partial\xi}\,\ell(\xi,y)$,

$$= \begin{cases} \dfrac{\partial\ell(\xi,y)}{\partial\beta_j} \\[2ex] \dfrac{\partial\ell(\xi,y)}{\partial\lambda} \\[2ex] \dfrac{\partial\ell(\xi,y)}{\partial\sigma} \end{cases} = \begin{cases} \dfrac{1}{\sigma}\sum_{i=1}^{n}x_{ij}\,[\varepsilon_i(y_i,\xi) - E_\xi\varepsilon_i(y_i,\xi)], \quad j = 1,\ldots,p, \\[2ex] \sum_{i=1}^{n}[v_i(y_i,\xi) - E_\xi v_i(y_i,\xi)], \\[2ex] \dfrac{1}{\sigma}\sum_{i=1}^{n}[\varepsilon_i^2(y_i,\xi) - E_\xi\varepsilon_i^2(y_i,\xi)], \end{cases} \qquad (3.25a)$$

or in matrix form

$$= \begin{cases} \dfrac{\partial \ell(\xi, y)}{\partial \beta} \\[2ex] \dfrac{\partial \ell(\xi, y)}{\partial \lambda} \\[2ex] \dfrac{\partial \ell(\xi, y)}{\partial \sigma} \end{cases} = \begin{cases} \dfrac{1}{\sigma} X^T\{\varepsilon(y, \xi) - \mathrm{E}_\xi[\varepsilon(y, \xi)]\}, \\[2ex] 1_n^T\{ \nu(y, \xi) - \mathrm{E}_\xi[\nu(y, \xi)]\}, \\[2ex] \dfrac{1}{\sigma}\{\varepsilon^T(y, \xi)\varepsilon(y, \xi) - \mathrm{E}_\xi[\varepsilon^T(y, \xi)\varepsilon(y, \xi)]\}. \end{cases} \tag{3.25b}$$

The ***Information matrix,*** $I_n(\xi) = \mathrm{E}_\xi[S(\xi, y)S^T(\xi, y)]$, is

$$I_n(\xi) = \frac{1}{\sigma^2}\begin{bmatrix} X^T D(\xi)X, & \sigma X^T b_1(\xi), & X^T b_2(\xi) \\ \sigma b_1^T(\xi)X, & k_1, & k_2 \\ b_2^T(\xi)X, & k_2, & k_3 \end{bmatrix}, \tag{3.26}$$

where

$$\begin{aligned} k_1 &= \sigma^2\sum_{i=1}^{n}\mathrm{VAR}_\xi[\nu_i(y_i, \xi)], \\ k_2 &= \sigma\sum_{i=1}^{n}\mathrm{COV}_\xi[\varepsilon_i^2(y_i, \xi), \nu_i(y_i, \xi)], \\ k_3 &= \sum_{i=1}^{n}\mathrm{VAR}_\xi[\varepsilon_i^2(y_i, \xi)], \\ D(\xi) &= \mathrm{Diag}\{\mathrm{VAR}_\xi[\varepsilon_i(y_i, \xi)], i = 1, \ldots, n\}, \\ b_1(\xi) &= \{\mathrm{COV}_\xi[\varepsilon_i(y_i, \xi), \nu_i(y_i, \xi)]\}_{n\times 1}, \end{aligned}$$

and

$$b_2(\xi) = \{\mathrm{COV}_\xi[\varepsilon_i(y_i, \xi), \varepsilon_i^2(y_i, \xi)]\}_{n\times 1}.$$

The ***Observed information matrix,*** $OI_n(\xi) = -\dfrac{\partial}{\partial \xi^T} S(\xi, y) = -\dfrac{\partial^2}{\partial \xi\, \partial \xi^T}\ \ell(\xi, y)$, is

$$OI_n(\xi) = \frac{1}{\sigma^2}\begin{bmatrix} X^T D(\xi) X, & \sigma X^T[b_1(\xi) - u_1(y,\xi)], & X^T[b_2(\xi) + u_2(y,\xi)] \\ —, & k_1 - w_1, & k_2 - w_2 \\ —, & —, & k_3 + w_3 \end{bmatrix}, \qquad (3.27)$$

where

$$u_1 = \frac{1}{\sigma}\{h_\lambda(y_i, \lambda) - E_\xi[h_\lambda(y_i, \lambda)]\}_{n\times 1},$$

$$u_2 = 2\{\varepsilon_i(y_i, \xi) - E_\xi[\varepsilon_i(y_i, \xi)]\}_{n\times 1},$$

and

$$w_1 = \sum_{i=1}^{n} w_{1i}, \; w_2 = \sum_{i=1}^{n} w_{2i}, \text{ and } w_3 = \sum_{i=1}^{n} w_{3i},$$

where

$$\begin{aligned} w_{1i} &= \sigma^2 \frac{h_{y\lambda\lambda}(y_i, \lambda) h_y(y_i, \lambda) - [h_{y\lambda}(y_i, \lambda)]^2}{[h_y(y_i, \lambda)]^2} - [h_\lambda(y_i, \lambda)]^2 \\ &\quad - \sigma\varepsilon_i(y_i, \xi) h_{\lambda\lambda}(y_i, \lambda) - E_\xi\left\{\sigma^2 \frac{h_{y\lambda\lambda}(y_i, \lambda) h_y(y_i, \lambda) - [h_{y\lambda}(y_i, \lambda)]^2}{[h_y(y_i, \lambda)]^2}\right\} \\ &\quad - E_\xi\{[h_\lambda(y_i, \lambda)]^2 - \sigma\varepsilon_i(y_i, \xi) h_{\lambda\lambda}(y_i, \lambda)\}, \\ w_{2i} &= 2\{\varepsilon_i(y_i, \xi) h_\lambda(y_i, \lambda) - E_\xi[\varepsilon_i(y_i, \xi) h_\lambda(y_i, \lambda)]\}, \end{aligned}$$

and

$$w_{3i} = 3\{\varepsilon_i^2(y_i, \xi) - E_\xi[\varepsilon_i^2(y_i, \xi)]\}.$$

The lower triangle part of (3.27) denoted by '—' can be obtained by symmetry.

The ***Inverse of the information matrix,*** $\Sigma_n(\xi) = [\sigma^2 I_n(\xi)]^{-1}$, is

$$= \begin{bmatrix} (X^T D_* X)^{-1} + \frac{1}{k_0} H H^T, & -\frac{1}{k_0} H, & -\frac{1}{k_3 k_0}(k_0 H_2 + H H^T X^T b_2 - k_2 H) \\ —, & \frac{1}{k_0}, & -\frac{1}{k_3 k_0}(H^T X^T b_2 - k_2) \\ —, & —, & k \end{bmatrix} \qquad (3.28)$$

where

$$D_* = D_*(\xi) = D(\xi) - \frac{\sigma^2}{k_3} b_2(\xi) b_2^T(\xi),$$

$$H = H(\xi) = (X^T D_* X)^{-1} X^T b,$$

$$b = b(\xi) = b_1(\xi) - \frac{k_2}{k_3}\sigma b_2(\xi),$$

$$k_0 = k_0(\xi) = k_1 - \frac{k_2^2}{k_3} - b^T X H,$$

$$H_2 = H_2(\xi) = (X^T D_* X)^{-1} X^T b_2,$$

and

$$k = k(\xi) = -\frac{1}{k_3^2 k_0}[k_0 b_2^T X H_2 + (H^T X^T b_2 - k_2)^2] + \frac{1}{k_3}.$$

When λ_0 is known, the information matrix $I_n(\beta, \sigma)$ and the observed information matrix $OI_n(\beta, \sigma)$ are just the submatrices of (3.26) and (3.27) respectively, i.e.,

$$I_n(\beta, \sigma) = \frac{1}{\sigma^2}\begin{bmatrix} X^T D(\xi) X, & X^T b_2(\xi) \\ b_2^T(\xi) X, & k_3 \end{bmatrix}, \tag{3.29}$$

and

$$OI_n(\beta, \sigma) = \frac{1}{\sigma^2}\begin{bmatrix} X^T D(\xi) X, & X^T[b_2(\xi) + u_2(y,\xi)] \\ [b_2^T(\xi) + u_2^T(y,\xi)]X, & k_3 + w_3 \end{bmatrix}. \tag{3.30}$$

Thus,

$$\Sigma_n(\beta, \sigma) = [\sigma^2 I_n(\beta, \sigma)]^{-1}$$

$$= \begin{bmatrix} (X^T D_* X)^{-1}, & -\frac{1}{k_3} H_2 \\ -\frac{1}{k_3} H_2^T, & \frac{1}{k_3^2}(b_2^T X H_2 + k_3) \end{bmatrix}. \tag{3.31}$$

3.4.3. Consistency

Consider the log-likelihood function $\ell(\xi, y)$ in (3.24). The MLE of ξ_0 is a point in Ξ denoted by $\hat{\xi}_n$ which maximizes $\ell(\xi, y)$. We will assume that this maximum is attained by solving

$$S(\hat{\xi}_n, y) = 0.$$

where $S(\xi, y)$ is the score function given in (3.25b).

***Notation*:** i) For any random variable X, let

$X^{(B)} = X$ if $X \geq -B$, and $= -B$ otherwise, $B \geq 0$;

ii) For a sequence of random variables $X_1, X_2, \ldots$, let

$\mathrm{E}|X_i|^{1+\delta} < K$ mean that there exist positive constants K and δ so that this holds for $i = 1, 2, \ldots$.

Our MLE $\hat{\xi}_n$ of ξ_0 falls within the framework of Hoadley (1971). If the regularity conditions of Hoadley's Theorem 1 hold, then $\hat{\xi}_n$ converges in probability to ξ_0, i.e., $\hat{\xi}_n \rightarrow_{\mathcal{P}} \xi_0$. The following are the conditions under which Hoadley's Theorem 1 is applicable.

B1. Ξ is a compact subset of $\mathbf{R}^{p+2}$, $a \leq \lambda \leq b, -\infty < a < 0 < b < \infty$.

B2. The x_i's belongs to a compact subset of $\mathbf{R}^p$ such that X^TX/n converges to a positive definite matrix.

B3. $h_t(t, \lambda)$ is positive and continuous in both t and λ , and is increasing in λ if $t \geq 1$ and decreasing if $t \leq 1$.

B4. $\mathrm{E}[H_{1i}(y_i)]^{1+\delta} \leq K_1$, and $\mathrm{E}[H_{2i}(y_i)]^{1+\delta} \leq K_2$,

where

$$H_{1i}(y_i) = \begin{cases} \log\left[\dfrac{h_y(y_i, b)}{h_y(y_i, \lambda_0)}\right], & \text{if } y_i \geq 1, \\ 0, & \text{if } y_i \leq 1, \end{cases}$$

and

$$H_{2i}(y_i) = \begin{cases} \log\left[\dfrac{h_y(y_i, a)}{h_y(y_i, \lambda_0)}\right], & \text{if } y_i \le 1, \\ 0, & \text{if } y_i \ge 1. \end{cases}$$

B5. There exists $B > 0$ for which

(i) $\limsup \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[R_i^{(B)}(\xi)] < 0, \quad \xi \neq \xi_0;$

(ii) $\limsup \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[V_i^{(B)}(r)] < 0,$ where $V_i(r) = \sup\{R_i(\xi)\colon \|\xi\| > r\}$.

Theorem 3.3. If conditions B1–B5 hold, then $\hat{\xi}_n \to_{\mathcal{P}} \xi_0$.

Proof. We simply check the conditions C1–C5 of Hoadley's Theorem 1. First B1 implies C1, and C2 follows from (3.21). For C3(i), we have

$$\begin{aligned} R_i(\xi) &= \log\left[\frac{f_i(y_i \mid \xi)}{f_i(y_i \mid \xi_0)}\right] \\ &= \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right] + \frac{1}{2}\,\varepsilon_i^2(y_i, \xi_0) - \frac{1}{2}\,\varepsilon_i^2(y_i, \xi) + \log\left[\frac{h_y(y_i, \lambda)}{h_y(y_i, \lambda_0)}\right], \end{aligned}$$

hence

$$\begin{aligned} R_i(\xi, \rho) &= \sup_{\|t-\xi\| \le \rho} [R_i(t)] \\ &\le \sup_{\xi} \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right] + \sup_{\xi}\left[\left(\frac{1}{2}\,\varepsilon_i^2(y_i, \xi_0) - \frac{1}{2}\,\varepsilon_i^2(y_i, \xi)\right] + \sup_{\xi} \log\left[\frac{h_y(y_i, \lambda)}{h_y(y_i, \lambda_0)}\right] \\ &\le \sup_{\xi} \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right] + \frac{1}{2}\,\varepsilon_i^2(y_i, \xi_0) - \frac{1}{2}\inf_{\xi}\,\varepsilon_i^2(y_i, \xi) + \sup_{\lambda} \log\left[\frac{h_y(y_i, \lambda)}{h_y(y_i, \lambda_0)}\right]. \end{aligned}$$

By B3, we have

$$\begin{aligned} \sup_{\lambda} \log\left[\frac{h_y(y_i, \lambda)}{h_y(y_i, \lambda_0)}\right] &= \begin{cases} \log\left[\dfrac{h_y(y_i, b)}{h_y(y_i, \lambda_0)}\right], & \text{if } y_i \ge 1, \\[2ex] \log\left[\dfrac{h_y(y_i, a)}{h_y(y_i, \lambda_0)}\right], & \text{if } y_i \le 1, \end{cases} \\ &= H_{1i}(y_i) + H_{2i}(y_i). \end{aligned}$$

Since $\varepsilon_i^2(y_i, \xi) \geq 0$ and $\sup_{\xi} \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right] \geq 0$, hence by the fact that $Z^{(0)} \leq X^{(0)} + Y^{(0)}$ if $Z = X + Y$, for any random variables X and Y, we have

$$
\begin{aligned}
R_i^{(0)}(\xi, \rho) &\leq \sup_{\xi} \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right] + \frac{1}{2}\varepsilon_i^2(y_i, \xi_0) + H_{1i}^{(0)}(y_i) + H_{2i}^{(0)}(y_i) \\
&= \sup_{\xi} \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right] + \frac{1}{2}\varepsilon_i^2(y_i, \xi_0) + H_{1i}(y_i) + H_{2i}(y_i).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\mathrm{E}[R_i^{(0)}(\xi, \rho)]^{1+\delta} \\
&\leq \mathrm{E}\left\{\sup_{\xi} \log\left[\frac{c_i(\xi_0)}{c_i(\xi)}\right]\right\}^{1+\delta} + \mathrm{E}[\tfrac{1}{2}\varepsilon_i^2(y_i, \xi_0)]^{1+\delta} + \mathrm{E}[H_{1i}(y_i)]^{1+\delta} + \mathrm{E}[H_{2i}(y_i)]^{1+\delta}.
\end{aligned}
$$

B1 and B2 implies that $\inf_{\xi} c_i(\xi) > 0$, hence the first term of $\mathrm{E}[R_i^{(0)}(\xi, \rho)]^{1+\delta}$ is bounded. Since the 3rd and 4th terms are also bounded, then $\mathrm{E}[R_i^{(0)}(\xi, \rho)]^{1+\delta}$ is bounded if we can show that $\mathrm{E}[\frac{1}{2}\varepsilon_i^2(y_i, \xi_0)]^{1+\delta} < K$, for some $\delta > 0$ and $K > 0$.

Now choosing $\delta = 1$, we have

$$
\begin{aligned}
&\mathrm{E}[(\tfrac{1}{2}\varepsilon_i^2(y_i, \xi_0)]^2 \\
&= \int_{I_3(\lambda_0)} \left[\frac{(z_i - x_i^T\beta_0)^2}{2\sigma_0^2}\right]^2 \frac{1}{c_i(\xi_0)} \exp[-\frac{1}{2\sigma_0^2}(z_i - x_i^T\beta_0)^2]dz_i \\
&\leq \frac{\sqrt{2\pi}\sigma_0}{c_i(\xi_0)} \int_{-\infty}^{\infty} \left[\frac{(z_i - x_i^T\beta_0)^2}{2\sigma_0^2}\right]^2 \frac{1}{\sqrt{2\pi}\sigma_0} \exp[-\frac{1}{2\sigma_0^2}(z_i - x_i^T\beta_0)^2]dz_i \\
&= \frac{\sigma_0}{4c_i(\xi_0)} \int_{-\infty}^{\infty} t^4 \exp(-\frac{1}{2}t^2)dt \\
&= \frac{3\sqrt{2\pi}\,\sigma_0}{4\,c_i(\xi_0)}.
\end{aligned}
$$

The same procedure leads to C3(ii), i.e., $\mathrm{E}[V_i^{(0)}(r)]^{1+\delta} \leq K$.

B5 is C4 of Hoadley. C5 is immediate. The proof is complete.

When λ_0 is known, the conditions for consistency are easy to check. Let Ξ_0 denoted the parameter space when λ_0 is known, and $\hat{\beta}_{n0}$ and $\hat{\sigma}_{n0}$ denote the MLE of β_0 and σ_0 when λ_0 is known.

Theorem 3.4. If conditions B1 and B2 hold with Ξ and λ replaced by Ξ_0 and λ_0, then we have

$$\begin{pmatrix} \hat{\beta}_{n0} \\ \hat{\sigma}_{n0} \end{pmatrix} \rightarrow_{\mathcal{P}} \begin{pmatrix} \beta_0 \\ \sigma_0 \end{pmatrix}.$$

Proof. Straightforward from the proof of Theorem 3.1.

3.4.4. Asymptotic normality

Conditions for asymptotic normality of $\hat{\xi}_n$ are based on Theorem 2 of Hoadley. The reduced set of regularity conditions are:

D1. Ξ is a compact subset of $\mathbb{R}^{p+2}$ where $a \le \lambda \le b, -\infty < a < 0 < b < \infty$.

D2. $\hat{\xi}_n \rightarrow_{\mathcal{P}} \xi_0$.

D3. The x_i's belongs to a compact subset of $\mathbb{R}^p$ such that X^TX/n converges to a positive definite matrix.

D4. $h(t, \lambda)$ is three times continuously differentiable, once with respect to t and twice with respect to λ.

D5. $E_\xi[v_i(y_i, \xi]^2 < \infty$, and $E_\xi[\frac{\partial}{\partial\lambda} v_i(y_i, \xi)]$ is finite.

D6. $\frac{1}{n} I_n(\xi) \rightarrow I(\xi)$, where $I(\xi)$ is a positive definite matrix.

D7. For some $\delta > 0$, $\frac{1}{n^{(2+\delta)/2}} \sum_{i=1}^{n} E|v_i(y_i, \xi_0)|^{2+\delta} \rightarrow 0$.

D8. (i) There exist $\varepsilon > 0$ and random variable $B(y_i)$ such that

$$\text{Sup}\ \{\left|\frac{\partial}{\partial\lambda} v_i(y_i, \xi)\right| : |\lambda - \lambda_0| < \varepsilon\ \} \le B(y_i), \text{ with } \mathrm{E}|\mathrm{B}(y_i)|^{1+\delta} \le K,$$

(ii) $\mathrm{E}[h^2(y_i, b)] \le K$, $\mathrm{E}\left|\text{Sup}_{\lambda}\, h(y_i, \lambda)h_\lambda(y_i, \lambda)\right| \le K$,

and $\mathrm{E}\left|\text{Sup}_{\lambda}\, h_\lambda(y_i, \lambda)\right| \le K$.

Theorem 3.5. Under the regularity conditions D1– D8, we have

$$\sqrt{n}\left(\frac{\hat{\xi}_n - \xi_0}{\sigma_0}\right) \to_{\mathcal{P}} N\left[0, \lim_{n\to\infty} n\Sigma_n(\xi_0)\right],$$

where $\Sigma_n(\xi_0)$ is given in (3.28) by replacing ξ by ξ_0.

Proof. The proof of the theorem follows by checking the conditions N1–N9 of Hoadley's Theorem 2.

First, D1 implies N1, D2 is equivalent to N2, the conditions D4 and D5 imply N3 and N4, (3.23a) and (3.23b) imply N5 and N6, and D6 is equivalent to N7.

Second, it is easy to show that $\mathrm{E}[\varepsilon_i(y_i, \xi_0)]^6 \le K$, which combined with D7 gives N8.

Now we prove N9. By (3.27) the condition N9 reduces to: there exists $\varepsilon > 0$ such that the random variables $u_{1i}, u_{2i}, w_{1i}, w_{2i}$, and w_{3i} are all uniformly integrable in the ε-neighborhood of ξ_0. The uniform integrability of w_{1i} is assured by D8 (i). For the others, it suffices to show that the random variables $h(y_i, \lambda)$, $h(y_i, \lambda)h_\lambda(y_i, \lambda)$, $h_\lambda(y_i, \lambda)$, and $h^2(y_i, \lambda) - x_i^T\beta\, h(y_i, \lambda)$ are uniformly integrable in the neighborhood of ξ_0. Those were given in D8 (ii). The proof is complete.

When λ_0 is known the corresponding theorem is as follows.

Theorem 3.6. Under the conditions of Theorem 3.5, if $\frac{1}{n}I_n(\beta, \sigma) \longrightarrow I(\beta, \sigma)$, where $I(\beta, \sigma)$ is positive definite, then

$$\frac{\sqrt{n}}{\sigma_0}\left[\begin{pmatrix}\hat{\beta}_{n0}\\ \hat{\sigma}_{n0}\end{pmatrix}-\begin{pmatrix}\beta_0\\ \sigma_0\end{pmatrix}\right] \longrightarrow_{\mathcal{P}} N[0, \lim_{n\to\infty} n\Sigma_n(\beta_0, \sigma_0)]$$

where $\Sigma_n(\beta_0, \sigma_0)$ is obtained from (3.14) by replacing β by β_0 and σ by σ_0.

Proof. It is straightforward from the proof of Theorem 3.5.

Asymptotic normality of the MLE for λ_0 known and unknown cases has been established, which provides a theoretical basis for the heuristic argument of Section 3.2. Now we derive some results concerning the effects of using a different definition of parameter of interest. Let $\beta(\hat{\lambda}_n)$ denote a parameter-vector of interest, given $\hat{\lambda}$.

Theorem 3.7. Under the conditions of Theorem 3.5, if

i) $\beta(\lambda_0) = \beta_0 + o(n^{-1/2})$,

ii) $\beta_\lambda(\lambda_0)$ exists and converges to a finite vector as $n \to \infty$, where $\beta_\lambda(\lambda_0)$ is the derivative of $\beta(\hat{\lambda}_n)$ with respect to $\hat{\lambda}_n$ evaluated at λ_0, then we have

$$\sqrt{n}[\hat{\beta}_n - \beta(\hat{\lambda}_n)]/\sigma_0 \longrightarrow_{\mathcal{P}} N(0, \lim_{n\to\infty} n\Pi_n),$$

where

$$\Pi_n = (X^T D_* X)^{-1} + \Lambda_n,$$

$\Lambda_n = \frac{1}{k_0}[H + \beta_\lambda(\lambda_0)][H + \beta_\lambda(\lambda_0)]^T$, and Λ_n is non-negative definite,

where k_0, D_* and H are defined in (3.28).

Proof. Consider a first-order Taylor expansion of $\beta(\hat{\lambda}_n)$ around λ_0. We have

$$\beta(\hat{\lambda}_n) = \beta(\lambda_0) + (\hat{\lambda}_n - \lambda_0)[\ \beta_\lambda(\lambda_0) + R_n], \tag{3.32}$$

where $R_n \longrightarrow 0$ as $\hat{\lambda}_n \longrightarrow \lambda_0$. From Theorem 3.3, $\hat{\lambda}_n \longrightarrow_{\mathcal{P}} \lambda_0$, hence $R_n \longrightarrow_{\mathcal{P}} 0$. Substituting (3.32) into $\frac{\sqrt{n}}{\sigma_0}[\hat{\beta}_n - \beta(\hat{\lambda}_n)]$ gives

$$\begin{aligned}\frac{\sqrt{n}}{\sigma_0}[\hat{\beta}_n - \beta(\hat{\lambda}_n)] &= \frac{\sqrt{n}}{\sigma_0}\{\hat{\beta}_n - \beta(\lambda_0) - (\hat{\lambda}_n - \lambda_0)[\ \beta_\lambda(\lambda_0) + R_n]\} \\ &= \frac{\sqrt{n}}{\sigma_0} B_n(\hat{\xi}_n - \xi_0) - \frac{\sqrt{n}}{\sigma_0}(\hat{\lambda}_n - \lambda_0)R_n + \frac{\sqrt{n}}{\sigma_0} o(n^{-1/2}),\end{aligned}$$

where $B_n = \{I_p : -\beta_\lambda(\lambda_0) : 0_{p\times 1}\}$.

Now the last term is $o(1)$. By the fact that $\frac{\sqrt{n}}{\sigma_0}(\hat{\lambda}_n - \lambda_0)$ converges in distribution, and $R_n \longrightarrow_{\mathcal{P}} 0$, the second last term also converges in probability to 0. Now

$$\frac{\sqrt{n}}{\sigma_0} B_n(\hat{\xi}_n - \xi_0) \longrightarrow_{\mathcal{P}} N(0, \lim_{n\to\infty} nB_n \Sigma_n(\xi_0)B_n^T),$$

so that

$$\frac{\sqrt{n}}{\sigma_0}[\hat{\beta}_n - \beta(\hat{\lambda}_n)] \longrightarrow_{\mathcal{P}} N(0, \lim_{n\to\infty} nB_n \Sigma_n(\xi_0)B_n^T).$$

It remains to show that

$$B_n \Sigma_n(\xi_0)B_n^T = (X^T D_* X)^{-1} + \frac{1}{k_0}[H + \beta_\lambda(\lambda_0)][H + \beta_\lambda(\lambda_0)]^T,$$

which is straightforward from (3.28).

Now, $\frac{1}{k_0}$ is the asymptotic variance of $\hat{\lambda}_n$ which is a positive number, and the matrix $[H + \beta_\lambda(\lambda_0)][H + \beta_\lambda(\lambda_0)]^T$ is nonnegative definite. Hence Λ_n is nonnegative definite, completing the proof.

Corollary 3.3. Under the conditions of Theorem 3.5 and the additional regularity conditions described in the proof below, if $\beta(\hat{\lambda}_n) = \beta_c(\hat{\lambda}_n)$ then

$$\beta_\lambda(\lambda_0) = -H + o(n^{-1/2}),$$

hence

$$\sqrt{n}[\hat{\beta}_n - \beta_c(\hat{\lambda}_n)]/\sigma_0 \sim_{\mathcal{D}} \sqrt{n}[\hat{\beta}_{n0} - \beta_0]/\sigma_0.$$

Proof. From Theorem 3.5, we have under some additional regularity conditions

$$\mathrm{E}\{\sqrt{n}(\hat{\beta}_n - \beta_0) \mid \sqrt{n}\,(\hat{\lambda} - \lambda_0)\} = 0 + (-H/k_0)(1/k_0)^{-1}\sqrt{n}(\hat{\lambda} - \lambda_0) + o_p(1),$$

which gives

$$\beta(\hat{\lambda}_n) = \mathrm{E}[\hat{\beta}_n \mid \hat{\lambda}] = \beta_0 - H(\hat{\lambda}_n - \lambda_0) + o_p(n^{-1/2}).$$

We further assume that the derivative of the error term $o_p(n^{-1/2})$ with respect to $\hat{\lambda}$ at λ is $o(n^{-1/2})$. We then have

$$\beta_\lambda(\lambda_0) = -H + o(n^{-1/2}),$$

completing the proof.

We have seen in Theorem 3.7 that a different definition for $\beta(\hat{\lambda})$ which is not within $o_p(n^{-1/2})$ of $\beta_u(\hat{\lambda})$ also inflates the asymptotic variance of the λ-unknown pivotal quantity (1.6). Theorem 3.7 is important in the sense that it gives a general expression for the variance inflation factor of the λ-unknown pivotal quantity associated with a general definition of parameter of interest. The variance inflation over the λ-known quantity is represented by a nonnegative definite matrix $\Lambda_n = \frac{1}{k_0}H + \beta_\lambda(\lambda_0)][H + \beta_\lambda(\lambda_0)]^T$. If $\beta(\hat{\lambda}_n) = \beta_c(\hat{\lambda}_n)$ then $\Lambda_n = 0$, which means that there is no variance inflation in this case. If $\beta(\hat{\lambda}_n) = \beta_0$ then $\Lambda_n = \frac{1}{k_0}HH^T$, which is the variance inflation factor of for the Bickel-Doksum λ-unknown pivotal quantity (1.7).

3.5. Example: Simple Linear Regression Model With the Intercept Treated as a Transformation Parameter

Asymptotic results have been obtained to evaluate the analysis following the Box-Cox transformation. In this section we study an example where exact results are available to further demonstrate the consequences of a Box-Cox analysis and a Bickel-Doksum analysis. Consider the following simple linear regression model

$$h(y_i, \lambda) = y_i - \lambda = \beta(\lambda)x_i + \sigma(\lambda)e_i(\lambda), \text{ where } e_i(\lambda) \sim N(0, 1), i = 1, \ldots, n.$$

Note that here the x_i are real-valued. We consider a Box-Cox analysis with the intercept λ treated as a transformation parameter. The Box-Cox analysis interprets the slope conditionally, for a given value of intercept. The Bickel-Doksum analysis interprets the slope in the usual manner, without regard to the intercept. The Bickel-Doksum analysis is generally prefered here because fixing the intercept does not simplify interpretation of the slope.

When λ is known, the MLE or the least square estimators (LSE) of $\beta(\lambda)$ and $\sigma(\lambda)$ are

$$\hat{\beta}(\lambda) = \frac{\Sigma x_i(y_i - \lambda)}{\Sigma x_i^2}, \text{ and } \hat{\sigma}^2(\lambda) = \frac{1}{n}\Sigma(y_i - \lambda - \hat{\beta}(\lambda)x_i)^2.$$

We have

$$E(\hat{\beta}(\lambda)) = \beta(\lambda), \text{ Var}((\hat{\beta}(\lambda)) = \sigma^2(\lambda)/\Sigma x_i^2, \text{ and } \hat{\sigma}^2(\lambda) \sim \frac{1}{n}\sigma^2(\lambda)\chi^2_{n-1},$$

where χ^2_{n-1} denotes a χ^2 random variable with $n-1$ degrees of freedom.

The λ-known pivotal quantity

$$T(\lambda) = \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} \sim \frac{1}{\sqrt{\frac{n-1}{n}\Sigma x_i^2}} t_{n-1},$$

where t_{n-1} represents a t random variable with $n-1$ degrees of freedom.

Letting $\hat{\lambda}$ denote the least squares estimator of λ, we have by Definition (2.7),

$$\beta_u(\hat{\lambda}) = \beta(\lambda) + \frac{(\lambda - \hat{\lambda})\Sigma x_i}{\Sigma x_i^2}.$$

The usual estimators and the pivotal quantity obtained by pretending $\hat{\lambda} = \lambda$ are

$$\hat{\beta}(\hat{\lambda}) = \frac{\Sigma x_i(y_i - \hat{\lambda})}{\Sigma x_i^2}, \quad \hat{\sigma}^2(\hat{\lambda}) = \frac{1}{n}\Sigma(y_i - \hat{\lambda} - \hat{\beta}(\hat{\lambda})x_i)^2,$$

and

$$T(\hat{\lambda}) = \frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}.$$

Now,

$$\begin{aligned}\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda}) &= \frac{\Sigma x_i(y_i - \hat{\lambda})}{\Sigma x_i^2} - \beta(\lambda) - \frac{(\lambda - \hat{\lambda})\Sigma x_i}{\Sigma x_i^2} \\ &= \frac{\sigma(\lambda)\Sigma x_i e_i}{\Sigma x_i^2} \\ &= \hat{\beta}(\lambda) - \beta(\lambda),\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2(\hat{\lambda}) &= \frac{1}{n}\Sigma(y_i - \hat{\lambda} - \hat{\beta}(\hat{\lambda})x_i)^2 \\ &\sim \frac{1}{n}\sigma^2(\lambda)\chi^2_{n-2}.\end{aligned}$$

Since $\hat{\sigma}^2(\hat{\lambda})$ is independent of $(\hat{\beta}(\hat{\lambda}), \hat{\lambda})$, we have

$$T(\hat{\lambda}) \sim \sqrt{\frac{n-1}{n-2}} \bullet \frac{1}{\sqrt{\frac{n-1}{n}\Sigma x_i^2}} \bullet t_{n-2}.$$

Hence $T^2(\hat{\lambda})$ is stochastically greater than $T^2(\lambda)$, which implies that the confidence interval for the data-based parameter $\hat{\beta}(\hat{\lambda})$ is slightly liberal unconditionally.

Now consider the effect of replacing $\hat{\lambda}$ by a constant l. As above we have

$$\hat{\beta}(l) - \beta_u(l) = \hat{\beta}(\lambda) - \beta(\lambda).$$

Now

$$\begin{aligned}
\hat{\sigma}^2(l) &= \frac{1}{n}\Sigma(y_i - l - \hat{\beta}(l))x_i)^2 \\
&= \frac{1}{n} \|(I_n - \frac{1}{\|x\|^2} xx^T)(y - l1_n)\|^2 \\
&= \frac{1}{n} \|(I_n - \frac{1}{\|x\|^2} xx^T)(y - \lambda 1_n + (\lambda - l)1_n)\|^2 \\
&= \frac{1}{n} \|(I_n - \frac{1}{\|x\|^2} xx^T)(y - \lambda 1_n - \beta(\lambda)x + (\lambda - l)1_n)\|^2 \\
&= \frac{1}{n} \|(I_n - \frac{1}{\|x\|^2} xx^T)(e + \frac{\lambda - l}{\sigma(\lambda)}1_n)\|^2 \, \sigma^2(\lambda) \\
&\sim \frac{1}{n} \sigma^2(\lambda)\chi^2_{n-1}(\gamma),
\end{aligned}$$

where $\chi^2_{n-1}(\gamma)$ denotes a noncentral χ^2 random variable with $n-1$ degrees of freedom and noncentrality parameter γ. In this case we have

$$\begin{aligned}
\gamma &= \|(I_n - \frac{1}{\|x\|^2} xx^T) \, \frac{\lambda - l}{\sigma(\lambda)}1_n\|^2 \\
&= \frac{(\lambda - l)^2}{\sigma^2(\lambda)} (n - \frac{(\Sigma x_i)^2}{\Sigma x_i^2}).
\end{aligned}$$

The estimate $\hat{\sigma}^2(l)$ is stochastically larger than $\hat{\sigma}^2(\lambda)$, so $T^2(l)$ is stochastically smaller than $T^2(\lambda)$. The confidence interval for $\hat{\beta}(l)$ based on $T(l)$ is thus conservative.

When λ is estimated by $\hat{\lambda}$, the relevant method for evaluating the Box-Cox analysis is to evaluate the conditional distribution of $T(\hat{\lambda})$ given $\hat{\lambda}$. Using the joint normality of $(\hat{\beta}(\hat{\lambda}), \hat{\lambda})$ we can easily show that

$$\mathrm{E}\{\hat{\beta}(\hat{\lambda}) \mid \hat{\lambda}\} = \beta(\lambda) + \frac{(\lambda - \hat{\lambda})\Sigma x_i}{\Sigma x_i^2}.$$

Hence the two definitions $\beta_c(\hat{\lambda})$ and $\beta_u(\hat{\lambda})$ agree. Further it is also easy to show that $\hat{\beta}(\hat{\lambda}) - \mathrm{E}\{\hat{\beta}(\hat{\lambda}) \mid \hat{\lambda}\}$ and $\hat{\sigma}^2(\hat{\lambda})$ are independent of $\hat{\lambda}$. Thus $T(\hat{\lambda})$ is independent of $\hat{\lambda}$, so that the conditional distribution of $T(\hat{\lambda})$ given $\hat{\lambda}$ is the same as the unconditional distribution of $T(\hat{\lambda})$. Thus the Box-Cox confidence intervals are slightly liberal.

Finally suppose, as is usually the case here, that we are interested in $\beta(\lambda)$ after estimating λ by $\hat{\lambda}$. The distribution of the pivotal quantity

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta(\lambda)}{\hat{\sigma}(\hat{\lambda})} \sim \frac{1}{\sqrt{\frac{n-2}{n}\Sigma(x_i - \bar{x})^2}} t_{n-2}$$

$$= \sqrt{\frac{n-1}{n-2}} \cdot \sqrt{\frac{\Sigma x_i^2}{\Sigma(x_i - \bar{x})^2}} \cdot \frac{1}{\sqrt{\frac{n-1}{n}\Sigma x_i^2}} t_{n-2},$$

is more dispersed than that of $T(\hat{\lambda})$. It is thus easier to estimate the slope when the intercept is specified than when it is not.

We have seen from this example that there are some effects of making inferences based on the selected model. But if we have the ' right ' parameter of interest, the effect is very small, especially in the case of the transformation being estimated from the same set of data. When the transformation estimation is based on the prior information, then the cost could be large if the information is poor.

3.6. A Comparison of $\beta_c(\hat{\lambda})$ and $\beta_u(\hat{\lambda})$

Now we are back to the usual notation. We compare the two definitions $\beta_u(\hat{\lambda})$ and $\beta_c(\hat{\lambda})$ when the estimators are the MLEs defined in Section 3.4. The ψ_i function is

$$\begin{cases} \psi_{1i}(y_i, \xi(\lambda)) = -\dfrac{1}{\sigma(\lambda)} x_i^T[\varepsilon_i(y_i, \xi(\lambda)) - E_\xi\{\varepsilon_i(y_i, \xi(\lambda))\}], \\ \psi_{2i}(y_i, \xi(\lambda)) = -v_i(y_i, \xi(\lambda)) + E_\xi\{v_i(y_i, \xi(\lambda))\}, \\ \psi_{3i}(y_i, \xi(\lambda)) = -\dfrac{1}{\sigma(\lambda)}[\varepsilon_i^2(y_i, \xi(\lambda)) - E_\xi\{\varepsilon_i^2(y_i, \xi(\lambda))\}], \end{cases} \tag{3.33}$$

where

$$\varepsilon_i(y_i, \xi(\lambda)) = [h(y_i, \lambda) - x_i^T\beta]/\sigma(\lambda),$$

and

$$v_i(y_i, \xi(\lambda)) = \frac{h_{y\lambda}(y_i, \lambda)}{h_y(y_i, \lambda)} - \frac{1}{\sigma(\lambda)}\varepsilon_i(y_i, \xi(\lambda))h_\lambda(y_i, \lambda).$$

Hence the vector ψ, and the matrices ψ_ξ and A_ξ are the score function, observed information matrix, and information matrix respectively.

Now for Theorem 3.1, we have

$$\begin{aligned} A_{1\beta} &= \frac{1}{n}X^TDX, \\ A_{1\lambda} &= \frac{1}{n}\sigma(\lambda)X^Tb_1 \\ A_{1\sigma} &= \frac{1}{n}\sigma(\lambda)X^Tb_2, \\ A_{3\beta} &= A_{1\sigma}^T, \\ A_{3\lambda} &= \frac{k_2}{n}, \text{ and} \\ A_{3\sigma} &= \frac{k_3}{n}. \end{aligned}$$

Hence

$$
\begin{aligned}
B &= (I_p - B_{1\beta} A_{1\sigma} B_{3\sigma} A_{3\beta})^{-1} \\
&= [I_p - (X^T DX)^{-1} X^T b_2 b_2^T X / k_3]^{-1} \\
&= (X^T DX)(X^T D_* X)^{-1},
\end{aligned}
$$

$$
\begin{aligned}
A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda} &= \frac{1}{n}[\sigma(\lambda) X^T b_1 - \frac{k_2}{k_3} X^T b_2] \\
&= \frac{1}{n} X^T b,
\end{aligned}
$$

and

$$
BB_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) = -H.
$$

Theorem 3.1 thus becomes

$$
\hat{\beta}(\hat{\lambda}) - \beta(\lambda) = n^{-1/2} B g_1 - H(\hat{\lambda} - \lambda) + n^{-1/2} BB_{1\beta} A_{1\sigma} g_3 + O_p(n^{-1}), \qquad (3.34)
$$

Hence

$$
\beta_u(\hat{\lambda}) = \beta(\lambda) - H(\hat{\lambda} - \lambda) + O_p(n^{-1}),
$$

which may be compared with

$$
\beta_c(\hat{\lambda}) = \beta(\lambda) - H(\hat{\lambda} - \lambda) + o_p(n^{-1/2}),
$$

given in Section 3.4. Thus we have $\beta_u(\hat{\lambda}) = \beta_c(\hat{\lambda}) + o_p(n^{-1/2})$.

Under the general framework of Section 3.3, if $E(g_1 | \hat{\lambda})$ and $E(g_3 | \hat{\lambda})$ are both $o_p(1)$, then following Theorem 3.1 we have $\beta_u(\hat{\lambda}) = \beta_c(\hat{\lambda}) + o_p(n^{-1/2})$ provided that the conditional expectation of the error term of (3.9) given $\hat{\lambda}$ is still $o_p(n^{-1/2})$.

CHAPTER 4

SECOND-ORDER EXPANSIONS FOR M-ESTIMATORS

4.1 Introduction

The asymptotic validity of applying usual inference methods to a transformed linear model has been studied in Chapter 3. In particular, we have shown in Theorem 3.2 that if n is large the Box-Cox λ-unknown quantity (1.6) can be approximated by the corresponding λ-known quantity (1.5). We refer to this result as the first-order approximation to the Box-Cox λ-unknown quantity. When n is moderate to small, this first-order result may not be adequate in some cases. Our intuition suggests that in the one sample case, the smaller the $\sigma(\lambda)$, the harder it is to estimate λ. When $\sigma(\lambda)$ is identically zero, the parameter λ is not identifiable hence the estimate of λ and thereafter the effect of estimating λ could be anything. Intuition also tells that there are some similarities between one sample (one mean) and one-way layout (at least two means) when $\sigma(\lambda)$ is small, but the situation in one-way layout should be improved if the means are moved farther apart because increasing the distance between the means will increase the information about λ. We feel that the first-order approximation should in general be good for the structured model. In any case the accuracy of the first-order approximation is worth of study.

The present chapter deals only with general theory. In Section 4.2, we carry out general second-order expansions to the M-estimator and the related λ-unknown pivotal quantity. First we carry out the expansions in terms of $\hat{\lambda}$ which allows for the study of the Box-Cox transformation methodology when $\hat{\lambda}$ is independent of y. Second we obtain full expansions in terms of error vector $e(\lambda)$ only to study the unconditional behavior of the λ-unknown pivotal quantity. The full expansions are achieved by combining the expansions in terms of $\hat{\lambda}$ with a first-order expansion of

$\hat{\lambda}$ since $\hat{\lambda}$ is only involved in the second-order term. Finally we carry out an expansion for the variance of the λ-unknown quantity.

Since the proofs of some results are lengthy and straightforward, we put them into the Appendix B

4.2. General Second-Order Asymptotic Expansions

We continue with the general setup of Section 3.3 and add the following notation and assumption.

Notation: Let $\psi_{1i\beta\beta}$, $\psi_{1i\beta\lambda}$, $\psi_{1i\beta\sigma}$, $\psi_{2i\beta\beta}$, $\psi_{2i\beta\lambda}$, $\psi_{2i\beta\sigma}$, $\psi_{3i\beta\beta}$, $\psi_{3i\beta\lambda}$, $\psi_{3i\beta\sigma}$, $\psi_{1i\lambda\lambda}$, $\psi_{1i\sigma\sigma}$, $\psi_{1i\lambda\sigma}$, $\psi_{2i\lambda\lambda}$, $\psi_{2i\sigma\sigma}$, $\psi_{2i\lambda\sigma}$, $\psi_{3i\lambda\lambda}$, $\psi_{3i\sigma\sigma}$, and $\psi_{3i\lambda\sigma}$ denote the second-order partial derivatives of ψ_{1i}, ψ_{2i}, and ψ_{3i}, e.g.,

$$\psi_{1i\beta\beta} = \psi_{1i\beta\beta}\,(y_i, \xi) = \frac{\partial^2}{\partial\beta^T\partial\beta}\,\psi_{1i}(y_i, \xi),$$

$$\psi_{1i\beta\lambda} = \psi_{1i\beta\lambda}(y_i, \xi) = \frac{\partial^2}{\partial\beta^T\partial\lambda}\,\psi_{1i}(y_i, \xi), \text{ etc},$$

where, for example, $\frac{\partial^2}{\partial\beta^T\partial\beta}\,\psi_{1i}(y_i, \xi)$ means that we first apply $\frac{\partial}{\partial\beta^T}$ to each element of the $p\times 1$ vector $\psi_{1i}(y_i, \xi)$ which results in a $p\times p$ matrix, then we apply $\frac{\partial}{\partial\beta}$ to each element of the resulting matrix. Equivalently, we apply $\frac{\partial^2}{\partial\beta^T\partial\beta}$ to each element of $\psi_{1i}(y_i, \xi)$.

Hence, $\psi_{1i\beta\beta}$ is a $p^2\times p$ matrix; $\psi_{1i\beta\lambda}$, $\psi_{1i\beta\sigma}$, $\psi_{2i\beta\beta}$ and $\psi_{3i\beta\beta}$ are $p\times p$ matrices; $\psi_{1i\lambda\lambda}$, $\psi_{1i\sigma\sigma}$, and $\psi_{1i\lambda\sigma}$ are $p\times 1$ column vectors; $\psi_{2i\beta\lambda}$, $\psi_{2i\beta\sigma}$, $\psi_{3i\beta\lambda}$, and $\psi_{3i\beta\sigma}$, are $1\times p$ row vectors; and $\psi_{2i\lambda\lambda}$, $\psi_{2i\sigma\sigma}$, $\psi_{2i\lambda\sigma}$, $\psi_{3i\lambda\lambda}$, $\psi_{3i\sigma\sigma}$, and $\psi_{3i\lambda\sigma}$ are scalars.

Let $\psi_{1\beta\beta}$, $\psi_{1\beta\lambda}$, $\psi_{1\beta\sigma}$, $\psi_{2\beta\beta}$, $\psi_{2\beta\lambda}$, $\psi_{2\beta\sigma}$, $\psi_{3\beta\beta}$, $\psi_{3\beta\lambda}$, $\psi_{3\beta\sigma}$, $\psi_{1\lambda\lambda}$, $\psi_{1\sigma\sigma}$, $\psi_{1\lambda\sigma}$, $\psi_{2\lambda\lambda}$, $\psi_{2\sigma\sigma}$, $\psi_{2\lambda\sigma}$, $\psi_{3\lambda\lambda}$, $\psi_{3\sigma\sigma}$, and $\psi_{3\lambda\sigma}$ represent the averages, e.g.,

$$\psi_{1\beta\beta} = \frac{1}{n}\sum_{i=1}^{n}\psi_{1i\beta\beta}, \text{ and } \psi_{1\lambda\lambda} = \frac{1}{n}\sum_{i=1}^{n}\psi_{1i\lambda\lambda}, \text{ etc.}$$

Now, we replace assumption A2 of Section 3.3 by

Assumption **A2′.** $\psi_\xi = A_\xi + n^{1/2}\chi_\xi$ where χ_ξ is $O_p(1)$ and partitioned

$$\chi_\xi = \begin{bmatrix} \chi_{1\beta} & \chi_{1\lambda} & \chi_{1\sigma} \\ \chi_{2\beta} & \chi_{2\lambda} & \chi_{2\sigma} \\ \chi_{3\beta} & \chi_{3\lambda} & \chi_{3\sigma} \end{bmatrix}.$$

We assume further that

Assumption **A5.** The quantities $\psi_{1\beta\beta}$, $\psi_{1\beta\lambda}$, $\psi_{1\beta\sigma}$, $\psi_{2\beta\beta}$, etc., are $O_p(1)$.

Second-order expansion for $\hat{\beta}(\hat{\lambda})$

Before we get to the main theorems we need the following Lemma. Let g_1 and g_3 be defined in Section 3.3 and define

$$g_1^* = B_{1\beta}[\chi_{1\beta}g_1 + \frac{1}{2}(I_p \otimes g_1^T)\psi_{1\beta\beta}\, g_1],$$

$$H_{1\lambda} = B_{1\beta}[\chi_{1\lambda} + \chi_{1\beta}B_{1\beta}A_{1\lambda} + (I_p \otimes g_1^T)\psi_{1\beta\beta}B_{1\beta}A_{1\lambda} + \psi_{1\beta\lambda}\, g_1],$$

$$H_{1\sigma} = B_{1\beta}[\chi_{1\sigma} + \chi_{1\beta}B_{1\beta}A_{1\sigma} + (I_p \otimes g_1^T)\psi_{1\beta\beta}B_{1\beta}A_{1\sigma} + \psi_{1\beta\sigma}\, g_1],$$

$$H_{1\lambda\lambda} = B_{1\beta}\{\frac{1}{2}[I_p \otimes (B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}B_{1\beta}A_{1\lambda} + \frac{1}{2}\psi_{1\lambda\lambda} + \psi_{1\beta\lambda}B_{1\beta}A_{1\lambda}\},$$

$$H_{1\sigma\sigma} = B_{1\beta}\{\frac{1}{2}[I_p \otimes (B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}B_{1\beta}A_{1\sigma} + \frac{1}{2}\psi_{1\sigma\sigma} + \psi_{1\beta\lambda}B_{1\beta}A_{1\sigma}\},$$

$$H_{1\lambda\sigma} = B_{1\beta}\{[I_p \otimes (B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}B_{1\beta}A_{1\sigma} + \psi_{1\beta\lambda}B_{1\beta}A_{1\sigma} + \psi_{1\beta\sigma}B_{1\beta}A_{1\lambda} + \psi_{1\lambda\sigma}\},$$

$$g_3^* = B_{3\sigma}(\chi_{3\sigma}g_3 + \frac{1}{2}\psi_{3\sigma\sigma}\, g_3^2),$$

$$H_{3\beta} = B_{3\sigma}(\chi_{3\beta} + \chi_{3\sigma}B_{3\sigma}A_{3\beta} + \psi_{3\sigma\sigma}\, g_3 B_{3\sigma}A_{3\beta} + \psi_{3\beta\sigma}\, g_3),$$

$$H_{3\lambda} = B_{3\sigma}(\chi_{3\lambda} + \chi_{3\sigma}B_{3\sigma}A_{3\lambda} + \psi_{3\sigma\sigma}\, g_3 B_{3\sigma}A_{3\lambda} + \psi_{3\lambda\sigma}\, g_3),$$

$$
\begin{aligned}
H_{3\beta\beta} &= B_{3\sigma}\left(\frac{1}{2}\psi_{3\beta\beta} + \frac{1}{2}\psi_{3\sigma\sigma} B_{3\sigma}^2 A_{3\beta}^T A_{3\beta} + \psi_{3\sigma\beta} B_{3\sigma} A_{3\beta}\right), \\
H_{3\lambda\lambda} &= B_{3\sigma}\left(\frac{1}{2}\psi_{2\lambda\lambda} + \frac{1}{2}\psi_{3\sigma\sigma} B_{3\sigma}^2 A_{3\lambda}^2 + \psi_{3\lambda\sigma} B_{3\sigma} A_{3\lambda}\right), \\
H_{3\beta\lambda} &= B_{3\sigma}\left(\psi_{3\sigma\sigma} B_{3\sigma}^2 A_{3\beta} A_{3\lambda} + \psi_{3\beta\sigma} B_{3\sigma} A_{3\lambda} + \psi_{3\lambda\sigma} B_{3\sigma} A_{3\beta} + \psi_{3\beta\lambda}\right).
\end{aligned}
$$

So g_1^*, $H_{1\lambda}$, $H_{1\sigma}$, $H_{1\lambda\lambda}$, $H_{1\sigma\sigma}$, and $H_{1\lambda\sigma}$ are $p\times 1$ random column vectors; $H_{3\beta}$ and $H_{3\beta\lambda}$ are $1\times p$ random row vectors; $H_{3\beta\beta}$ is a $p\times p$ random matrix; and the others are random scalars. Note that the H-quantities such as $H_{3\beta}$ do not represent derivatives.

Lemma 4.1. If g_1 and g_3 are $O_p(1)$, then as $n \to \infty$, we have

$$
\begin{aligned}
\hat{\beta}(\hat{\lambda}) &= \beta(\lambda) + n^{-1/2} g_1 + B_{1\beta} A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta} A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + n^{-1} g_1^* \\
&\quad + n^{-1/2} H_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1/2} H_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + H_{1\lambda\lambda}(\hat{\lambda} - \lambda)^2 \\
&\quad + H_{1\sigma\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 + H_{1\lambda\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-3/2}). \qquad (4.1)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma}(\hat{\lambda}) &= \sigma(\lambda) + n^{-1/2} g_3 + B_{3\sigma} A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + B_{3\sigma} A_{3\lambda}(\hat{\lambda} - \lambda) + n^{-1} g_3^* \\
&\quad + n^{-1/2} H_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + n^{-1/2} H_{3\lambda}(\hat{\lambda} - \lambda) + H_{3\lambda\lambda}(\hat{\lambda} - \lambda)^2 \\
&\quad + (\hat{\beta}(\hat{\lambda}) - \beta(\lambda))^T H_{3\beta\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + H_{3\beta\lambda}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\lambda} - \lambda) \\
&\quad + O_p(n^{-3/2}). \qquad (4.2)
\end{aligned}
$$

where g_1^*, $H_{1\lambda}$, $H_{1\sigma}$, $H_{1\lambda\lambda}$, $H_{1\sigma\sigma}$, $H_{1\lambda\sigma}$, $H_{3\beta}$, $H_{3\beta\lambda}$, and $H_{3\beta\beta}$ are $O_p(1)$; and g_3^*, $H_{3\lambda}$, and $H_{3\lambda\lambda}$ are $O_p(1)$.

Proof. A second-order Taylor series expansion of ψ_{1i} around ξ combined with (3.10) gives (4.1), and a second-order Taylor series expansion of ψ_{3i} combined with (3.11) leads to (4.2). The detailed proofs involve tedious algebraic manipulation and hence are put in Appendix B.

Now define

$$
\begin{aligned}
D &= (1 - B_{3\sigma} A_{3\beta} B_{1\beta} A_{1\sigma})^{-1} \\
H_1 &= g_1^* + H_{1\sigma} D(g_3 + B_{3\sigma} A_{3\beta} g_1) + H_{1\sigma\sigma}[D(g_3 + B_{3\sigma} A_{3\beta} g_1)]^2 \\
&\quad + B_{1\beta} A_{1\sigma}[\, g_3^* + H_{3\beta} B(g_1 + B_{1\beta} A_{1\sigma} g_3)] + (g_1 + B_{1\beta} A_{1\sigma} g_3)^T B^T H_{3\beta\beta} \\
&\quad\; B(g_1 + B_{1\beta} A_{1\sigma} g_3); \\
H_2 &= H_{1\lambda} + B_{1\beta} A_{1\sigma}[H_{3\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) + H_{3\lambda}] \\
&\quad + B_{1\beta} A_{1\sigma} H_{3\beta\lambda} B(g_1 + B_{1\beta} A_{1\sigma} g_3) + 2B_{1\beta} A_{1\sigma}[B(g_1 + B_{1\beta} A_{1\sigma} g_3)]^T \\
&\quad \bullet H_{3\beta\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) + H_{1\sigma} D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda}) \\
&\quad + 2H_{1\sigma\sigma} D^2 (g_3 + B_{3\sigma} A_{3\beta} g_1) B_{3\sigma} (A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda}) \\
&\quad + H_{1\lambda\sigma} D(g_3 + B_{3\sigma} A_{3\beta} g_1); \\
H_3 &= H_{1\lambda\lambda} + B_{1\beta} A_{1\sigma} H_{3\lambda\lambda} + B_{1\beta} A_{1\sigma}[B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})]^T H_{3\beta\beta} [\cdots] \\
&\quad + B_{1\beta} A_{1\sigma} H_{3\beta\lambda} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) \\
&\quad + H_{1\sigma\sigma}[D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})]^2 \\
&\quad + H_{1\lambda\sigma} D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda}).
\end{aligned}
$$

So D is a fixed scalar, H_1, H_2, and H_3 are $p \times 1$ random vectors.

Theorem 4.1. Assume that g_1 and g_3 are $O_p(1)$. Then as $n \to \infty$, we have the following asymptotic expansion for the $\hat{\beta}(\hat{\lambda})$

$$
\begin{aligned}
\hat{\beta}(\hat{\lambda}) &= \beta(\lambda) + n^{-1/2} B g_1 + n^{-1/2} B B_{1\beta} A_{1\sigma} g_3 + B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda) \\
&\quad + n^{-1} B H_1 + n^{-1/2} B H_2(\hat{\lambda} - \lambda) + B H_3(\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2}),
\end{aligned}
\tag{4.3}
$$

where H_1, H_2, and H_3 are all of order $O_p(1)$.

Proof. Analogous to Theorem 3.1, we have

$$
\begin{aligned}
\hat{\sigma}(\hat{\lambda}) &= \sigma(\lambda) + n^{-1/2} D g_3 + D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})(\hat{\lambda} - \lambda) \\
&\quad + n^{-1/2} D B_{3\sigma} A_{3\beta} g_1 + O_p(n^{-1}).
\end{aligned}
\tag{4.4}
$$

Now based on (4.1), (4.2), (4.4) and Theorem (3.1), some substitutions and some algebra lead to (4.3). The detailed proof is given in Appendix B.

Corollary 4.1. Under the framework of Theorem 4.1, if λ is known then $\hat{\beta}(\lambda)$ has the asymptotic expansion

$$\hat{\beta}(\lambda) = \beta(\lambda) + n^{-1/2}Bg_1 + n^{-1/2}BB_{1\beta}A_{1\sigma}g_3 + n^{-1}BH_1 + O_p(n^{-3/2}). \qquad (4.5)$$

Proof. Simply equate $(\hat{\lambda} - \lambda)$ to zero in (4.3).

Corollary 4.2. Under the conditions of Theorem 4.1, $\beta_u(\hat{\lambda})$ has the asymptotic expansion

$$\begin{aligned}\beta_u(\hat{\lambda}) &= \beta(\lambda) + BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})(\hat{\lambda} - \lambda) + n^{-1}B\mathrm{E}(H_1) \\ &\quad + n^{-1/2}B\mathrm{E}(H_2)(\hat{\lambda} - \lambda) + B\mathrm{E}(H_3)(\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2}). \qquad (4.6)\end{aligned}$$

Proof. Take the expectation of both sides of (4.3) pretending that $\hat{\lambda}$ is fixed.

Corollary 4.3. From Theorem 4.1 and it's corollaries we have

$$\begin{aligned}\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda}) &= \hat{\beta}(\lambda) - \beta(\lambda) - n^{-1}B\mathrm{E}(H_1) + n^{-1/2}B(H_2 - \mathrm{E}H_2)(\hat{\lambda} - \lambda) \\ &\quad + B(H_3 - \mathrm{E}H_3)(\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2}), \qquad (4.7)\end{aligned}$$

$$\begin{aligned}\hat{\beta}(\hat{\lambda}) - \beta(\lambda) &= \hat{\beta}(\lambda) - \beta(\lambda) + BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})(\hat{\lambda} - \lambda) \\ &\quad + n^{-1/2}BH_2(\hat{\lambda} - \lambda) + BH_3(\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2}) \qquad (4.8)\end{aligned}$$

Proof. Straightforward from Theorem 4.1, Corollary 4.1, and Corollary 4.2.

Second-order expansion for $[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]/\hat{\sigma}(\hat{\lambda})$

Here we study the effect of estimating λ on the pivotal quantity $[\hat{\beta}(\hat{\lambda})-\beta_u(\hat{\lambda})]/\hat{\sigma}(\hat{\lambda})$ used to obtain tests and confidence regions. To this end we reparameterize the model

$$h(y, \lambda) = X\beta(\lambda) + \frac{1}{\tau(\lambda)} e(\lambda), \quad \text{where} \quad \tau(\lambda) = 1/\sigma(\lambda), \tag{4.9}$$

and change the definition of the ψ_i function accordingly. Now write $\xi(\lambda) = (\beta^T(\lambda), \lambda, \tau(\lambda))^T$, and $\psi_i = \psi_i(y_i, \xi(\lambda))$. All expansions obtained in the earlier part maintain the same form except one change in the notation: $\sigma(\lambda)$ is replaced by $\tau(\lambda)$. From now on we will use the expansions in the previous part assuming the notation has already been changed.

Theorem 4.2. Under the conditions of Theorem 4.1, the Box-Cox λ-unknown quantity $[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]/\hat{\sigma}(\hat{\lambda})$ has the following asymptotic expansion as $n \to \infty$,

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})} = \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} + n^{-1}d(\hat{\lambda}) + O_p(n^{-3/2}), \tag{4.10}$$

where

$$\begin{aligned} d(\hat{\lambda}) &= -\tau(\lambda)\, B\mathrm{E}(H_1) + n^{1/2} D B_{3\tau}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda}) B\big[(g_1 + B_{1\beta} A_{1\tau}\, g_3) \\ &\quad + \tau(\lambda)(H_2 - \mathrm{E}H_2)\big](\hat{\lambda} - \lambda) + n\tau(\lambda)\, B(H_3 - \mathrm{E}H_3)(\hat{\lambda} - \lambda)^2 \\ &= O_p(1). \end{aligned}$$

Proof. From (4.4) we have

$$\begin{aligned} \hat{\tau}(\hat{\lambda}) - \tau(\lambda) &= n^{-1/2} D g_3 + D B_{3\tau}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})(\hat{\lambda} - \lambda) \\ &\quad + n^{-1/2} D B_{3\tau} A_{3\beta}\, g_1 + O_p(n^{-1}), \end{aligned}$$

hence the corresponding λ-known expansion for $\hat{\tau}(\lambda)$ is

$$\hat{\tau}(\lambda) - \tau(\lambda) = n^{-1/2}Dg_3 + n^{-1/2}DB_{3\tau}A_{3\beta}\,g_1 + O_p(n^{-1}),$$

so that

$$\hat{\tau}(\hat{\lambda}) = \hat{\tau}(\lambda) + DB_{3\tau}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda) + O_p(n^{-1}). \qquad (4.11)$$

Now multiplying (4.7) by (4.11) and collecting the terms of order $O_p(n^{-3/2})$, we have

$$\begin{aligned}
&[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]\,\hat{\tau}(\hat{\lambda}) \\
&\quad = [\hat{\beta}(\lambda) - \beta(\lambda)]\,\hat{\tau}(\lambda) - n^{-1}\hat{\tau}(\lambda)B\mathrm{E}(H_1) + n^{-1/2}\hat{\tau}(\lambda)B[H_2 - \mathrm{E}(H_2)](\hat{\lambda} - \lambda) \\
&\quad + \hat{\tau}(\lambda)B[H_3 - \mathrm{E}(H_3)](\hat{\lambda} - \lambda)^2 \\
&\quad + [\hat{\beta}(\lambda) - \beta(\lambda)]DB_{3\tau}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda) + O_p(n^{-3/2});
\end{aligned}$$

Using (4.5) for the last term gives

$$\begin{aligned}
&[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]\,\hat{\tau}(\hat{\lambda}) \\
&\quad = [\hat{\beta}(\lambda) - \beta(\lambda)]\,\hat{\tau}(\lambda) - n^{-1}\hat{\tau}(\lambda)B\mathrm{E}(H_1) + n^{-1/2}\hat{\tau}(\lambda)B[H_2 - \mathrm{E}(H_2)](\hat{\lambda} - \lambda) \\
&\quad + \hat{\tau}(\lambda)B[H_3 - \mathrm{E}(H_3)](\hat{\lambda} - \lambda)^2 \\
&\quad + n^{-1/2}B(g_1 + B_{1\beta}A_{1\tau}g_3)DB_{3\tau}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda) + O_p(n^{-3/2}).
\end{aligned}$$

Substituting $\hat{\tau}(\hat{\lambda}) = 1/\hat{\sigma}(\hat{\lambda}))$ and $\hat{\tau}(\lambda) = 1/\hat{\sigma}(\lambda))$ into above expression, and using the fact that $\hat{\sigma}(\lambda)$ is root-n consistent gives the result (4.10).

We have obtained a major result (Theorem 4.2) concerning the second-order asymptotic expansion for the Box-Cox λ-unknown quantity. It is unified in the sense that it holds for any smooth ψ_i function. The theoretical importance of Theorem 4.2 is that, like the general conclusion from Theorem 3.2, it also concludes generally that there is no first-order ($O_p(n^{-1/2})$) effecting term, and it adds more, giving a general expression of the second-order ($O_p(n^{-1})$) effecting term. Hence using the λ-known quantity together with $n^{-1}d(\hat{\lambda})$ to approximate the λ-unknown quantity (1.6) results in an error of order $O_p(n^{-3/2})$, while using the λ-known quantity only gives an error of order $O_p(n^{-1})$. This second-order expansion allows for the detailed study of the effects for small sample case for a given ψ_i function.

It should be pointed out that Theorem 4.2 depends only on the root-n consistency of $\hat{\lambda}$ to λ, not on the way in which $\hat{\lambda}$ is defined. In other words, $\hat{\lambda}$ could be a least square estimator or an estimator obtained from a preliminary data set while $\hat{\beta}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$ are M-estimators for the given $\hat{\lambda}$.

Since Theorem 4.2 is an expansion in terms of $\hat{\lambda}$, it can be used directly to study the conditional behavior of the Box-Cox λ-unknown quantity if $\hat{\lambda}$ is independent of y. The unconditional behavior of the Box-Cox λ-unknown quantity can be studied by a full second-order asymptotic expansion which can be obtained using Theorem 4.2 together with a full first-order expansion for $\hat{\lambda}$.

First-order expansion of $\hat{\lambda}-\lambda$

To study the behavior of the Box-Cox λ-unknown quantity unconditionally, a full asymptotic expansion in terms of the error vector $e(\lambda)$ is needed. From Theorem 4.2 we see that it suffices to have a first-order full asymptotic expansion for $\hat{\lambda}$.

Suppose that $\hat{\lambda}$ is defined by (3.6) jointly with $\hat{\beta}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})$, and define

$$g_2 = n^{1/2} B_{2\lambda} \psi_2, \quad B_{2\lambda} = -A_{2\lambda}^{-1},$$

and

$$c_0 = [A_{2\lambda} + A_{2\beta} BB_{1\beta}(A_{1\lambda} + A_{1\tau} B_{3\tau} A_{3\lambda}) + A_{2\tau} DB_{3\tau}(A_{3\beta} B_{1\beta} A_{1\lambda} + A_{3\lambda})].$$

Theorem 4.3. Assume that g_1, g_2 and g_3 are $O_p(1)$. As $n \to \infty$, $\hat{\lambda}$ has the first-order asymptotic expansion

$$\hat{\lambda} - \lambda = n^{-1/2} M_1 g_1 + n^{-1/2} M_2 g_2 + n^{-1/2} M_3 g_3 + O_p(n^{-1}), \tag{4.12}$$

where

$$M_1 = -\frac{1}{c_0}(A_{2\beta}B + A_{2\tau}DB_{3\tau}A_{3\beta}) = O(1),$$
$$M_2 = \frac{1}{c_0}A_{2\lambda} = O(1),$$

and

$$M_3 = -\frac{1}{c_0}(A_{2\tau}D + A_{2\beta}BB_{1\beta}A_{1\tau}) = O(1).$$

Proof. A first-order Taylor series expansion of ψ_{2i} gives

$$0 = \frac{1}{n}\sum_{i=1}^{n}\{\psi_{2i} + \psi_{2i\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \psi_{2i\lambda}(\hat{\lambda} - \lambda) + \psi_{2i\tau}(\hat{\tau}(\hat{\lambda}) - \tau(\lambda))\} + O_p(n^{-1}),$$
$$= \psi_2 + \psi_{2\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \psi_{2\lambda}(\hat{\lambda} - \lambda) + \psi_{2\tau}(\hat{\tau}(\hat{\lambda}) - \tau(\lambda)) + O_p(n^{-1}).$$

By Assumption A2 we have

$$-A_{2\lambda}(\hat{\lambda} - \lambda) = \psi_2 + A_{2\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + A_{2\tau}(\hat{\tau}(\hat{\lambda}) - \tau(\lambda)) + O_p(n^{-1}). \quad (4.13)$$

Substituting (3.9) and (4.4) into (4.13), we have

$$\begin{aligned}
&-A_{2\lambda}(\hat{\lambda} - \lambda) \\
&\quad = \psi_2 + A_{2\beta}[n^{1/2}Bg_1 + BB_{1\beta}(A_{1\lambda} + A_{1\tau}B_{3\tau}A_{3\lambda})(\hat{\lambda} - \lambda) \\
&\qquad + n^{-1/2}BB_{1\beta}A_{1\tau}g_3 + O_p(n^{-1})] + A_{2\tau}[n^{-1/2}Dg_3 + DB_{3\tau}(A_{3\lambda} \\
&\qquad + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda) + n^{-1/2}DB_{3\tau}A_{3\beta}g_1 + O_p(n^{-1})] + O_p(n^{-1}) \\
&\quad = \psi_2 + n^{1/2}A_{2\beta}B\,g_1 + n^{-1/2}A_{2\beta}BB_{1\beta}A_{1\tau}\,g_3 \\
&\qquad + A_{2\beta}BB_{1\beta}(A_{1\lambda} + A_{1\tau}B_{3\tau}A_{3\lambda})(\hat{\lambda} - \lambda) + n^{-1/2}A_{2\tau}Dg_3 \\
&\qquad + n^{-1/2}A_{2\tau}DB_{3\tau}A_{3\beta}g_1 + A_{2\tau}DB_{3\tau}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda) + O_p(n^{-1}).
\end{aligned}$$

Collecting the terms with $(\hat{\lambda} - \lambda)$, we obtain

$$[A_{2\lambda} + A_{2\beta}BB_{1\beta}(A_{1\lambda} + A_{1\tau}B_{3\tau}A_{3\lambda}) + A_{2\tau}DB_{3\tau}(A_{3\beta}B_{1\beta}A_{1\lambda} + A_{3\lambda})](\hat{\lambda} - \lambda)$$
$$= -\psi_2 - n^{1/2}(A_{2\beta}B + A_{2\tau}DB_{3\tau}A_{3\beta})g_1 - n^{-1/2}(A_{2\tau}D + A_{2\beta}BB_{1\beta}A_{1\tau})g_3 + O_p(n^{-1}),$$

completing the proof.

Note that if the ψ function is specified we can use Theorem 4.3 to study the behavior of $\hat{\lambda}$ for large n. If we are interested in the behavior of $\hat{\lambda}$ for small-to-moderate n, a higher order expansion may be required.

Expansion for the Variance

Following the results in Section 4.2, we can easily obtain asymptotic expansions for the variance of the λ-unknown quantity.

Proposition 4.1. Under the assumptions of Theorem 4.2 and 4.3, we have

$$\mathrm{COV}\left[\frac{\hat{\beta}(\hat{\lambda})-\beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}\right]=\mathrm{COV}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)}\right]+n^{-3/2}\tau(\lambda)\,DB_{3\tau}(A_{3\lambda}+A_{3\beta}B_{1\beta}A_{1\lambda})\bullet$$
$$\mathrm{E}\left\{B(g_1+B_{1\beta}A_{1\tau}g_3)\left[(g_1+B_{1\beta}A_{1\tau}g_3)^T+\tau(\lambda)(H_2-\mathrm{E}H_2)^T\right]B^T(M_1g_1+M_2g_2+M_3g_3)\right\}$$
$$+n^{-3/2}\tau^2(\lambda)\left[B(g_1+B_{1\beta}A_{1\tau}g_3)(H_3-\mathrm{E}H_3)^TB^T(M_1g_1+M_2g_2+M_3g_3)^2\right]+O(n^{-2}).$$

Proof. From (4.10), we have

$$\mathrm{COV}\left[\frac{\hat{\beta}(\hat{\lambda})-\beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}\right]$$
$$=\mathrm{COV}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)}\right]+2\,\mathrm{COV}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)},\ n^{-1}d(\hat{\lambda})\right]+O_p(n^{-2}).$$

From (3.5) we have

$$\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)}=n^{-1/2}\tau(\lambda)B(g_1+B_{1\beta}A_{1\tau}g_3)+O_p(n^{-1}),$$

which implies that

$$\mathrm{E}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)}\right]=O(n^{-1}),$$

and

$$\mathrm{Cov}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)},\ n^{-1}d(\hat{\lambda})\right]=\mathrm{E}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)}\bullet\frac{d^T(\hat{\lambda})}{n}\right]+\ O_p(n^{-2}).$$

Now combination of (4.10) and (4.12) gives

$$\begin{aligned}
d(\hat{\lambda}) &= -\ \tau(\lambda)B\mathrm{E}(H_1) +\ n^{1/2}DB_{3\tau}(A_{3\lambda}+A_{3\beta}B_{1\beta}A_{1\lambda})B\big[(g_1+B_{1\beta}A_{1\tau}g_3)\\
&\quad +\ \tau(\lambda)(H_2-\mathrm{E}H_2)\big](\hat{\lambda}-\lambda)+n\tau(\lambda)B(H_3-\mathrm{E}H_3)(\hat{\lambda}-\lambda)^2\\
&= -\ \tau(\lambda)B\mathrm{E}(H_1)+DB_{3\tau}(A_{3\lambda}+A_{3\beta}B_{1\beta}A_{1\lambda})B\big[(g_1+B_{1\beta}A_{1\tau}g_3)\\
&\quad +\ \tau(\lambda)(H_2-\mathrm{E}H_2)\big](M_1g_1+M_2g_2+M_3g_3)\\
&\quad +\ \tau(\lambda)B(H_3-\mathrm{E}H_3)(M_1g_1+M_2g_2+M_3g_3)^2+O_p(^{-1/2}),
\end{aligned}$$

hence

$$\mathrm{E}\left[\frac{\hat{\beta}(\lambda)-\beta(\lambda)}{\hat{\sigma}(\lambda)}\bullet\frac{d^T(\hat{\lambda})}{n}\right]$$

$$\begin{aligned}
&= n^{-3/2}\tau(\lambda))DB_{3\tau}(A_{3\lambda}+A_{3\beta}B_{1\beta}A_{1\lambda})B\times\\
&\mathrm{E}\Big\{(g_1+B_{1\beta}A_{1\tau}g_3)\big[(g_1+B_{1\beta}A_{1\tau}g_3)^T+\tau(\lambda)(H_2-\mathrm{E}H_2)^T\big]B^T(M_1g_1+M_2g_2+M_3g_3)\Big\}\\
&+\frac{n^{-3/2}}{\sigma^2(\lambda)}\big[(Bg_1+BB_{1\beta}A_{1\tau}g_3)(H_3-\mathrm{E}H_3)^TB^T(M_1g_1+M_2g_2+M_3g_3)^2\big]+O(n^{-2}),
\end{aligned}$$

completing the proof.

CHAPTER 5
SECOND-ORDER EXPANSIONS FOR THE BOX-COX ESTIMATORS

5.1. Introduction

In Chapter 4, we derived expansions for M-estimators and related pivotal quantities for a general smooth ψ_i function. More detailed results are possible when the ψ_i function is specified. We consider the reparameterized model (4.9) with ψ_i corresponding to the Box-Cox estimator, i.e., the estimator $\hat{\xi}(\hat{\lambda})$ of $\xi(\lambda) = (\beta^T(\lambda), \lambda, \tau(\lambda))^T$ defined by

$$\psi_i = \begin{cases} \psi_{1i}(y_i, \xi(\lambda)) = -\tau^2(\lambda)x_i[h(y_i, \lambda) - x_i^T\beta(\lambda)], \\ \psi_{2i}(y_i, \xi(\lambda)) = \tau^2(\lambda)[h(y_i, \lambda) - x_i^T\beta(\lambda)]h_\lambda(y_i, \lambda) - \dfrac{h_{y\lambda}(y_i, \lambda)}{h_y(y_i, \lambda)}, \\ \psi_{3i}(y_i, \xi(\lambda)) = -\dfrac{1}{\tau(\lambda)} + \tau(\lambda)[h(y_i, \lambda) - x_i^T\beta(\lambda)]^2. \end{cases} \tag{5.1}$$

For h defined by (1.1), the last term of ψ_{2i} is $\log y_i$.

In Section 5.2 we evaluate the expansions of Chapter 4 using (5.1). In Section 5.3 we make some approximations to the function $h(y, l)$ for fixed l in terms of $z = h(y, \lambda) = X\beta(\lambda) + (1/\tau(\lambda))e(\lambda)$, and use this approximation to work out the approximate asymptotic expansions in terms of $\sigma(\lambda)$ and $e(\lambda)$. The approximate asymptotic expansions enable us to draw some further conclusions concerning the effect of estimating transformation on the λ-unknown quantity (1.6) under different structures of the model and different values of $\sigma(\lambda)$. We measure this effect by the magnitude of the second-order term in the expansion. First for certain structured models with small $\sigma(\lambda)$ the second-order term is small hence the approximation of (1.5) to (1.6) should be very good. In unstructured models with more than one mean the effect is small if the means are far apart and $\sigma(\lambda)$ is small, but if the means are close together the effect could be large when $\sigma(\lambda)$ is small. In unstructured models with only one mean (one-sample) the effect is large when $\sigma(\lambda)$ is small. To provide empirical evidences to back up the theoretical results in Section 5.3, a simulation study was performed and the results are given in Section 5.4. A method for checking the adequacy of the second-order expansion and related simulation are described in Section 5.4. The latter results support the use of the second-order expansion. In Section 5.5, we discuss the consequences of the theorems and results obtained and suggest some directions for further research.

5.2. Second-order Expansions Related to the Box-Cox Estimators

We evaluate the expansions of Chapter 4 assuming that the ψ_i function defined by (5.1) satisfies the Assumptions A1-A5 of Chapter 4. Hernandez and Johnson (1980) showed that for one-sample case the assumption A1 holds, in particular, the Box-Cox estimator $\hat{\xi}(\hat{\lambda})$ is strongly consistent and has an asymptotic normal distribution. Hernandez (1978) showed that the condition A4 holds when y_i has lognormal, Gamma, Weibull, inverse Gaussian, and Pareto distributions. There might be some problem with the root-n consistency of $\hat{\xi}(\hat{\lambda})$ to $\xi(\lambda)$ but generally $\hat{\xi}(\hat{\lambda})$ will converge to a limit $\xi^*=(\beta^*,\lambda^*,\sigma^*)$. If the latter is true our expansions remain valid if we replace $\xi(\lambda)$ in the expansions by ξ^* and the λ-known quantity by $\{\hat{\beta}(\hat{\lambda})-\beta^*\}/\hat{\sigma}(\hat{\lambda})$, provided that the Assumptions A1-A5 are true corresponding to ξ^*. For more information on the asymptotic properties of the Box-Cox estimators, see Hinkley (1975), Taylor (1985a) and Taylor (1985b).

Computation of the quantities defined in Chapter 4

Write $e = e(\lambda)$. For the ψ_i function defined in (5.1) with the Box-Cox power transformation we have:

$$\begin{aligned}
\psi_1 &= -\frac{\tau(\lambda)}{n} X^T e, \\
\psi_2 &= n^{-1}[\tau(\lambda) e^T h_\lambda(y, \lambda) - \sum_{i=1}^{n} \log y_i], \\
\psi_3 &= -\frac{1}{\tau(\lambda)} + \frac{1}{n\tau(\lambda)} e^T e.
\end{aligned}$$

To avoid confusion we note that when calculating partial derivatives of ψ_{ki} with respect to λ we treat $\beta(\lambda)$ and $\sigma(\lambda)$ as constants since they are just the other parameters to be jointly estimated. Now the first-order partial derivatives of ψ_{1i}, ψ_{2i}, and ψ_{3i} are

$$\begin{aligned}
\psi_{1i\beta} &= \tau^2(\lambda) x_i x_i^T, \\
\psi_{1i\lambda} &= -\tau^2(\lambda) x_i h_\lambda(y_i, \lambda), \\
\psi_{1i\tau} &= -2\tau(\lambda) x_i [h(y_i, \lambda) - x_i^T \beta(\lambda)], \\
\psi_{2i\beta} &= -\tau^2(\lambda) x_i^T h_\lambda(y_i, \lambda) \\
&= \psi_{1i\lambda}^T, \\
\psi_{2i\lambda} &= \tau^2(\lambda)[h_\lambda^2(y_i, \lambda) + (h(y_i, \lambda) - x_i^T \beta(\lambda)) h_{\lambda\lambda}(y_i, \lambda)], \\
\psi_{2i\tau} &= 2\tau(\lambda)[h(y_i, \lambda) - x_i^T \beta(\lambda)] h_\lambda(y_i, \lambda), \\
\psi_{3i\beta} &= -2\tau(\lambda) x_i^T [h(y_i, \lambda) - x_i^T \beta(\lambda)] \\
&= \psi_{1i\tau}^T, \\
\psi_{3i\lambda} &= 2\tau(\lambda)[h(y_i, \lambda) - x_i^T \beta(\lambda)] h_\lambda(y_i, \lambda) \\
&= \psi_{2i\tau}, \\
\psi_{3i\tau} &= \frac{1}{\tau^2(\lambda)} + [h(y_i, \lambda) - x_i^T \beta(\lambda)]^2.
\end{aligned}$$

The elements of $\psi_\xi = \frac{1}{n}\sum_{i=1}^{n} \psi_{i\xi}$ are

$$\psi_{1\beta} = \frac{\tau^2(\lambda)}{n} X^T X,$$

$$\psi_{1\lambda} = -\frac{\tau^2(\lambda)}{n} X^T h_\lambda(y, \lambda),$$

$$\psi_{1\tau} = -\frac{2\tau(\lambda)}{n} X^T [h(y, \lambda) - X\beta(\lambda)],$$

$$\psi_{2\beta} = -\frac{\tau^2(\lambda)}{n} h_\lambda^T(y, \lambda) X^T$$
$$= \psi_{1\lambda}^T,$$

$$\psi_{2\lambda} = \frac{\tau^2(\lambda)}{n} [h_\lambda^T(y, \lambda) h_\lambda(y, \lambda) + (h(y, \lambda) - X\beta(\lambda))^T h_{\lambda\lambda}(y, \lambda)],$$

$$\psi_{2\tau} = \frac{2\tau(\lambda)}{n} [h(y, \lambda) - X\beta(\lambda)]^T h_\lambda(y, \lambda),$$

$$\psi_{3\beta} = -\frac{2\tau(\lambda)}{n} [h(y, \lambda) - X\beta(\lambda)]^T X$$
$$= \psi_{1\tau}^T,$$

$$\psi_{3\lambda} = \frac{2\tau(\lambda)}{n} [h(y, \lambda) - X\beta(\lambda)]^T h_\lambda(y, \lambda)$$
$$= \psi_{2\tau},$$

$$\psi_{3\tau} = \frac{1}{\tau^2(\lambda)} + \frac{1}{n} [h(y, \lambda) - X\beta(\lambda)]^T [h(y, \lambda) - X\beta(\lambda)].$$

The elements of $A_\xi = \mathrm{E}\psi_\xi$ are

$$A_{1\beta} = \frac{\tau^2(\lambda)}{n} X^T X,$$

$$A_{1\lambda} = -\frac{\tau^2(\lambda)}{n} X^T \mathrm{E}[h_\lambda(y, \lambda)],$$

$$A_{1\tau} = -\frac{2\tau(\lambda)}{n} X^T \mathrm{E}[h(y, \lambda) - X\beta(\lambda)]$$
$$= 0_{p\times 1},$$

$$A_{2\beta} = -\frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda^T(y, \lambda)] X^T$$
$$= A_{1\lambda}^T,$$

$$A_{2\lambda} = \frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda^T(y, \lambda) h_\lambda(y, \lambda) + (h(y, \lambda) - X\beta(\lambda))^T h_{\lambda\lambda}(y, \lambda)],$$

$$A_{2\tau} = \frac{2\tau(\lambda)}{n} \mathrm{E}[(h(y, \lambda) - X\beta(\lambda))^T h_\lambda(y, \lambda)],$$

$$
\begin{aligned}
A_{3\beta} &= -\frac{2\tau(\lambda)}{n} \mathrm{E}[h(y,\lambda) - X\beta(\lambda)]^T X \\
&= A_{1\tau}^T \\
&= 0_{p\times 1}^T, \\
A_{3\lambda} &= \frac{2\tau(\lambda)}{n} \mathrm{E}[h(y,\lambda) - X\beta(\lambda)]^T h_\lambda(y,\lambda) \\
&= A_{2\tau}, \\
A_{3\tau} &= \frac{1}{\tau^2(\lambda)} + \frac{1}{n} \mathrm{E}[h(y,\lambda) - X\beta(\lambda)]^T [h(y,\lambda) - X\beta(\lambda)] \\
&= \frac{1}{\tau^2(\lambda)} + \frac{1}{n}\mathrm{E}[\frac{1}{\tau} e^T \frac{1}{\tau} e] \\
&= \frac{2}{\tau^2(\lambda)}.
\end{aligned}
$$

The elements of χ_ξ are

$$
\begin{aligned}
\chi_{1\beta} &= 0, \\
\chi_{1\lambda} &= \sqrt{n}(\psi_{1\lambda} - A_{1\lambda}) = -n^{-1/2}\tau^2(\lambda)X^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)], \\
\chi_{1\tau} &= \sqrt{n}(\psi_{1\tau} - A_{1\tau}) = -2n^{-1/2}\tau(\lambda)X^T [h(y,\lambda) - X\beta(\lambda)] \\
&\qquad = -2n^{-1/2}X^T e, \\
\chi_{2\beta} &= \sqrt{n}(\psi_{2\beta} - A_{2\beta}) = -n^{-1/2}\tau^2(\lambda)[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)]^T X, \\
\chi_{2\lambda} &= \sqrt{n}(\psi_{2\lambda} - A_{2\lambda}) = n^{-1/2}\tau^2(\lambda)\{h_\lambda^T(y,\lambda)h_\lambda(y,\lambda) + \frac{1}{\tau(\lambda)} e^T h_{\lambda\lambda}(y,\lambda) \\
&\qquad -\mathrm{E}[h_\lambda^T(y,\lambda)h_\lambda(y,\lambda)] + \mathrm{E}[\frac{1}{\tau(\lambda)} e^T h_{\lambda\lambda}(y,\lambda)\}, \\
\chi_{2\tau} &= \sqrt{n}(\psi_{2\tau} - A_{2\tau}) = 2n^{-1/2}\{e^T h_\lambda(y,\lambda) - \mathrm{E}[e^T h_\lambda(y,\lambda)]\}, \\
\chi_{3\beta} &= \sqrt{n}(\psi_{3\beta} - A_{3\beta}) = -2n^{-1/2}e^T X, \\
\chi_{3\lambda} &= \sqrt{n}(\psi_{3\lambda} - A_{3\lambda}) = 2n^{-1/2}\{e^T h_\lambda(y,\lambda) - \mathrm{E}[e^T h_\lambda(y,\lambda)]\} \\
&\qquad = \chi_{2\tau}, \\
\chi_{3\tau} &= \sqrt{n}(\psi_{3\tau} - A_{3\tau}) \\
&= \sqrt{n}\{\frac{1}{\tau^2(\lambda)} + \frac{1}{n}[h(y,\lambda) - X\beta(\lambda)]^T[h(y,\lambda) - X\beta(\lambda)] - \frac{2}{\tau^2(\lambda)}\} \\
&= \frac{\sqrt{n}}{\tau^2(\lambda)}(\frac{1}{n}e^T e - 1).
\end{aligned}
$$

The second-order derivatives of ψ_{1i}, ψ_{2i}, and ψ_{3i} are

$$\psi_{1i\beta\beta} = 0,$$
$$\psi_{1i\beta\lambda} = 0,$$
$$\psi_{1i\beta\tau} = 2\tau(\lambda)x_i x_i^T,$$
$$\psi_{1i\lambda\beta} = 0,$$
$$\psi_{1i\lambda\lambda} = -\tau^2(\lambda)x_i h_{\lambda\lambda}(y_i, \lambda),$$
$$\psi_{1i\lambda\tau} = -2\tau(\lambda)x_i h_\lambda(y_i, \lambda),$$
$$\psi_{1i\tau\beta} = 2\tau(\lambda)x_i x_i^T,$$
$$\psi_{1i\tau\lambda} = -2\tau(\lambda)x_i h_\lambda(y_i, \lambda),$$
$$\psi_{1i\tau\tau} = -2x_i [h(y_i, \lambda) - x_i^T\beta(\lambda)],$$
$$\psi_{2i\beta\beta} = 0,$$
$$\psi_{2i\beta\lambda} = -\tau^2(\lambda)x_i^T h_{\lambda\lambda}(y_i, \lambda),$$
$$\psi_{2i\beta\tau} = -2\tau(\lambda)x_i^T h_\lambda(y_i, \lambda),$$
$$\psi_{2i\lambda\beta} = -\tau^2(\lambda)x_i h_{\lambda\lambda}(y_i, \lambda),$$
$$\psi_{2i\lambda\lambda} = 3\tau^2(\lambda)h_\lambda(y_i, \lambda)h_{\lambda\lambda}(y_i, \lambda) - \tau^2(\lambda)[h(y_i, \lambda) - x_i^T\beta(\lambda)]h_{\lambda\lambda\lambda}(y_i, \lambda),$$
$$\psi_{2i\lambda\tau} = 2\tau(\lambda)\{h_\lambda^2(y_i, \lambda) + [h(y_i, \lambda) - x_i^T\beta(\lambda)]h_{\lambda\lambda}(y_i, \lambda)\},$$
$$\psi_{2i\tau\beta} = -2\tau(\lambda)x_i^T h_\lambda(y_i, \lambda),$$
$$\psi_{2i\tau\lambda} = 2\tau(\lambda)\{h_\lambda^2(y_i, \lambda) + [h(y_i, \lambda) - x_i^T\beta(\lambda)]h_{\lambda\lambda}(y_i, \lambda)\},$$
$$\psi_{2i\tau\tau} = 2[h(y_i, \lambda) - x_i^T\beta(\lambda)]h_\lambda(y_i, \lambda),$$
$$\psi_{3i\beta\beta} = 2\tau(\lambda)x_i x_i^T,$$
$$\psi_{3i\beta\lambda} = -2\tau(\lambda)x_i^T h_\lambda(y_i, \lambda),$$
$$\psi_{3i\beta\tau} = -2x_i^T[h(y_i, \lambda) - x_i^T\beta(\lambda)],$$
$$\psi_{3i\lambda\beta} = -2\tau(\lambda)x_i h_\lambda(y_i, \lambda),$$
$$\psi_{3i\lambda\lambda} = 2\tau(\lambda)\{x_i^T(y_i, \lambda) + [h(y_i, \lambda) - x_i^T\beta(\lambda)]h_{\lambda\lambda}(y_i, \lambda)\},$$
$$\psi_{3i\lambda\tau} = 2[h(y_i, \lambda) - x_i^T\beta(\lambda)]h_\lambda(y_i, \lambda),$$
$$\psi_{3i\tau\beta} = -2x_i[h(y_i, \lambda) - x_i^T\beta(\lambda)],$$

$$\psi_{3i\tau\lambda} = 2[h(y_i, \lambda) - x_i^T\beta(\lambda)]h_\lambda(y_i, \lambda),$$
$$\psi_{3i\tau\tau} = -\frac{2}{\tau^3(\lambda)}.$$

T' average of second derivatives, e.g., $\psi_{1\beta\beta} = \frac{1}{n}\sum_{i=1}^{n}\psi_{1\beta\beta}$ are

$$\psi_{1\beta\beta} = 0_{p^2\times p},$$
$$\psi_{1\beta\lambda} = 0_{p\times p},$$
$$\psi_{1\beta\tau} = 2\tau(\lambda)X^TX/n,$$
$$\psi_{1\lambda\beta} = 0_{p\times p},$$
$$\psi_{1\lambda\lambda} = -\tau^2(\lambda)X^Th_{\lambda\lambda}(y, \lambda)/n,$$
$$\psi_{1\lambda\tau} = -2\tau(\lambda)X^Th_\lambda(y, \lambda)/n,$$
$$\psi_{1\tau\beta} = 2\tau(\lambda)X^TX/n,$$
$$\psi_{1\tau\lambda} = -2\tau(\lambda)X^Th_\lambda(y, \lambda)/n,$$
$$\psi_{1\tau\tau} = -2X^T[h(y, \lambda) - X\beta(\lambda)]/n,$$
$$\psi_{2\beta\beta} = 0_{p\times p},$$
$$\psi_{2\beta\lambda} = -\tau^2(\lambda)h_{\lambda\lambda}^T(y, \lambda)X/n,$$
$$\psi_{2\beta\tau} = -2\tau(\lambda)h_\lambda^T(y, \lambda)X/n,$$
$$\psi_{2\lambda\beta} = -\tau^2(\lambda)X^Th(y, \lambda)/n,$$
$$\psi_{2\lambda\lambda} = \tau^2(\lambda)\{3h_\lambda^T(y, \lambda)h_{\lambda\lambda}(y, \lambda) - [h(y, \lambda) - X\beta(\lambda)]^Th_{\lambda\lambda\lambda}(y, \lambda)\}/n,$$
$$\psi_{2\lambda\tau} = 2\tau(\lambda)\{h_\lambda^T(y, \lambda)h_\lambda(y, \lambda) + [h(y, \lambda) - X\beta(\lambda)]^Th_{\lambda\lambda}(y, \lambda)\}/n,$$
$$\psi_{2\tau\beta} = -2\tau(\lambda)h_\lambda^T(y, \lambda)X/n,$$
$$\psi_{2\tau\lambda} = 2\tau(\lambda)\{h_\lambda^T(y, \lambda)h_\lambda(y, \lambda) + [h(y, \lambda) - X\beta(\lambda)]^Th_{\lambda\lambda}(y, \lambda)\}/n,$$
$$\psi_{2\tau\tau} = 2[h(y, \lambda) - X\beta(\lambda)]^Th_\lambda(y, \lambda)/n,$$
$$\psi_{3\beta\beta} = 2\tau(\lambda)X^TX/n,$$
$$\psi_{3\beta\lambda} = -2\tau(\lambda)h_\lambda^T(y, \lambda)X/n,$$
$$\psi_{3\beta\tau} = -2[h(y, \lambda) - X\beta(\lambda)]^TX/n,$$
$$\psi_{3\lambda\beta} = -2\tau(\lambda)X^Th_\lambda(y, \lambda)/n,$$

$$\psi_{3\lambda\lambda} = 2\tau(\lambda)\{h_\lambda^T(y,\lambda)h_\lambda(y,\lambda) + [h(y,\lambda) - X\beta(\lambda)]^T h_{\lambda\lambda}(y,\lambda)\}/n,$$

$$\psi_{3\lambda\tau} = 2[h(y,\lambda) - X\beta(\lambda)]^T h_\lambda(y,\lambda)/n,$$

$$\psi_{3\tau\beta} = -\,2X^T[h(y,\lambda) - X\beta(\lambda)]/n,$$

$$\psi_{3\tau\lambda} = 2[h(y,\lambda) - X\beta(\lambda)]^T h_\lambda(y,\lambda)/n,$$

$$\psi_{3\tau\tau} = -\,2/\tau^3(\lambda).$$

The quantities in the theorems of Chapter 4 are

$$\begin{aligned}
g_1 &= n^{1/2}B_{1\beta}\psi_1 \\
&= n^{1/2}[-(\frac{\tau^2(\lambda)}{n}X^TX)^{-1}][-\frac{\tau(\lambda)}{n}X^Te] \\
&= \frac{n^{1/2}}{\tau(\lambda)}(X^TX)^{-1}X^Te, \\
g_1^* &= B_{1\beta}[\chi_{1\beta}\, g_1 + \frac{1}{2}(I_p\otimes g_1^T)\psi_{1\beta\beta}\, g_1] \\
&= 0, \\
H_{1\lambda} &= B_{1\beta}[\chi_{1\lambda} + \chi_{1\beta}B_{1\beta}A_{1\lambda} + (I_p\otimes g_1^T)\psi_{1\beta\beta}\, B_{1\beta}A_{1\lambda} + \psi_{1\beta\lambda}\, g_1] \\
&= B_{1\beta}\chi_{1\lambda} \\
&= (-(\tau^2(\lambda)/n)X^TX)^{-1}\{-n^{-1/2}\tau^2(\lambda)X^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)]\} \\
&= n^{1/2}(X^TX)^{-1}X^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)], \\
H_{1\tau} &= B_{1\beta}[\chi_{1\tau} + \chi_{1\beta}B_{1\beta}A_{1\tau} + (I_p\otimes g_1^T)\psi_{1\beta\beta}\, B_{1\beta}A_{1\tau} + \psi_{1\beta\tau}\, g_1] \\
&= B_{1\beta}(\chi_{1\tau} + \psi_{1\beta\tau}\, g_1) \\
&= B_{1\beta}(-\,2n^{-1/2}X^Te + \frac{2\tau(\lambda)}{n}X^TX\,\frac{n^{1/2}}{\tau(\lambda)}(X^TX)^{-1}X^Te) \\
&= 0, \\
H_{1\lambda\lambda} &= B_{1\beta}\{\frac{1}{2}[I_p\otimes(B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}\, B_{1\beta}A_{1\lambda} + \frac{1}{2}\psi_{1\lambda\lambda} + \psi_{1\beta\lambda}B_{1\beta}A_{1\lambda}\} \\
&= \frac{1}{2}B_{1\beta}\,\psi_{1\lambda\lambda} \\
&= \frac{1}{2}[-(\tau^2(\lambda)/n)X^TX]^{-1}[-\,\tau^2(\lambda)X^Th_{\lambda\lambda}(y,\lambda)/n] \\
&= \frac{1}{2}(X^TX)^{-1}Xh_{\lambda\lambda}(y,\lambda), \\
H_{1\tau\tau} &= B_{1\beta}\{\frac{1}{2}[I_p\otimes(B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}\, B_{1\beta}A_{1\tau} + \frac{1}{2}\psi_{1\sigma\sigma} + \psi_{1\beta\lambda}B_{1\beta}A_{1\tau}\} \\
&= \frac{1}{2}B_{1\beta}\psi_{1\tau\tau}
\end{aligned}$$

$$= \frac{1}{2}\,[-\frac{n}{\tau^2(\lambda)}\,(X^TX)^{-1}][-\frac{2}{n}X^T(h(y,\lambda) - X^T\beta(\lambda))]$$

$$= \frac{1}{\tau^3(\lambda)}\,(X^TX)^{-1}X^Te,$$

$$H_{1\lambda\tau} = B_{1\beta}\{[I_p \otimes (B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}B_{1\beta}A_{1\tau} + \psi_{1\beta\lambda}B_{1\beta}A_{1\tau}$$

$$+ \psi_{1\beta\tau}B_{1\beta}A_{1\lambda} + \psi_{1\lambda\tau}\}$$

$$= B_{1\beta}(\psi_{1\beta\tau}B_{1\beta}A_{1\lambda} + \psi_{1\lambda\tau})$$

$$= (-\frac{\tau^2(\lambda)}{n}X^TX)^{-1}\{\frac{2\tau(\lambda)}{n}X^TX(-\frac{\tau^2(\lambda)}{n}X^TX)^{-1}[-\frac{\tau^2(\lambda)}{n}X^T\mathrm{E}h_\lambda(y,\lambda)]$$

$$-\frac{2\tau(\lambda)}{n}X^Th_\lambda(y,\lambda)]\}$$

$$= \frac{2}{\tau(\lambda)}\,(X^TX)^{-1}X^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)],$$

$$g_3 = n^{1/2}B_{3\tau}\psi_3$$

$$= n^{1/2}(-\frac{\tau^2(\lambda)}{2})(-\frac{1}{\tau(\lambda)} + \frac{1}{n\tau(\lambda)}\,e^Te)$$

$$= \frac{1}{2}\,n^{1/2}\tau(\lambda)\,(1 - \frac{1}{n}e^Te),$$

$$g_3^* = B_{3\tau}(\chi_{3\tau}g_3 + \frac{1}{2}\psi_{3\tau\tau}g_3^2)$$

$$= -\frac{\tau^2(\lambda)}{2}\,[\frac{n^{1/2}}{\tau^2(\lambda)}(\frac{1}{n}e^Te - 1)\frac{1}{2}\,n^{1/2}\tau(\lambda)(1 - \frac{1}{n}e^Te)$$

$$+\frac{1}{2}(-\frac{2}{\tau^3(\lambda)})[\frac{1}{2}n^{1/2}\tau(\lambda)(1 - \frac{1}{n}e^Te)]^2$$

$$= \frac{3}{8}\,n\tau(\lambda)(1 - \frac{1}{n}e^Te)^2,$$

$$H_{3\beta} = B_{3\tau}(\chi_{3\beta} + \chi_{3\tau}B_{3\tau}A_{3\beta} + \psi_{3\tau\tau}g_3B_{3\tau}A_{3\beta} + \psi_{3\beta\tau}g_3)$$

$$= B_{3\tau}(\chi_{3\beta} + \psi_{3\beta\tau}g_3),$$

$$H_{3\lambda} = B_{3\tau}(\chi_{3\lambda} + \chi_{3\tau}B_{3\tau}A_{3\lambda} + \psi_{3\tau\tau}g_3B_{3\tau}A_{3\lambda} + \psi_{3\lambda\tau}g_3),$$

$$H_{3\beta\beta} = B_{3\tau}(\frac{1}{2}\,\psi_{3\beta\beta} + \frac{1}{2}\,\psi_{3\tau\tau}B_{3\tau}^2A_{3\beta}^TA_{3\beta} + \psi_{3\tau\beta}B_{3\tau}A_{3\beta})$$

$$= \frac{1}{2}\,B_{2\tau}\psi_{3\beta\beta},$$

$$H_{3\lambda\lambda} = B_{3\tau}(\frac{1}{2}\,\psi_{2\lambda\lambda} + \frac{1}{2}\,\psi_{3\tau\tau}B_{3\tau}^2A_{3\lambda}^2 + \psi_{3\lambda\tau}B_{3\tau}A_{3\lambda}),$$

$$H_{3\beta\lambda} = B_{3\tau}(\psi_{3\tau\tau}B_{3\tau}^2A_{3\beta}A_{3\lambda} + \psi_{3\beta\tau}B_{3\tau}A_{3\lambda} + \psi_{3\lambda\tau}B_{3\tau}A_{3\beta} + \psi_{3\beta\lambda})$$

$$= B_{3\tau}(\psi_{3\beta\tau}B_{3\tau}A_{3\lambda} + \psi_{3\beta\lambda}),$$

$$B = (I_p - B_{1\beta}A_{1\tau}B_{3\tau}A_{3\beta})^{-1}$$

$$= I_p,$$

$$D = (1 - B_{3\tau} A_{3\beta} B_{1\beta} A_{1\sigma})^{-1}$$

$$= 1,$$

$$H_1 = g_1^* + H_{1\tau} D(g_3 + B_{3\tau} A_{3\beta} g_1) + H_{1\tau\tau}[D(g_3 + B_{3\tau} A_{3\beta} g_1)]^2$$

$$+ B_{1\beta} A_{1\tau}[g_3^* + H_{3\beta} B(g_1 + B_{1\beta} A_{1\tau} g_3) + (g_1 + B_{1\beta} A_{1\tau} g_3)^T B^T H_{3\beta\beta}$$

$$\bullet B(g_1 + B_{1\beta} A_{1\tau} g_3),$$

$$= H_{1\tau\tau} g_3^2$$

$$= \frac{1}{2} B_{1\beta} \psi_{1\tau\tau} g_3^2$$

$$= \frac{1}{2} [- \frac{n}{\tau^2(\lambda)} (X^T X)^{-1}][- \frac{2}{n} X^T (h(y, \lambda) - X^T \beta(\lambda))] g_3^2$$

$$= \frac{1}{\tau^3(\lambda)} (X^T X)^{-1} X^T e g_3^2),$$

$$H_2 = H_{1\lambda} + B_{1\beta} A_{1\sigma}[H_{3\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) + H_{3\lambda}]$$

$$+ B_{1\beta} A_{1\sigma} H_{3\beta\lambda} B(g_1 + B_{1\beta} A_{1\sigma} g_3) + 2 B_{1\beta} A_{1\sigma}[B(g_1 + B_{1\beta} A_{1\sigma} g_3)]^T$$

$$\bullet H_{3\beta\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) + H_{1\sigma} D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})$$

$$+ 2 H_{1\sigma\sigma} D^2 (g_3 + B_{3\sigma} A_{3\beta} g_1) B_{3\sigma} (A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})$$

$$+ H_{1\lambda\sigma} D(g_3 + B_{3\sigma} A_{3\beta} g_1)$$

$$= H_{1\lambda} + H_{1\lambda\tau} g_3,$$

$$H_3 = H_{1\lambda\lambda} + B_{1\beta} A_{1\sigma} H_{3\lambda\lambda} + B_{1\beta} A_{1\sigma}[B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})]^T H_{3\beta\beta} [\cdots]$$

$$+ B_{1\beta} A_{1\sigma} H_{3\beta\lambda} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})$$

$$+ H_{1\sigma\sigma}[D B_{3\sigma} (A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})]^2$$

$$+ H_{1\lambda\sigma} D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})$$

$$= H_{1\lambda\lambda} + H_{1\tau\tau}(B_{3\tau} A_{3\lambda})^2 + H_{1\lambda\tau} B_{3\tau} A_{3\lambda}.$$

Now we state and prove some results for the Box-Cox estimator based on the assumptions made in Chapter 4.

Lemma 5.1. Under the assumptions A1–A5 of Chapter 4 and the assumptions of $g_1 = O_p(1)$ and $g_3 = O_p(1)$, for the ψ_i function defined by (5.1), we have

i) H_1 is of order $O_p(n^{-1/2})$;

ii) $H_2 = H_{1\lambda} + O_p(n^{-1/2})$;

iii) $H_3 = H_{1\lambda\lambda} + O_p(n^{-1/2})$.

Proof. Since $H_1 = \frac{1}{\tau^3(\lambda)}(X^TX)^{-1}X^Teg_3^2$, and g_3 is $O_p(1)$, therefore H_1 is $O_p(n^{-1/2})$. Now $H_2 = H_{1\lambda} + H_{1\lambda\tau}g_3$ and $H_3 = H_{1\lambda\lambda} + H_{1\tau\tau}(B_{3\tau}A_{3\lambda})^2 + H_{1\lambda\tau}B_{3\tau}A_{3\lambda}$. Clearly $H_{1\tau\tau} = \frac{1}{\tau^3(\lambda)}(X^TX)^{-1}X^Te$ is $O_p(n^{-1/2})$. Since $H_{1\lambda}$ is $O_p(1)$, we have $H_{1\lambda\tau} = n^{-1/2}\frac{2}{\tau(\lambda)}H_{1\lambda} = O_p(n^{-1/2})$, completing ii) and iii).

Theorem 5.1. Under the assumptions of Lemma 5.1, we have the following asymptotic expansions

$$\begin{aligned}\hat{\beta}(\hat{\lambda}) - \beta(\lambda) &= \hat{\beta}(\lambda) - \beta(\lambda) + (X^TX)^{-1}X^Th_\lambda(y,\lambda)(\hat{\lambda}-\lambda) \\ &\quad + \frac{1}{2}(X^TX)^{-1}X^Th_{\lambda\lambda}(y,\lambda)(\hat{\lambda}-\lambda)^2 + O_p(n^{-3/2}) \end{aligned} \tag{5.2}$$

$$\begin{aligned}\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda}) &= \hat{\beta}(\lambda) - \beta(\lambda) + (X^TX)^{-1}X^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)](\hat{\lambda}-\lambda) \\ &\quad + \frac{1}{2}(X^TX)^{-1}X^T[h_{\lambda\lambda}(y,\lambda) - \mathrm{E}h_{\lambda\lambda}(y,\lambda)](\hat{\lambda}-\lambda)^2 + O_p(n^{-3/2}) \end{aligned} \tag{5.3}$$

Proof. Since $B = I_p$ and $A_{1\tau} = 0$, we have from Theorem 4.1 and Lemma 5.1 that

$$\hat{\beta}(\hat{\lambda}) = \beta(\lambda) + n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda}-\lambda) + n^{-1/2}H_{1\lambda}(\hat{\lambda}-\lambda) + H_{1\lambda\lambda}(\hat{\lambda}-\lambda)^2 + O_p(n^{-3/2})$$

Substituting the expressions for $B_{1\beta}$, $A_{1\lambda}$, $H_{1\lambda}$, and $H_{1\lambda\lambda}$ yields (5.2). Now (5.3) follows directly from (5.2) by applying the definition of $\beta_u(\hat{\lambda})$.

Remark 5.1. Under the ψ_i function defined in (5.1) we have

$$\begin{aligned}
\hat{\beta}(\hat{\lambda}) &= (X^TX)^{-1}X^T h(y, \hat{\lambda}), \\
\beta_u(l) &= (X^TX)^{-1}X^T \mathrm{E}[h(y, l)], \text{ for fixed } l, \\
\hat{\beta}(\lambda) &= (X^TX)^{-1}X^T h(y, \lambda), \\
\hat{\sigma}^2(\hat{\lambda}) &= \frac{1}{n}\sum_{i=1}^{n}\{h(y_i, \hat{\lambda}) - x_i^T\hat{\beta}(\hat{\lambda})\}^2, \\
\hat{\sigma}^2(\lambda) &= \frac{1}{n}\sum_{i=1}^{n}\{h(y_i, \lambda) - x_i^T\hat{\beta}(\lambda)\}^2,
\end{aligned} \tag{5.4}$$

Hence the asymptotic expansions (5.2) and (5.3) are equivalent to the expansions obtained by directly applying the second-order Taylor expansion of $h(y, \hat{\lambda})$ around λ to $\hat{\beta}(\hat{\lambda})$ and $\beta_u(\hat{\lambda})$ in (5.4).

Theorem 5.2. Under the assumptions of Lemma 5.1, $[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]/\hat{\sigma}(\hat{\lambda})$ has the following asymptotic expansion

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})} = \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} - n^{-1}(X^TX)^{-1}X^T d_1(\hat{\lambda}) + O_p(n^{-3/2}), \tag{5.5}$$

where

$$\begin{aligned}
d_1(\hat{\lambda}) = {} & \tau(\lambda)e\mathrm{E}[e^T h_\lambda(y,\lambda)](\hat{\lambda} - \lambda) + n\tau(\lambda)[h_\lambda(y, \lambda) - \mathrm{E}h_\lambda(y, \lambda)](\hat{\lambda} - \lambda) \\
& + \frac{n}{2}\tau(\lambda)[h_{\lambda\lambda}(y, \lambda) - \mathrm{E}h_{\lambda\lambda}(y, \lambda)](\hat{\lambda} - \lambda)^2.
\end{aligned}$$

Proof. It is straightforward from Theorem 4.2.

To compare (1.6) with (1.5) in terms of the coverage probability of the confidence region of $\beta_u(\hat{\lambda})$ obtained from (1.6), we need to introduce an F-ratio quantity. Let $\mathcal{F}(\hat{\lambda})$ denote the usual F-ratio quantity after estimating the transformation, i.e.,

$$\mathcal{F}(\hat{\lambda}) = \frac{[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]^T X^T X [\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]}{\sum_{i=1}^{n} \{h(y_i, \hat{\lambda}) - x_i^T \hat{\beta}\}^2} \bullet \frac{n-p}{p}. \tag{5.6}$$

Note that if e_i's are approximately i.i.d. standard normal, then $\mathcal{F}(\lambda)$ has an approximate F-distribution with p and $n-p$ degrees of freedom.

Corollary 5.1. Under the assumptions of Theorem 5.2 and $\frac{1}{n}(X^T X) = O(1)$, $\mathcal{F}(\hat{\lambda})$ has the following asymptotic expansion

$$\mathcal{F}(\hat{\lambda}) = \mathcal{F}(\lambda) + 2\frac{n-p}{n^2 p} e^T P d_1(\hat{\lambda}) + O_p(n^{-1}), \tag{5.7}$$

where

$$P = X(X^T X)X^T.$$

Proof. Since

$$\mathcal{F}(\hat{\lambda}) = \frac{[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]^T X^T X [\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]}{\hat{\sigma}^2(\hat{\lambda})} \bullet \frac{n-p}{np},$$

a direct application of Theorem 5.2 gives the result.

Now we apply Theorem 4.3 to the ψ function (5.1). Since $B = I_p, D = 1$, $A_{1\tau} = 0$, and $A_{3\beta} = 0$, we have

$$\begin{aligned}
c_0 &= A_{2\lambda} + A_{2\beta} B_{1\beta} A_{1\lambda} + A_{2\lambda} B_{3\tau} A_{3\lambda} \\
&= \frac{\tau^2(\lambda)}{n} \mathrm{E}\{[h_\lambda^T(y, \lambda) h_\lambda(y, \lambda)] + [h(y, \lambda) - X\beta(\lambda)]^T h_{\lambda\lambda}(y, \lambda)\} \\
&\quad + \frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda(y, \lambda)]^T X (-\frac{\tau^2(\lambda)}{n} X^T X)^{-1} \frac{\tau^2(\lambda)}{n} X^T \mathrm{E}[h_\lambda(y, \lambda)] \\
&\quad + \frac{2\tau(\lambda)}{n} \mathrm{E}[\frac{1}{\tau(\lambda)} e^T h_\lambda(y, \lambda)](-\frac{\tau^2(\lambda)}{2}) \frac{2\tau(\lambda)}{n} \mathrm{E}[\frac{1}{\tau(\lambda)} e^T h_\lambda(y, \lambda)] \\
&= \frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda^T(y, \lambda) h_\lambda(y, \lambda)] + \frac{\tau(\lambda)}{n} \mathrm{E}[e^T h_{\lambda\lambda}(y, \lambda)] \\
&\quad - \frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda(y, \lambda)]^T P \mathrm{E}[h_\lambda(y, \lambda)] - \frac{2\tau^2(\lambda)}{n^2} \{\mathrm{E}[e^T h_\lambda(y, \lambda)]\}^2
\end{aligned}$$

$$= \frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)]^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)]$$
$$+ \frac{\tau(\lambda)}{n} \mathrm{E}[e^T h_{\lambda\lambda}(y,\lambda)]$$
$$+ \frac{\tau^2(\lambda)}{n} \mathrm{E}[h_\lambda^T(y,\lambda)]Q\mathrm{E}[h_\lambda(y,\lambda)] - \frac{2\tau^2(\lambda)}{n^2} \{\mathrm{E}[e^T h_\lambda(y,\lambda)]\}^2,$$

where $Q = I_n - P$ and I_n is an $n{\times}n$ identity matrix.

Theorem 5.3. Under the assumptions of Theorem 4.3, if the ψ_i function is defined by (5.1), then $\hat{\lambda}$ has the first-order asymptotic expansion

$$\hat{\lambda} - \lambda = \frac{\tau(\lambda)}{nc_0} \{- e^T h_\lambda(y,\lambda) + \frac{1}{\tau(\lambda)} \sum_i \log y_i + \mathrm{E}[h_\lambda^T(y,\lambda)]Pe$$
$$- \mathrm{E}[e^T h_\lambda(y,\lambda)](1 - \frac{1}{n} e^T e)\} + O_p(n^{-1}). \tag{5.8}$$

Proof. Since $B = I_p, D = 1, A_{1\tau} = 0,$ and $A_{3\beta} = 0$, we have from Theorem 4.3

$$\hat{\lambda} - \lambda = -\{\frac{1}{c_0}\psi_2 + n^{-1/2}\frac{1}{c_0}(A_{2\beta}\, g_1 + A_{2\tau}\, g_3)\} + O_p(n^{-1}).$$

Substituting the expressions for ψ_2, $A_{2\beta}$, g_1, $A_{2\tau}$, and g_3, we obtain (5.8).

The results related to variance can be easily obtained from Propositions 4.1.

Proposition 5.1. Under the assumptions of Proposition 4.1 and assuming further that $(X^TX)^{-1}X^T[h_{\lambda\lambda}(y,\lambda) - \mathrm{E}h_{\lambda\lambda}(y,\lambda)] = O_p(n^{-1/2})$, we have

$$\mathrm{Cov}\left[\frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})}\right] = \mathrm{Cov}\left[\frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)}\right]$$
$$+ \frac{2\tau^2(\lambda)}{nc_0}(X^TX)^{-1}X^T\Lambda X(X^TX)^{-1} + O(n^{-2}), \tag{5.9}$$

where $\Lambda = \mathrm{E}\Big\{ e[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda) - \frac{1}{n} e\mathrm{E}(e^T h_\lambda(y,\lambda))]^T [- e^T h_\lambda(y,\lambda)$

$$+ \frac{1}{\tau(\lambda)} \sum_i \log y_i + (\mathrm{E} h_\lambda^T(y, \lambda))Pe - (\mathrm{E} e^T h_\lambda(y, \lambda))(1 - \frac{1}{n} e^T e)] \} .$$

Proof. Straightforward.

Note that similar theorems corresponding to other h function can be easily obtained by making changes related to the last term of ψ_{2i} in (5.1).

5.3. Further Approximation to the Asymptotic Expansion Formulas

Based on the asymptotic expansions in Section 5.2, we make some further approximations in order to draw some specific conclusions in different situations, e.g., $\sigma(\lambda)$ is small or large, and the model is structured or unstructured. Let $\eta = X\beta(\lambda)$, and define

$$g(z, \hat{\lambda}, \lambda) = h(h^{-1}(z, \lambda), \hat{\lambda}) = h(y, \hat{\lambda}) \qquad (5.10)$$

Let $g_z(z, \hat{\lambda}, \lambda)$ and $g_{zz}(z, \hat{\lambda}, \lambda)$ denote the first and second-order partial derivatives of $g(z, \hat{\lambda}, \lambda)$ with respect to z obtained componentwise. Then we have for the Box-Cox power transformation

$$g(z, \hat{\lambda}, \lambda) = \begin{cases} \dfrac{(1 + \lambda z)^{\hat{\lambda}/\lambda} - 1}{\hat{\lambda}}, & \lambda \neq 0, \quad \hat{\lambda} \neq 0, \\ \dfrac{1}{\lambda} \log(1 + \lambda z), & \lambda \neq 0, \quad \hat{\lambda} = 0, \\ \dfrac{e^{\hat{\lambda} z} - 1}{\hat{\lambda}}, & \lambda = 0, \quad \hat{\lambda} \neq 0, \\ z, & \lambda = 0, \quad \hat{\lambda} = 0, \end{cases} \qquad (5.11)$$

$$g_z(z, \hat{\lambda}, \lambda) = \begin{cases} (1 + \lambda z)^{\hat{\lambda}/\lambda - 1}, & \lambda \neq 0, \\ e^{\hat{\lambda} z}, & \lambda = 0, \end{cases} \qquad (5.12)$$

and

$$g_{zz}(z, \hat{\lambda}, \lambda) = \begin{cases} (\hat{\lambda} - \lambda)(1 + \lambda z)^{\hat{\lambda}/\lambda - 2}, & \lambda \neq 0, \\ \hat{\lambda} e^{\hat{\lambda} z}, & \lambda = 0. \end{cases} \tag{5.13}$$

Let # denote a componentwise vector multiplication operator, i.e., for any two column vectors a and b of the same length, $a\#b$ is a new vector with the ith element $a_i b_i$. The common functions such as square and log applied to a vector a are operated componentwise, e.g., $a^2 = \{ a_i^2 \}_{n\times 1}$ and $\log a = \{ \log a_i \}_{n\times 1}$.

Now a second-order Taylor expansion of g around η gives

$$\begin{aligned} h(y, \hat{\lambda}) &= g(z, \hat{\lambda}, \lambda) \\ &\approx g(\eta, \hat{\lambda}, \lambda) + \sigma(\lambda)\, g_\eta(\eta, \hat{\lambda}, \lambda)\#e + \frac{\sigma^2(\lambda)}{2}\, g_{\eta\eta}(\eta, \hat{\lambda}, \lambda)\#e^2 \\ &= a_0(\hat{\lambda}) + \sigma(\lambda) a_1(\hat{\lambda})\#e + \frac{\sigma^2(\lambda)}{2} a_2(\hat{\lambda})\#e^2, \end{aligned} \tag{5.14}$$

where

$$\begin{aligned} a_0(\hat{\lambda}) &= g(\eta, \hat{\lambda}, \lambda), \\ a_1(\hat{\lambda}) &= g_\eta(\eta, \hat{\lambda}, \lambda), \\ a_2(\hat{\lambda}) &= g_{\eta\eta}(\eta, \hat{\lambda}, \lambda), \\ g_\eta(\eta, \hat{\lambda}, \lambda) &= g_z(z, \hat{\lambda}, \lambda)|_{z=\eta}, \end{aligned}$$

and

$$g_{\eta\eta}(\eta, \hat{\lambda}, \lambda) = g_{zz}(z, \hat{\lambda}, \lambda)|_{z=\eta}.$$

Note that the approximation (5.14) gives a second-order approximation to $g(z,\hat{\lambda},\lambda)$ locally around each mean $x_i^T\beta(\lambda)$, $i = 1, \ldots, n$. For $\sigma(\lambda)$ small to moderate, this approximation should be adequate. For $\sigma(\lambda)$ large the approximation may not be good enough, hence the conclusions based on this should be treated with caution.

Let $a_{i\lambda}(\lambda)$ and $a_{i\lambda\lambda}(\lambda)$, $i = 0, 1, 2$, denote the first and second-order derivatives of $a_i(\hat{\lambda})$, $i = 0, 1, 2$, respectively evaluated at $\hat{\lambda} = \lambda$. Let $h_\lambda(y, \lambda)$ and $h_{\lambda\lambda}(y, \lambda)$ denote the first and second-order partial derivatives of $h(y, \hat{\lambda})$ with respect to $\hat{\lambda}$ evaluated at $\hat{\lambda} = \lambda$.

Using the above we have

$$h_\lambda(y, \lambda) \approx a_{0\lambda}(\lambda) + \sigma(\lambda)a_{1\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)\#e^2, \tag{5.15}$$

$$h_{\lambda\lambda}(y, \lambda) \approx a_{0\lambda\lambda}(\lambda) + \sigma(\lambda)\, a_{1\lambda\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2} a_{2\lambda\lambda}(\lambda)\#e^2, \tag{5.16}$$

and

$$\log y = g(z, 0, \lambda) \approx a_0(0) + \sigma(\lambda)a_1(0)\#e + \frac{\sigma^2(\lambda)}{2} a_2(0)\#e^2, \tag{5.17}$$

where

$$a_0(0) = g(\eta, 0, \lambda),$$
$$a_1(0) = g_\eta(\eta, 0, \lambda),$$

and

$$a_2(0) = g_{\eta\eta}(\eta, 0, \lambda).$$

Using (5.11)-(5.13), we can easily have for $\lambda \neq 0$,

$$\begin{aligned}
a_0(0) &= \lambda^{-1}\log(1 + \lambda\eta),\\
a_1(0) &= (1 + \lambda\eta)^{-1},\\
a_2(0) &= -\lambda(1 + \lambda\eta)^{-2},\\
a_{0\lambda}(\lambda) &= \lambda^{-2}[(1 + \lambda\eta)\#(\log(1 + \lambda\eta)) - \lambda\eta],\\
a_{1\lambda}(\lambda) &= \lambda^{-1}\log(1 + \lambda\eta),\\
a_{1\lambda\lambda}(\lambda) &= [\lambda^{-1}\log(1 + \lambda\eta)]^2\\
&= [a_{1\lambda}(\lambda)]^2;\\
a_{2\lambda}(\lambda) &= (1 + \lambda\eta)^{-1},\\
a_{2\lambda\lambda}(\lambda) &= 2\lambda^{-1}(1 + \lambda\eta)^{-1}\log(1 + \lambda\eta).
\end{aligned} \tag{5.18}$$

Based on the above we have the following results corresponding to the theorems of Section 5.2. We refer to these as results rather than theorems because the effect of the approximation (5.14) is not quantified in a rigorous manner. Nevertheless, we expect the error introduced by the approximation (5.14) to be small relative to the remainder terms in the previous asymptotic expansions. We convert $\tau(\lambda)$ to $\sigma(\lambda)$ in the final expressions.

Result 5.1. Assume that the first four moments of e_i are the same as the first four moments of a standard normal random variable, i.e., $Ee_i = 0$, $Ee_i^2 = 1$, $Ee_i^3 = 0$, and $Ee_i^4 = 3$. Under the approximation (5.14) and the assumptions of Theorem 5.3, $\hat{\lambda}$ has the following approximate asymptotic expansion

$$\hat{\lambda} - \lambda \approx -\frac{1}{nc_0^*}\{\frac{1}{\sigma(\lambda)} a_{0\lambda}^T(\lambda)Qe + e^T a_{1\lambda}(\lambda)\#e + 1_n^T a_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e) - 1_n^T a_0(0)$$
$$+ \sigma(\lambda)[\frac{1}{2}e^T a_{2\lambda}(\lambda)\#e^2 - 1_n^T a_1(0)\#e - \frac{1}{2}e^T P a_{2\lambda}(\lambda)] - \frac{\sigma^2(\lambda)}{2} 1_n^T a_2(0)\#e^2\}$$
$$+ O_p(n^{-1}), \tag{5.19}$$

where

$$c_0^* = \frac{1}{2n}\sigma^2(\lambda)1_n^T a_{2\lambda}^2(\lambda) + \frac{2}{n}[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)]^T[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)]$$
$$+ \frac{1}{n\sigma^2(\lambda)}[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2}a_{2\lambda}(\lambda)]^T Q[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2}a_{2\lambda}(\lambda)],$$

and

$$\bar{a}_{1\lambda}(\lambda) = n^{-1}1_n^T a_{1\lambda}(\lambda).$$

Proof. Substituting (5.15)-(5.17) into Theorem 5.3 we have

$$-(\hat{\lambda} - \lambda)$$
$$= \frac{\tau(\lambda)}{nc_0}\{e^T h_\lambda(y, \lambda) - \frac{1}{\tau(\lambda)}\sum_i \log y_i - E[h_\lambda^T(y, \lambda)]Pe$$
$$+ E[e^T h_\lambda(y, \lambda)](1 - \frac{1}{n}e^T e)\} + O_p(n^{-1})$$

$$= \frac{\tau(\lambda)}{nc_0}\{e^T a_{0\lambda}(\lambda) + \sigma(\lambda)e^T a_{1\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2}e^T a_{2\lambda}(\lambda)\#e^2 - \sigma(\lambda)1_n^T a_0(0)$$

$$- \sigma^2(\lambda)1_n^T a_1(0)\#e - \frac{\sigma^3(\lambda)}{2}1_n^T a_2(0)\#e^2 - a_{0\lambda}^T(\lambda)Pe - \frac{\sigma^2(\lambda)}{2}[a_{2\lambda}(\lambda)]^T Pe$$

$$+ \sigma(\lambda)1_n^T a_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e)\} + O_p(n^{-1})$$

$$= \frac{1}{nc_0}\{\frac{1}{C(\lambda)}a_{0\lambda}^T(\lambda)Qe + e^T a_{1\lambda}(\lambda)\#e + 1_n^T a_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e) - 1_n^T a_0(0)$$

$$+ \sigma(\lambda)[\frac{1}{2}e^T a_{2\lambda}(\lambda)\#e^2 - 1_n^T a_1(0)\#e - \frac{1}{2}e^T P a_{2\lambda}(\lambda)] - \frac{\sigma^2(\lambda)}{2}1_n^T a_2(0)\#e^2\}$$

$$+ O_p(n^{-1}).$$

and

$$c_0 = \frac{\tau^2(\lambda)}{n}\mathrm{E}\{[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)]^T[h_\lambda(y,\lambda) - \mathrm{E}h_\lambda(y,\lambda)]\} + \frac{\tau(\lambda)}{n}\mathrm{E}[e^T h_{\lambda\lambda}(y,\lambda)]$$

$$+ \frac{\tau^2(\lambda)}{n}\mathrm{E}[h_\lambda^T(y,\lambda)]Q\mathrm{E}[h_\lambda(y,\lambda)] - \frac{2\tau^2(\lambda)}{n^2}\{\mathrm{E}[e^T h_\lambda(y,\lambda)]\}^2$$

$$= \frac{\tau^2(\lambda)}{n}\mathrm{E}\{[\sigma(\lambda)a_{1\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2}a_{2\lambda}(\lambda)\#(e^2-1)]^T[\sigma(\lambda)a_{1\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2}$$

$$\bullet a_{2\lambda}(\lambda)\#(e^2-1)]\} + \frac{1}{n}1_n^T a_{1\lambda\lambda}(\lambda) + \frac{\tau^2(\lambda)}{n}[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{n}a_{2\lambda}(\lambda)]^T Q[a_{0\lambda}(\lambda)$$

$$+ \frac{\sigma^2(\lambda)}{n}a_{2\lambda}(\lambda)] - \frac{2}{n^2}[1_n^T a_{1\lambda}(\lambda)]^2$$

$$= \frac{\tau^2(\lambda)}{n}[\sigma^2(\lambda)1_n^T a_{1\lambda}^2(\lambda) + \frac{1}{2}\sigma^4(\lambda)1_n^T a_{2\lambda}^2(\lambda)] + \frac{1}{n}1_n^T a_{1\lambda\lambda}(\lambda) - \frac{2}{n^2}[1_n^T a_{1\lambda}(\lambda)]^2$$

$$+ \frac{\tau^2(\lambda)}{n}[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{n}a_{2\lambda}(\lambda)]^T Q[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{n}a_{2\lambda}(\lambda)]$$

$$= \frac{1}{2n}\sigma^2(\lambda)1_n^T a_{2\lambda}^T(\lambda) + \frac{1}{n}1_n^T a_{1\lambda}^T(\lambda) + \frac{1}{n}1_n^T a_{1\lambda\lambda}(\lambda) - \frac{2}{n^2}[1_n^T a_{1\lambda}(\lambda)]^2$$

$$+ \frac{1}{n\sigma^2(\lambda)}[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2}a_{2\lambda}(\lambda)]^T Q[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2}a_{2\lambda}(\lambda)].$$

Now (5.18) gives

$$\frac{1}{n}1_n^T a_{1\lambda}^2(\lambda) + \frac{1}{n}1_n^T a_{1\lambda\lambda}(\lambda) - \frac{2}{n^2}[1_n^T a_{1\lambda}(\lambda)]^2$$

$$= \frac{2}{n}1_n^T a_{1\lambda}^2(\lambda) - \frac{2}{n^2}[1_n^T a_{1\lambda}(\lambda)]^2$$

$$= \frac{2}{n}[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)]^T[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)].$$

Hence $c_0 \approx c_0^*$, completing the proof.

Result 5.2. Under the assumptions of Theorem 5.2 and Result 5.1, the Box-Cox λ-unknown quantity $[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})]/\hat{\sigma}(\hat{\lambda})$ has the following approximate asymptotic expansion

$$\frac{\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})}{\hat{\sigma}(\hat{\lambda})} \approx \frac{\hat{\beta}(\lambda) - \beta(\lambda)}{\hat{\sigma}(\lambda)} + n^{-1}(X^TX)^{-1}X^Td^* + O_p(n^{-3/2}), \tag{5.20}$$

where

$$\begin{aligned} d^* &= [-1_n^T a_{1\lambda}(\lambda)e + n a_{1\lambda}(\lambda)\#e + \frac{\sigma(\lambda)}{2} n a_{2\lambda}(\lambda)\#(e^2 - 1)]\delta_\lambda \\ &+ \frac{1}{2} n[a_{1\lambda\lambda}(\lambda)\#e + \frac{\sigma(\lambda)}{2} a_{2\lambda\lambda}(\lambda)\#(e^2 - 1)]\delta_\lambda^2, \end{aligned}$$

and δ_λ is given in (5.19) by omitting the term $O_p(n^{-1})$.

Proof. From (5.15) and (5.16) we have

$$h_\lambda(y, \lambda) - \mathrm{E}h_\lambda(y, \lambda) \approx \sigma(\lambda)a_{1\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)\#(e^2 - 1),$$
$$h_{\lambda\lambda}(y, \lambda) - \mathrm{E}h_{\lambda\lambda}(y, \lambda) \approx \sigma(\lambda)a_{1\lambda\lambda}(\lambda)\#e + \frac{\sigma^2(\lambda)}{2} a_{2\lambda\lambda}(\lambda)\#(e^2 - 1),$$

and

$$\mathrm{E}[e^T h_\lambda(y, \lambda)] \approx \sigma(\lambda) 1_n^T a_{1\lambda}(\lambda).$$

Substituting those expressions into (5.5) and using Result 5.1 yields the result.

Result 5.3. Under the assumptions of Result 5.2, we have

$$\mathcal{F}(\hat{\lambda}) \approx \mathcal{F}(\lambda) + 2\,\frac{n-p}{n^2 p}\, e^T P d^* + O_p(n^{-1}). \tag{5.21}$$

Note that the n-vectors $a_{0\lambda}(\lambda)$, $a_{0\lambda\lambda}(\lambda)$, $a_{1\lambda}(\lambda)$, $a_{1\lambda\lambda}(\lambda)$, $a_{2\lambda}(\lambda)$, $a_{2\lambda\lambda}(\lambda)$, $a_0(0)$, $a_1(0)$, and $a_2(0)$ are all functions of η and λ only, hence the formulas (5.20) and (5.21) are expressed explicitly in terms of $\sigma(\lambda)$ and e. Thus allows us to discuss the behavior of the second-order term as $\sigma(\lambda)$ varies with e and n fixed. We base our arguments on Result 5.2. Since the second-order term in (5.20) is $n^{-1}(X^TX)^{-1}X^Td*$ in which only $d*$ involves $\sigma(\lambda)$, we initially consider how $d*$ varies with $\sigma(\lambda)$.

Before we introducing the next result we clarify our definition of *structured* and *unstructured models*. A model is structured if $h(y,\lambda)=X\beta$ and $h(y,\lambda^*)=X\beta^*$ together imply $(\lambda,\beta)=(\lambda^*,\beta^*)$. Since we assume that X has full rank, this is equivalent to saying that $Qg(X\beta,\lambda^*,\lambda)=0$ implies that $\lambda^*=\lambda$. A model is unstructured if it is equivalent to a one-way layout. This implies that $Qf(X\beta)=0$ for all functions f defined componentwise.

Result 5.4. Based on the Result 5.2, we have for fixed $e(\lambda)$ and n:

i) For a structured model such that $Qa_{0\lambda}(\lambda)\neq 0$, $d*$ approaches zero as $\sigma(\lambda)$ approaches zero;

ii) For nontrivial unstructured models such as one-way layout with at least two means, as $\sigma(\lambda)$ approaches zero, $d*$ approaches a nonzero quantity;

iii) For one-sample case, $d*$ approaches infinity as $\sigma(\lambda)$ approaches zero.

Proof. From Result 5.1 and Result 5.2 we have

$$d* = [-1_n^Ta_{1\lambda}(\lambda)e + na_{1\lambda}(\lambda)\#e + \frac{\sigma(\lambda)}{2} na_{2\lambda}(\lambda)\#(e^2-1)]\delta_\lambda + \frac{1}{2} n[a_{1\lambda\lambda}(\lambda)\#e + \frac{\sigma(\lambda)}{2} a_{2\lambda\lambda}(\lambda)\#(e^2-1)]\delta_\lambda{}^2,$$

where

$$\delta_\lambda = -\frac{1}{nc_0^*}\{\frac{1}{\sigma(\lambda)} a_{0\lambda}^T(\lambda)Qe + e^Ta_{1\lambda}(\lambda)\#e + 1_n^Ta_{1\lambda}(\lambda)(1-\frac{1}{n}e^Te) - 1_n^Ta_0(0)$$

$$+ \sigma(\lambda)[\frac{1}{2} e^T a_{2\lambda}(\lambda)\#e^2 - 1_n^T a_1(0)\#e - \frac{1}{2} e^T P a_{2\lambda}(\lambda)] - \frac{\sigma^2(\lambda)}{2} 1_n^T a_2(0)\#e^2\},$$

$$c_0^* = \frac{1}{2n} \sigma^2(\lambda) 1_n^T a_{2\lambda}^2(\lambda) + \frac{2}{n} [a_{1\lambda}(\lambda) - 1_n \bar{a}_{1\lambda}(\lambda)]^T [a_{1\lambda}(\lambda) - 1_n \bar{a}_{1\lambda}(\lambda)]$$
$$+ \frac{1}{n\sigma^2(\lambda)} [a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)]^T Q [a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)],$$

and

$$\bar{a}_{1\lambda}(\lambda) = n^{-1} 1_n^T a_{1\lambda}(\lambda).$$

For a structured model such that $Qa_{0\lambda}(\lambda) \neq 0$, the quantity in the last term of c_0^*,

$$[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)]^T Q [a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)] \neq 0,$$

and the quantity in the first term of δ_λ,

$$a_{0\lambda}^T(\lambda) Q e \neq 0.$$

We have

$$\text{as } \sigma(\lambda) \longrightarrow 0,\ \sigma^2(\lambda) c_0 \longrightarrow \frac{1}{n} a_{0\lambda}^T(\lambda) Q a_{0\lambda}(\lambda),$$

hence for fixed e and n

$$\delta_\lambda / \sigma(\lambda) \ \rightarrow\ a_{0\lambda}^T Q e / a_{0\lambda}^T(\lambda) Q a_{0\lambda}(\lambda),$$

and

$$d^* \longrightarrow 0.$$

. For an unstructured model with at least two means, we have

$$[a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)]^T Q [a_{0\lambda}(\lambda) + \frac{\sigma^2(\lambda)}{2} a_{2\lambda}(\lambda)] = 0,$$

$$a_{0\lambda}^T(\lambda) Q e = 0,$$

and

$$[a_{1\lambda}(\lambda) - 1_n \bar{a}_{1\lambda}(\lambda)]^T [a_{1\lambda}(\lambda) - 1_n \bar{a}_{1\lambda}(\lambda)] \neq 0,$$

as it is for a one-way layout with unequal means. Then as $\sigma(\lambda) \longrightarrow 0$,

$$c_0 \longrightarrow \frac{2}{n}\,[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)]^T\,[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)],$$

so for fixed e and n

$$\delta_\lambda \longrightarrow -\frac{e^T a_{1\lambda}(\lambda)\#e + 1_n^T a_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e) - 1_n^T a_0(0)}{2[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)]^T\,[a_{1\lambda}(\lambda) - 1_n\bar{a}_{1\lambda}(\lambda)]}$$

implying that

$$d^* \longrightarrow (X^T X)^{-1} X^T [- n^{-1} 1_n^T a_{1\lambda}(\lambda)^T e\delta_\lambda + a_{1\lambda\lambda}(\lambda)\#e\delta_\lambda + a_{1\lambda\lambda}(\lambda)\#e\delta_\lambda^2].$$

Further for the unstructured model with only one mean (one-sample case), we have

$$[a_{1\lambda}(\lambda) - 1_n\,\bar{a}_{1\lambda}(\lambda)]^T[a_{1\lambda}(\lambda) - 1_n\,\bar{a}_{1\lambda}(\lambda)] = 0,$$

implying that

$$c_0 = \frac{1}{2n}\,\sigma^2(\lambda) 1_n^T a_{2\lambda}^2(\lambda).$$

Hence for fixed e and n

as $\sigma(\lambda) \longrightarrow 0$, $|\delta_\lambda| \longrightarrow \infty$, implying that $|d^*| \longrightarrow \infty$.

The proof is complete.

When n is large the Result 5.1 allows us to discuss the behavior of $\hat{\lambda}$ in different situations. Since $\hat{\lambda}-\lambda \approx \delta_\lambda$ when n is large, then following the proof of Result 5.4 we have for fixed e and n:

i) In a structured model with $Qa_{0\lambda}(\lambda) \neq 0$, $\hat{\lambda} - \lambda \longrightarrow 0$ as $\sigma(\lambda) \longrightarrow 0$;

ii) In unstructured model, as $\sigma(\lambda) \to 0$,

$$\hat{\lambda} - \lambda \;\to\; -\frac{e^T a_{1\lambda}(\lambda)\#e + 1_n^T a_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e) - 1_n^T a_0(0)}{2[a_{1\lambda}(\lambda) - 1_n \bar{a}_{1\lambda}(\lambda)]^T [a_{1\lambda}(\lambda) - 1_n \bar{a}_{1\lambda}(\lambda)]},$$

when the means are far apart, and $|\hat{\lambda} - \lambda| \to \infty$ when the means are equal.

So the structured model is able to provide better estimate of λ than the unstructured model, particularly when $\sigma(\lambda)$ is small; the unstructured model can give a very poor estimate of λ when the means are close together and n is not large. The above conclusions partially answer the research problem proposed in Box and Cox (1982): "There are numerous aspects of transformations that merit further study, These include in particular the further development of simple ways of assessing *transformation potential*; that is, of providing some more formal measure of the ability of particular data to provide useful information about a class of transformations. ..." When n is moderate-to-small a higher order expansion is needed to study the behavior of $\hat{\lambda}$.

Now we investigate the approximate forms of Propositions 5.1, basically to compute Λ in (5.9) .] $\sigma(\lambda)$, i.e., using the first-order Taylor approximation $h(y, \hat{\lambda}) = g(z, \hat{\lambda}, \lambda) \approx a_0(\hat{\lambda}) + \sigma(\lambda)\, a_1(\hat{\lambda})\#e$.

Result 5.5.. When $\sigma(\lambda)$ is small, the quantity Λ in (5.9) becomes $\Lambda \approx 0$.

Proof. When $\sigma(\lambda)$ is small, (5.14) reduces to $h(y, \hat{\lambda}) \approx a_0(\hat{\lambda}) + \sigma(\lambda) a_1(\hat{\lambda})\#e$. Substituting this into the expressions for Λ, we have

$$\Lambda = \mathrm{E}\{e[h_\lambda(y, \lambda) - \mathrm{E}h_\lambda(y, \lambda) - \frac{1}{n}e\mathrm{E}(e^T h_\lambda(y, \lambda))]^T [-e^T h_\lambda(y, \lambda)$$

$$+\frac{1}{\tau}\sum_i \log y_i + (\mathrm{E}h_\lambda^T(y, \lambda))Pe - (\mathrm{E}e^T h_\lambda(y, \lambda))(1 - \frac{1}{n}e^T e)]\}$$

$$\approx E\{\sigma(\lambda)e[(a_{1\lambda}(\hat{\lambda}) - \bar{a}_{1\lambda}(\lambda))\#e]^T[-e^T(a_{0\lambda}(\hat{\lambda}) + \sigma(\lambda)a_{1\lambda}(\hat{\lambda})\#e)$$
$$+ \sigma(\lambda)I_n^T(a_0(0) + \sigma(\lambda)a_1(0)\#)e + a_{0\lambda}^T(\lambda)Pe - n\sigma(\lambda)\bar{a}_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e)]\}$$
$$= E\{\sigma(\lambda)e[(a_{1\lambda}(\hat{\lambda}) - \bar{a}_{1\lambda}(\lambda))\#e]^T[-\sigma(\lambda)e^T a_{1\lambda}(\hat{\lambda})\#e + \sigma(\lambda)I_n^T a_0(0)$$
$$- n\sigma(\lambda)\bar{a}_{1\lambda}(\lambda)(1 - \frac{1}{n}e^T e)]\}$$
$$= \sigma^2(\lambda)\{[a_{1\lambda i}(\lambda) - \bar{a}_{1\lambda}(\lambda)][-2a_{1\lambda i}(\lambda) - n\bar{a}_{1\lambda}(\lambda) + n\bar{a}_0(0) - n\bar{a}_{1\lambda}(\lambda)$$
$$+ (2 + n)\bar{a}_{1\lambda}(\lambda)]\}$$
$$\approx n\sigma^2(\lambda)[a_{1\lambda i}(\lambda) - \bar{a}_{1\lambda}(\lambda)][\bar{a}_0(0) - \bar{a}_{1\lambda}(\lambda)].$$

Further from (5.18) we have $a_0(0) = a_{1\lambda}(\lambda) = \lambda^{-1}\log(1 + \lambda\eta)$, hence $\Lambda \approx 0$.

5.4. Simulations

In this section we present simulation results for the coverage probabilities of the usual confidence region of $\beta(\hat{\lambda})$, i.e. the confidence region obtained from $\mathcal{F}(\hat{\lambda})$ in (5.6). We are mainly interested in seeing the behavior of this confidence region under different situations. From the results of Section 5.3, the following factors are important: i) the type of model, ii) the spread of the means $x_i^T\beta(\lambda)$, $i = 1, \ldots, n$, iii) the variance of the model error $\sigma^2(\lambda)$, and iv) the sample size n.

Simulation Process. The simulation process can be described simply as follows: first generate a sample $z_1, \ldots, z_n$ of size n from standard normal distribution, check whether $(1+\lambda z_i) > 0$, $i = 1, \ldots, n$ (if $(1+\lambda z_i) \leq 0$ for some i, then generate another z_i), calculate the quantity $\mathcal{F}(\hat{\lambda})$ by using the approximation (5.21), and then compare the simulated value of $\mathcal{F}(\hat{\lambda})$ with the critical values.

Algorithm. We used ***FORTRAN*** on a Macintosh II to perform the simulations. The ***GFSR Algorithm*** was used to generate the uniform random numbers between 0 and 1 which were converted to the standard normal random numbers by the ***Box-Muller Algorithm***.

Simulation Errors. The standard errors of the simulated coverage probabilities can be approximated as follows. Treat each run (one F random number is generated) as a Bernolli trial with the probability of success (generated F value is less than the corresponding critical value) $1-\alpha$. Given cm runs the standard errors of the simulated coverage probabilities are $[(1-\alpha)\alpha/cm]^{1/2}$. With $cm = 5000$ and $1-\alpha = 0.75, 0.90, 0.95, 0.975$, and 0.99 the standard errors are respectively 0.0061, 0.0042 0.0031, 0.0023, and 0.0014.

Example 5.1. 3^3 factorial design with additive effects and with one or two replicates. The sample sizes n are thus 27 and 54. The number of runs cm=5000.

When $n = 27$ (one replicate), the design matrix is given as follows

$$X^T = \left[\begin{array}{ccccccccccccccccccccccccccc} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ -1&-1&-1&-1&-1&-1&-1&-1&-1&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1 \\ -1&-1&-1&0&0&0&1&1&1&-1&-1&-1&0&0&0&1&1&1&-1&-1&-1&0&0&0&1&1&1 \\ -1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1&-1&0&1 \end{array}\right]$$

When $n = 54$ (two replicates), the design matrix is obtained by repeating each column of the X^T matrix given above. We use the fitted values in the Box-Cox Textile Example for the model parameters, i.e., $\lambda = -0.06$, and $\beta^T(\lambda) = (\beta_0, \beta_1, \beta_2, \beta_3)$ with $\beta_0 = 5.2523$, $\beta_1 = 0.569$, $\beta_2 = -0.4313$, and $\beta_3 = -0.2682$. We look at the combinations of five different coverage probabilities, i.e., $1-\alpha = 0.75, 0.90, 0.95, 0.975$, and 0.99, and four different variances, i.e. $\sigma^2(\lambda) = 0.0001, 0.01, 1.0$, and 10.

The simulated coverage probabilities are given in Table 5.1. We see that the coverage probabilities are very close to the nominal level $1-\alpha$. All of the simulated values are within two standard errors of the corresponding nominal level $1-\alpha$.

Table 5.1. Simulation results for 3^3 factorial design with additive effects

$\sigma^2(\lambda)$	$n = 27$					$n = 54$				
			$1-\alpha$					$1-\alpha$		
	.75	.90	.95	.975	.99	.75	.90	.95	.975	.99
.0001	.7552	.9022	.9522	.9764	.9878	.7526	.9012	.9568	.9780	.9916
.01	.7480	.8974	.9440	.9718	.9878	.7510	.9032	.9508	.9752	.9896
1.0	.7430	.8898	.9442	.9740	.9896	.7346	.8962	.9492	.9728	.9892
10.	.7416	.9002	.9516	.9770	.9902	.7382	.8856	.9390	.9698	.9880

Example 5.2. **One-way layout**

The number of runs used in this example is 5000, and the true transformation is 0.1. We use three different sets of means: i) $(\beta_0, \beta_1, \beta_2) = (10, 8, 6)$, ii) $(\beta_0, \beta_1, \beta_2) = (9, 8, 7)$, and iii) $(\beta_0, \beta_1, \beta_2) = (8.1, 8.0, 7.9)$; three levels of sample sizes, $n = 9$, 18, and 36; five coverage probabilities, i.e., $1-\alpha = 0.75$, 0.90, 0.95, 0.975, and 0.99; four variances, i.e., $\sigma^2 = 0.0001$, 0.01, 1.0, and 10.

The simulation results are given in Table 5.2. The effect of the structure of the means, the sample sizes, and the error variance as predicted by the expansions are clearly demonstrated in this example. First if the means are far apart, the difference between the simulated confidence level and the nominal level is generally small but not negligible if the sample size is small, i.e., $n = 9$. If the means are close together the differences are much greater. Increasing the variance when the means are close together reduces the difference. Simulation results in this one-way layout indicate that the usual confidence regions are generally liberal.

Table 5.2. Simulation results for one-way layout

$\sigma^2(\lambda)$	$n = 9$					$n = 18$				
	$1-\alpha$					$1-\alpha$				
	.75	.90	.95	.975	.99	.75	.90	.95	.975	.99
	$(\beta_0, \beta_1, \beta_2) = (10, 8, 6)$					$(\beta_0, \beta_1, \beta_2) = (10, 8, 6)$				
.0001	.7092	.8864	.9396	.9658	.9860	.7448	.8964	.9474	.9726	.9900
.01	.7256	.8890	.9402	.9682	.9898	.7276	.8908	.9430	.9718	.9886
1.0	.7248	.8866	.9428	.9686	.9860	.7530	.8982	.9496	.9752	.9914
10.	.7102	.8848	.9398	.9672	.9860	.7360	.8936	.9468	.9708	.9872
	$(\beta_0, \beta_1, \beta_2) = (9, 8, 7)$					$(\beta_0, \beta_1, \beta_2) = (9, 8, 7)$				
.0001	.6354	.8174	.8864	.9324	.9668	.6946	.8676	.9206	.9560	.9780
.01	.6448	.8304	.8956	.9374	.9660	.7014	.8756	.9342	.9616	.9812
1.0	.6458	.8342	.9020	.9446	.9722	.7006	.8662	.9292	.9622	.9806
10.	.6780	.8576	.9244	.9608	.9792	.7162	.8834	.9378	.9656	.9856
	$(\beta_1, \beta_2, \beta_3) = (8.1, 8.0, 7.9)$					$(\beta_1, \beta_2, \beta_3) = (8.1, 8.0, 7.9)$				
.0001	.2298	.3500	.4260	.4816	.5422	.2668	.3960	.4754	.5230	.5842
.01	.2252	.3534	.4222	.4796	.5488	.2918	.4158	.4920	.5460	.5952
1.0	.4778	.6646	.7556	.8174	.8742	.5440	.7218	.8026	.8528	.8934
10.	.6614	.8392	.9078	.9450	.9726	.7008	.8698	.9324	.9650	.9846

$\sigma^2(\lambda)$	$n=36$									
	$(\beta_0, \beta_1, \beta_2) = (10, 8, 6)$					$(\beta_0, \beta_1, \beta_2) = (9, 8, 7)$				
.0001	.7574	.9022	.9504	.9758	.9896	.7290	.8970	.9492	.9716	.9882
.01	.7444	.9014	.9532	.9774	.9924	.7390	.8938	.9438	.9710	.9854
1.0	.7506	.9056	.9508	.9738	.9888	.7286	.8976	.9472	.9750	.9892
10.	.7464	.8998	.9492	.9736	.9894	.7376	.9000	.9510	.9764	.9912
	$(\beta_0, \beta_1, \beta_2) = (8.1, 8.0, 7.9)$									
.0001	.3688	.5278	.6022	.6548	.7104					
.01	.3732	.5290	.6036	.6608	.7094					
1.0	.6404	.8118	.8840	.9246	.9550					
10	.7306	.8856	.9390	.9692	.9886					

Example 5.3. **Linear Regression.**

The true transformation parameter used in this example is λ=0.06, and the same set of model parameters as in Example 5.1 are used. The same levels of variance and significance as in previous examples are chosen. Corresponding to the sample sizes 10 and 20, we have the design matrices

$$X^T=\begin{bmatrix} 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. \\ 80. & 80. & 75. & 62. & 62. & 62. & 62. & 62. & 58. & 58. \\ 27. & 27. & 25. & 24. & 22. & 23. & 24. & 24. & 23. & 18. \\ 89. & 88. & 90. & 87. & 87. & 87. & 93. & 93. & 87. & 80. \end{bmatrix}$$

and

$$X^T =$$

$$\begin{bmatrix} 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. \\ 80. & 80. & 75. & 62. & 62. & 62. & 62. & 62. & 58. & 58. & 58. & 58. & 58. & 58. & 50. & 50. & 50. & 50. & 50. & 56. \\ 27. & 27. & 25. & 24. & 22. & 23. & 24. & 24. & 23. & 18. & 18. & 17. & 18. & 19. & 18. & 18. & 19. & 19. & 20. & 20. \\ 89. & 88. & 90. & 87. & 87. & 87. & 93. & 93. & 87. & 80. & 89. & 88. & 82. & 93. & 89. & 86. & 72. & 79. & 80. & 82. \end{bmatrix}.$$

The simulation results are given in Table 5.3 and similar results to example 5.1 are observed.

Table 5.3. Simulation results for regression model

$\sigma^2(\lambda)$	$n = 10$					$n = 20$				
			$1-\alpha$					$1-\alpha$		
	.75	.90	.95	.975	.99	.75	.90	.95	.975	.99
.0001	.7486	.8958	.9528	.9754	.9912	.7466	.8978	.9526	.9776	.9912
.01	.7360	.8974	.9468	.9720	.9892	.7352	.8946	.9486	.9748	.9914
1.0	.7344	.8858	.9450	.9732	.9894	.7350	.8936	.9474	.9716	.9888
10	.7356	.8918	.9458	.9736	.9906	.7472	.9000	.9512	.9744	.9904

A Check on the Accuracy of the Approximation (5.21)

So far we have used the second-order approximate asymptotic expansion to $\mathcal{F}(\hat{\lambda})$ in (5.21) in simulating the coverage probabilities of the usual confidence regions. We believe that this approximation should be good in general but a check on the accuracy is desirable. A method for doing this is described as follows. We simulate both $\mathcal{F}(\hat{\lambda})$ and the approximation (5.21). To generate the exact values of $\mathcal{F}(\hat{\lambda})$, we generate $\hat{\lambda}$ by a numerical root finding method and then calculate $\mathcal{F}(\hat{\lambda})$ by its expression. The detailed process for finding $\hat{\lambda}$ for a given set of e_i's is as follows. From (5.1) we have for each λ the solutions of $\sum\psi_{1i} = 0$ and $\sum\psi_{3i} = 0$ as

$$\hat{\beta}(\lambda) = (X^T X)^{-1} X^T h(y, \lambda) \quad \text{and} \quad \hat{\tau}^2(\lambda) = n / h^T(y, \lambda) Q h(y, \lambda).$$

Substituting $\hat{\beta}(\lambda)$ and $\hat{\tau}^2(\lambda)$ into $\sum\psi_{2i} = 0$ we have $\hat{\lambda}$ as the solution of

$$f(\lambda) = -n \frac{h^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)} + \frac{n}{\lambda} + \sum_{i=1}^{n} \log y_i = 0,$$

where $u(y, \lambda) = \{\lambda^{-1} y_i^{\lambda} \log y_i, i = 1, \ldots, n\}^T$.

To find the root of $f(\lambda)$ numerically, we need to know the derivative $f_\lambda(\lambda)$ of $f(\lambda)$. We have,

$$f_\lambda(\lambda) = -n \frac{\frac{d}{d\lambda}[h^T(y, \lambda) Q u(y, \lambda)]}{h^T(y, \lambda) Q h(y, \lambda)} + n \frac{h^T(y, \lambda) Q u(y, \lambda) \frac{d}{d\lambda}[h^T(y, \lambda) Q h(y, \lambda)]}{[h^T(y, \lambda) Q h(y, \lambda)]^2} - \frac{n}{\lambda^2}.$$

$$\begin{aligned}
&\frac{d}{d\lambda}[h^T(y, \lambda) Q u(y, \lambda)] \\
&= \sum_i \sum_j \frac{d}{d\lambda} h^T(y_i, \lambda) q_{ij} u(y_j, \lambda) \\
&= \sum_i \sum_j [\frac{d}{d\lambda} h^T(y_i, \lambda)] q_{ij} u(y_j, \lambda) + \sum_i \sum_j h^T(y_i, \lambda) q_{ij} \frac{d}{d\lambda} u(y_j, \lambda)
\end{aligned}$$

$$= \sum_i \sum_j [u(y_i, \lambda) - \frac{1}{\lambda} h(y_i, \lambda)] q_{ij}\, u(y_j, \lambda)$$

$$+ \sum_i \sum_j h^T(y_i, \lambda) q_{ij} [\frac{1}{\lambda} y_j^{\lambda} (\log y_j)^2 - \frac{1}{\lambda^2} y_j^{\lambda} \log y_j]$$

$$= [u(y, \lambda) - \frac{1}{\lambda} h(y, \lambda)]^T Q u(y, \lambda) + h^T(y, \lambda) Q [u(y, \lambda) \# \log y - \frac{1}{\lambda} u(y, \lambda)]$$

$$= u^T(y, \lambda) Q u(y, \lambda) + u^T(y, \lambda) Q [u(y, \lambda) \# \log y] - \frac{2}{\lambda} h^T(y, \lambda) Q u(y, \lambda),$$

and

$$\frac{d}{d\lambda} [h^T(y, \lambda) Q h(y, \lambda)] = 2\, h^T(y, \lambda) Q \frac{d}{d\lambda} h(y, \lambda)$$

$$= 2[h^T(y, \lambda) Q u(y, \lambda)] - \frac{2}{\lambda} [h^T(y, \lambda) Q y(y, \lambda)].$$

Putting them together, we have

$$f_\lambda(\lambda) = -n \frac{u^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)} - n \frac{y^T(y, \lambda) Q [u(y, \lambda) \# \log y]}{h^T(y, \lambda) Q h(y, \lambda)} - \frac{n}{\lambda^2}$$

$$+ \frac{2n}{\lambda} \frac{h^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)} + 2n \left[\frac{y^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)} \right]^2 - \frac{2n}{\lambda} \frac{y^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)}$$

$$= -n \frac{u^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)} - n \frac{h^T(y, \lambda) Q [u(y, \lambda) \# \log y]}{h^T(y, \lambda) Q h(y, \lambda)}$$

$$+ 2n \left[\frac{h^T(y, \lambda) Q u(y, \lambda)}{h^T(y, \lambda) Q h(y, \lambda)} \right]^2 - \frac{n}{\lambda^2}.$$

Example 5.4. **(Continued on Example 5.1)**

Let the model and parameters be defined as in Example 5.1 with n=27. In this example, we verify the accuracy of using asymptotic expansions to perform the simulations. Each of Figure 5.1 to Figure 5.4 is associate with a different error variance, and two graphs are plotted based on 100 pairs of random numbers from $\hat{\lambda}$ and another 100 pairs of random numbers from $\mathcal{F}(\hat{\lambda})$ using two different methods with this particular error variance. $\hat{\lambda}_{\text{Root}}$ and $\mathcal{F}_{\text{Root}}$ denote the numbers from root-finding method, and $\hat{\lambda}_{\text{Asymp}}$ and $\mathcal{F}_{\text{Asymp}}$ the asymptotic expansion method. The graphs are presented in the forms of $\hat{\lambda}_{\text{Asymp}}-\hat{\lambda}_{\text{Root}}$ versus $\hat{\lambda}_{\text{Root}}$ and $F_{\text{Asymp}}-F_{\text{Root}}$ versus F_{Root}. Note that $\hat{\lambda}_{\text{Asymp}} = \lambda + \lambda^*$ where λ^* is the first-order term in the approximate expansion for $\hat{\lambda}-\lambda$ given in Result 5.1. The second half of the plots show a good agreement between $\mathcal{F}_{\text{Asymp}}$ and $\mathcal{F}_{\text{Root}}$ particularly when $\sigma(\lambda)$ is small. The first half of the plots show good agreement between λ_{Asymp} and λ_{Root} when $\sigma(\lambda)$ is small but substantial difference when $\sigma(\lambda)$ is large. The approximation to λ_{Asymp} is first-order and the approximation to $\mathcal{F}_{\text{Asymp}}$ is second-order. Within the context of our discussion only the approximation to $\mathcal{F}_{\text{Asymp}}$ is of interest. The Figures were produced using ***DATADESK*** on a Macintosh II machine.

5.5. Directions for Further Research

Most of our results concerns the unconditional behavior of the λ-unknown quantity, although some large sample heuristic arguments were given concerning the conditional behavior of that quantity. The expansions reveal that in some cases the approximation of the λ-unknown quantity by the λ-known quantity could be poor and hence the approximation of the conditional distribution of the λ-unknown quantity by the unconditional distribution of the λ-known quantity will typically be also poor. Good unconditional approximation does not always give a satisfactory conditional approximation. A formal method to investigate the conditional behavior of the λ-unknown quantity for small-to-moderate n is thus needed. The behavior of $\hat{\lambda}$ is studied only for large n. When n is moderate-to-small we need a higher order expansion which may be obtained based on our general settings of Chapter 4. Another problem is to apply our general theory to testing statistical hypothesis, in particular to show that, for a given contrast matrix C, $C\beta(\hat{\lambda})=0$ is (asymptotically) equivalent to $C\beta(\lambda)=0$, hence the test based on (1.7) when λ is unknown has (asymptotically) the same size as the test based on (1 5) when λ is known.

Figure 5.1. Plots of $\hat{\lambda}_{\text{Asymp}} - \hat{\lambda}_{\text{Root}}$ versus $\hat{\lambda}_{\text{Root}}$ and $F_{\text{Asymp}} - F_{\text{Root}}$ versus F_{Root} for $\sigma = 0.01$.

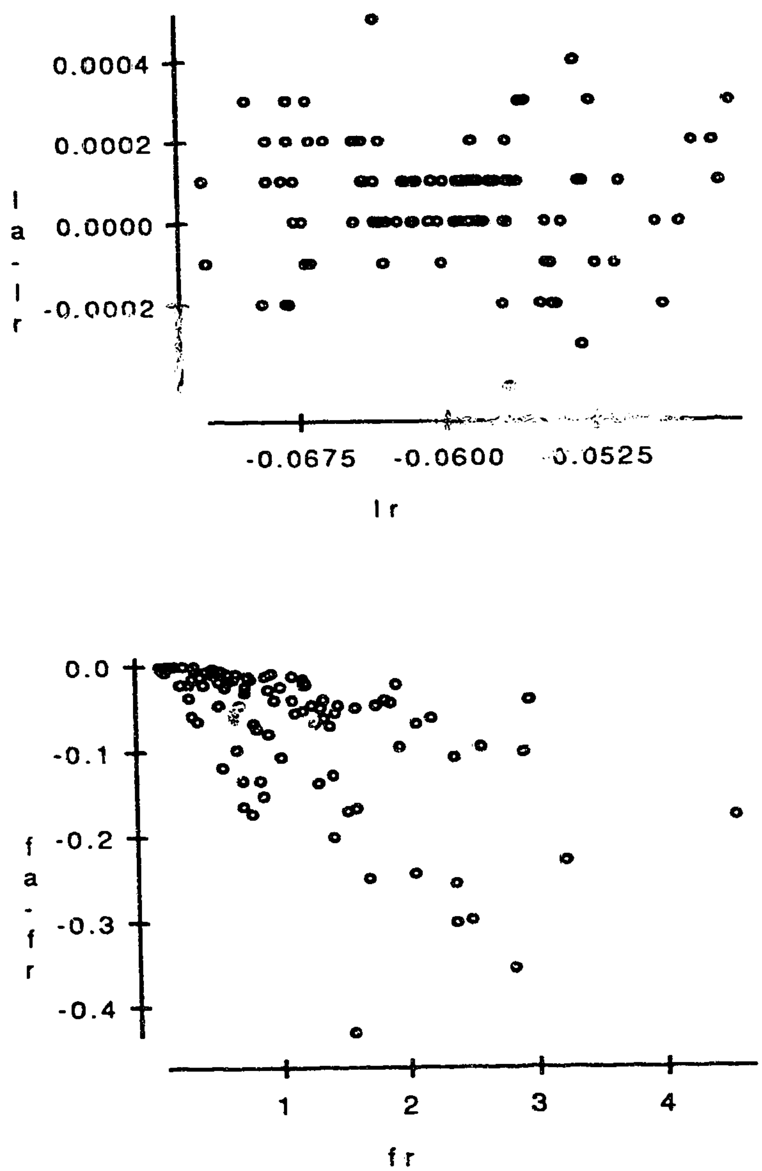
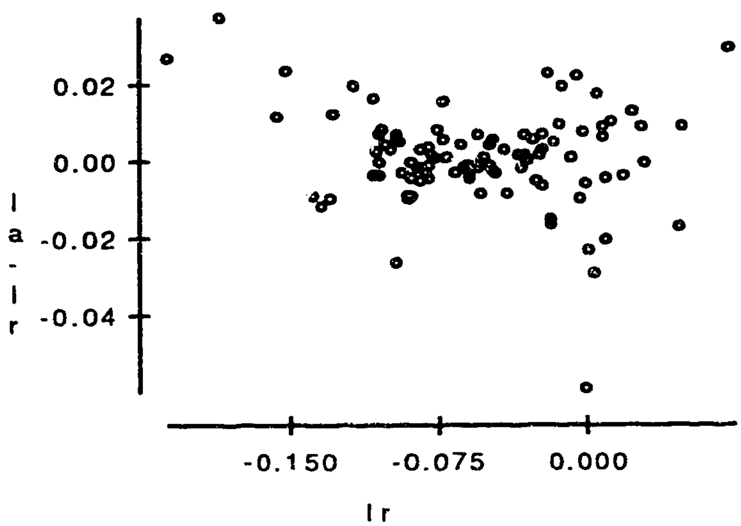


Figure 5.2. Plots of $\hat{\lambda}_{Asymp} - \hat{\lambda}_{Root}$ versus $\hat{\lambda}_{Root}$ and $F_{Asymp} - F_{Root}$ versus F_{Root} for $\sigma = 0.10$.



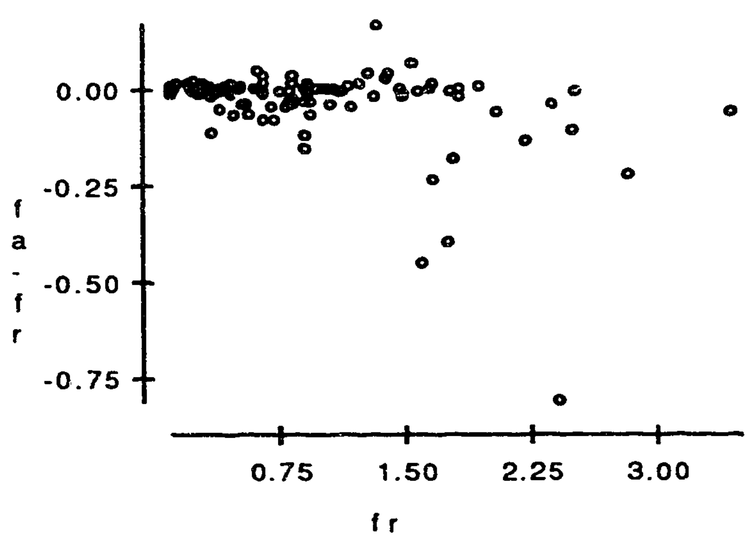
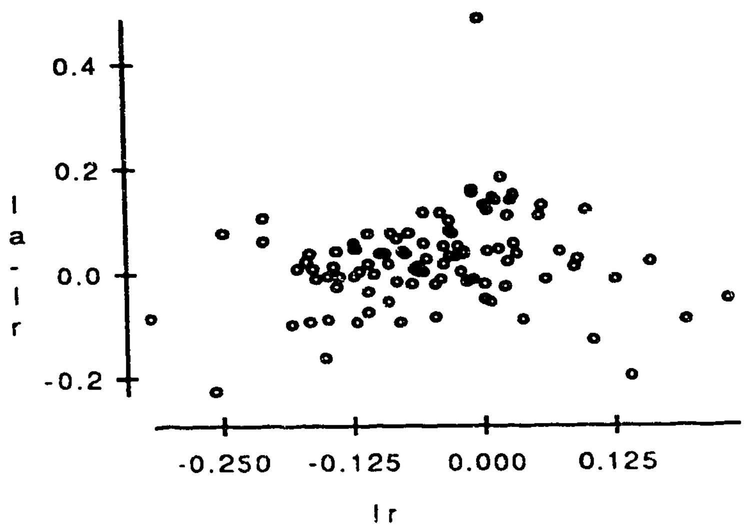


Figure 5.3. Plots of $\hat{\lambda}_{\text{Asymp}} - \hat{\lambda}_{\text{Root}}$ versus $\hat{\lambda}_{\text{Root}}$ and $F_{\text{Asymp}} - F_{\text{Root}}$ versus F_{Root} for $\sigma = 1.00$.



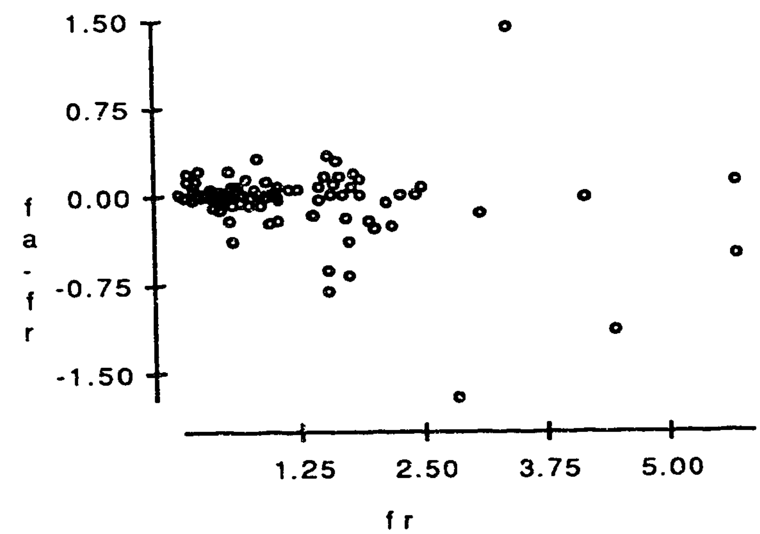
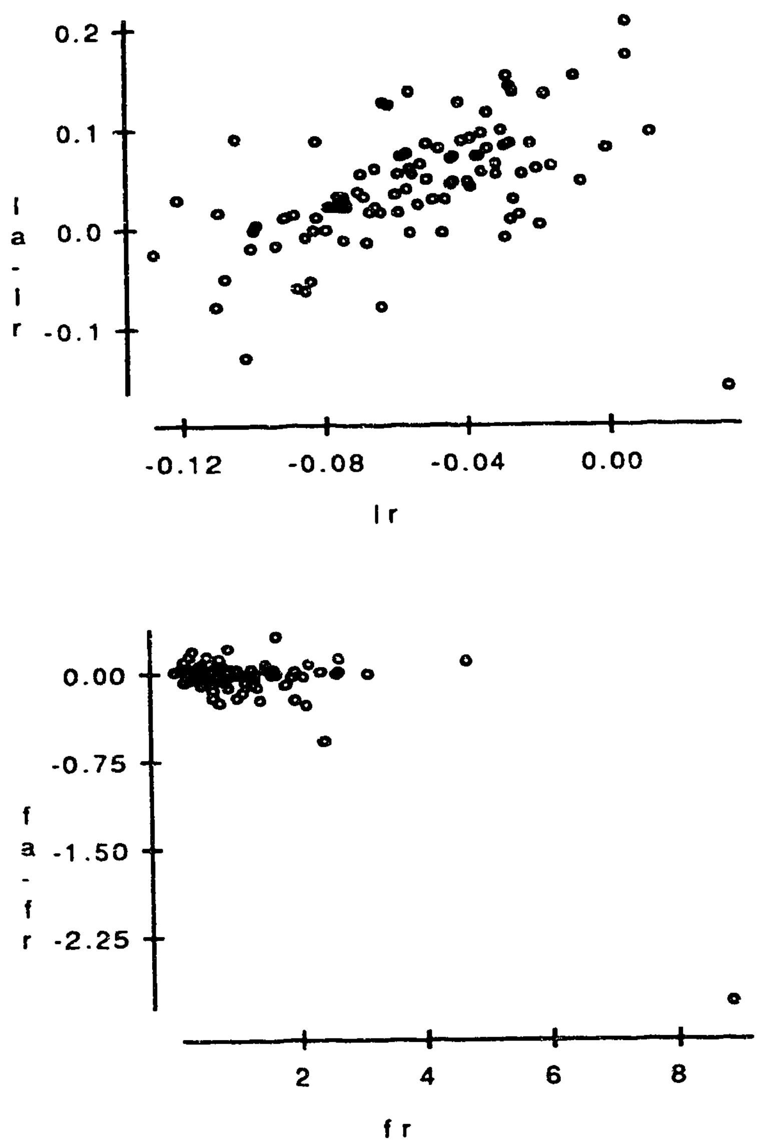


Figure 5.4. Plots of $\hat{\lambda}_{\text{Asymp}} - \hat{\lambda}_{\text{Root}}$ versus $\hat{\lambda}_{\text{Root}}$ and $F_{\text{Asymp}} - F_{\text{Root}}$ versus F_{Root} for $\sigma = 3.33$.



CHAPTER 6

ROBUSTNESS WHEN THE TRANSFORMATION IS SPECIFIED A PRIORI

6.1. Introduction

After selecting a transformation, one usually fits a normal-theory linear model to the transformed data as if the selected transformation were known in advance. There are various ways to select the transformation; e.g., maximum likelihood estimation of λ, rounding off the MLE to a convenient value, and a priori specification of a transformation. The first method has been investigated in the previous chapters. Evaluation of second method is more difficult since $\hat{\lambda}$ depends partially on the data and partially on prior information. The second method is, however, asymptotically equivalent to the third provided the set of "convenient values" is fixed as $n \to \infty$. In this chapter we study the third method, where the transformation is selected based on information independent of the data. The problem of parameter interpretation is less controversial here since one is typically interested in parameters related to the distribution of $h(y,\hat{\lambda})$. Since $\hat{\lambda}$ is constant, evaluation of the Box-Cox analysis now involves only the issues of robustness. We examine the robustness of normal-theory inference assuming a model of the form (1.2) but allowing the specified $\hat{\lambda}$ to differ from λ. Once again we assume that the parameters of interest are those associated with the selected transformation. In Section 6.2 we study the effects of misspecification of λ on the usual F-ratio and hence on the usual confidence region for $\beta(\hat{\lambda})$. Section 6.3 contains some simulation results supporting conclusions reached in Section 6.2. In Section 6.4 we describe a method for approximating coverage probabilities of the usual confidence regions.

Our conclusions regarding misspecification of λ may be summarized as follows: i) in structured model the confidence region is conservative for small $\sigma(\lambda)$;

ii) in unstructured model the effect of misspecification for small $\sigma(\lambda)$ is mainly increased heteroscedasticity.

6.2. The Effect of Misspecification of a Transformation

Let $\hat{\lambda}$ now be a fixed number representing the selected transformation based on prior information. The estimate $\hat{\lambda}$ is independent of the data y, the sample size n, and the underlining model parameters $\beta(\lambda)$ and $\sigma(\lambda)$. The results in the previous chapters require that $\hat{\lambda}$ be a root-n consistent estimator of λ and hence are not applicable here.

From the results of previous chapters we known that when $\hat{\lambda}$ is estimated from data we are able to estimate λ more precisely if we have a large data set or a certain structured model with small error variance. Hence the effect of estimating λ in those cases is small. The effect when $\hat{\lambda}$ is specified a priori depends on how much prior information we have. If we have relatively poor information when we have a large data set or a certain structured model with small error variance, the pre-specified value may not be accurate at all. In other words, in some cases one should be very careful on using the preassigned value $\hat{\lambda}$ and in some other cases one can rely on it. We will study those matters theoretically based on the least-squares estimators.

Consider the model (1.2), $z \equiv h(y, \lambda) = X\beta(\lambda) + \sigma(\lambda)e(\lambda)$, with a Box-Cox power transformation and approximate i.i.d. normal errors. The parameters associated with the selected transformation $\hat{\lambda}$ are defined as follows

$$\beta_u(\hat{\lambda}) = (X^TX)^{-1}X^T \, \mathrm{E}h(y, \hat{\lambda}) = (X^TX)^{-1}X^T\mathrm{E}g(z, \hat{\lambda}, \lambda), \qquad (6.1)$$

where

$$g(z, \hat{\lambda}, \lambda) = h(h^{-1}(z, \lambda), \hat{\lambda}) = h(y, \hat{\lambda}),$$

and the expression for the g function is given by (5.11).

After selecting the transformation $\hat{\lambda}$, we fit the model

$$h(y, \hat{\lambda}) = X\beta_u(\hat{\lambda}) + \sigma(\hat{\lambda})e(\hat{\lambda}), \tag{6.2}$$

and make inferences on $\beta_u(\hat{\lambda})$ pretending that $\hat{\lambda} = \lambda$. The least-squares estimates of $\beta_u(\hat{\lambda})$ and $\sigma(\hat{\lambda})$ are

$$\hat{\beta}(\hat{\lambda}) = (X^T X)^{-1} X^T h(y, \hat{\lambda}) = (X^T X)^{-1} X^T g(z, \hat{\lambda}, \lambda), \tag{6.3}$$

and

$$\begin{aligned} \hat{\sigma}(\hat{\lambda}) &= n^{-1/2} \| h(y, \hat{\lambda}) - X\hat{\beta}(\hat{\lambda}) \| \\ &= n^{-1/2} \| Qg(z, \hat{\lambda}, \lambda) \|, \end{aligned} \tag{6.4}$$

where

$$Q = I_n - X(X^T X)^{-1} X^T.$$

The usual F-ratio, i.e., the F-ratio obtained pretending $\hat{\lambda} = \lambda$, is

$$\mathcal{F}(\hat{\lambda}) = \frac{\| X[\hat{\beta}(\hat{\lambda}) - \beta_u(\hat{\lambda})] \|^2}{\| h(y, \hat{\lambda}) - X\hat{\beta}(\hat{\lambda}) \|^2} \cdot \frac{(n-p)}{p}. \tag{6.5}$$

Note that when $\hat{\lambda} = \lambda$, $\mathcal{F}(\hat{\lambda})$ has an approximate F-distribution with p and $n-p$ degrees of freedom in numerator and denominator respectively. The usual confidence region for $\beta_u(\hat{\lambda})$ is obtained from $\mathcal{F}(\hat{\lambda})$ pretending $\hat{\lambda} = \lambda$. The validity of treating the distribution of $\mathcal{F}(\hat{\lambda})$ as an F-distribution is investigated in this section. Now writing $\mathcal{F}(\hat{\lambda})$ in terms of the g function, we have

$$\mathcal{F}(\hat{\lambda}) = \frac{\| P[g(z, \hat{\lambda}, \lambda) - \mathrm{E}g(z, \hat{\lambda}, \lambda)] \|^2}{\| Qg(z, \hat{\lambda}, \lambda) \|^2} \cdot \frac{(n-p)}{p}, \tag{6.6}$$

where

$$P = X(X^T X)^{-1} X^T.$$

Using a second-order Taylor approximation to the g function around $\eta = X\beta(\lambda)$, i.e.,

$$g(z, \hat{\lambda}, \lambda) \approx g(\eta, \hat{\lambda}, \lambda) + \sigma(\lambda) g_\eta(\eta, \hat{\lambda}, \lambda) \# e + \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta, \hat{\lambda}, \lambda) \# e^2, \quad (6.7)$$

we have

$$\mathcal{F}(\hat{\lambda}) \approx \frac{\| P[\sigma(\lambda) g_\eta(\eta, \hat{\lambda}, \lambda) \# e + \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta, \hat{\lambda}, \lambda) \# (e^2 - 1)] \|^2}{\| Q[g(\eta, \hat{\lambda}, \lambda) + \sigma(\lambda) g_\eta(\eta, \hat{\lambda}, \lambda) \# e + \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta, \hat{\lambda}, \lambda) \# e^2] \|^2} \bullet \frac{(n-p)}{p}. \quad (6.8)$$

Note that the n-vectors $g(\eta, \hat{\lambda}, \lambda)$, $g_\eta(\eta, \hat{\lambda}, \lambda)$ and $g_{\eta\eta}(\eta, \hat{\lambda}, \lambda)$ in (6.8) are all independent of $\sigma(\lambda)$ and $e(\lambda)$.

Result 6.1. Under the approximation (6.7) we have

i) In structured model, for each e, $\mathcal{F}(\hat{\lambda}) \to 0$ as $\sigma(\lambda) \to 0$;

ii) In unstructured model, for each e,

$$\mathcal{F}(\hat{\lambda}) \to \frac{\| P[g_\eta(\eta, \hat{\lambda}, \lambda) \# e] \|^2}{\| Q[g_\eta(\eta, \hat{\lambda}, \lambda) \# e] \|^2} \bullet \frac{(n-p)}{p}$$

as $\sigma(\lambda) \to 0$.

Proof. If the model has structure, then for $\hat{\lambda} \neq \lambda$, $Qg(\eta, \hat{\lambda}, \lambda) \neq 0$. The $g(\eta, \hat{\lambda}, \lambda)$ is the only term in (6.8) not involving $\sigma(\lambda)$, hence $\mathcal{F}(\hat{\lambda}) \to 0$ as $\sigma(\lambda) \to 0$. If the model does not have structure, we have $Qg(\eta, \hat{\lambda}, \lambda) = 0$ and

$$\mathcal{F}(\hat{\lambda}) \approx \frac{\| P[g_\eta(\eta, \hat{\lambda}, \lambda) \# e + \frac{\sigma(\lambda)}{2} g_{\eta\eta}(\eta, \hat{\lambda}, \lambda) \# (e^2 - 1)] \|^2}{\| Q[g_\eta(\eta, \hat{\lambda}, \lambda) \# e + \frac{\sigma(\lambda)}{2} g_{\eta\eta}(\eta, \hat{\lambda}, \lambda) \# (e^2 - 1)] \|^2} \bullet \frac{(n-p)}{p}, \quad (6.9)$$

Letting $\sigma(\lambda) \to 0$ in (6.9) yields the result.

From (6.8) we also see that when $\sigma(\lambda)$ is la[illegible],

$$\mathcal{F}(\hat{\lambda}) \approx \frac{\|P[g_{\eta\eta}(\eta, \hat{\lambda}, \lambda)\#(e^2-1)]\|^2}{\|Q[g_{\eta\eta}(\eta, \hat{\lambda}, \lambda)\#(e^2-1)]\|^2} \cdot \frac{(n-p)}{p}, \tag{6.10}$$

for both structured and unstructured model. But (6.10) should be treated with caution since the approximation (6.7) may be poor in this case.

The implications of Result 6.1 are as follows. If the model has structure the usual confidence region for $\beta(\hat{\lambda})$ is conservative when $\sigma(\lambda)$ is small. In unstructured models with small $\sigma(\lambda)$, misspecification of λ introduces heteroscedasticity which depends on the derivatives $g_\eta(\eta_i, \hat{\lambda}, \lambda)$. The effect should be small if those derivatives are all close together. So the structured model is more sensitive to the departure of $\hat{\lambda}$ from λ than the unstructured model. This is in contrast to the conclusions drawn in the previous chapters where $\hat{\lambda}$ is estimated from data. The reason is that in the structured model the data are able to provide a accurate estimate of λ, whereas in the unstructured model it is not, particularly when $\sigma(\lambda)$ is small and the means are close together. We will investigate these conclusions using Mont Carlo simulations.

6.3. Simulations

The usual $100(1-\alpha)\%$ confidence region for $\beta(\hat{\lambda})$ is given by

$$C_\alpha(\hat{\lambda}) = \{\ \mu_u(\hat{\lambda}):\ \mathcal{F}(\hat{\lambda}) \le F_{p,\,n-p}(1-\alpha)\ \}, \tag{6.11}$$

where $\mathcal{F}(\hat{\lambda})$ is given by (6.6). In our simulations, $E\{g(z, \hat{\lambda}, \lambda)\}$ was approximated by $g(\eta, \hat{\lambda}, \lambda) + \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta, \hat{\lambda}, \lambda)$.

Two examples are considered in simulation study. Example 6.1 uses a structured model and Example 6.2 uses an unstructured model. In both examples, 5000 replication of $\mathcal{F}(\hat{\lambda})$ were generated for each combination of the selected values for $\sigma(\lambda)$ and $\hat{\lambda}$, hence the approximate standard errors of the simulated coverage probabilities are .0061, .0042, .0031, .0023, and .0014, respectively, corresponding to the coverage probabilities .75, .90, .95, .975, and .99.

Example 6.1. 3^3 factorial design with additive effects and $\lambda = -0.06$

In this example we simulated the coverage probabilities of the confidence region (6.11) using a 3^3 factorial design with additive effects. The design matrix is given in Example 5.1. The grand mean and the main effects are also given in Example 5.1. Seven values of $\hat{\lambda}$, four levels of $\sigma^2(\lambda)$ and five coverage probabilities were chosen. From the results in Table 6.1 we see that the usual confidence region is conservative when $\sigma(\lambda)$ is small and $\hat{\lambda}$ differs from λ but otherwise performs well. This agrees with the conclusions in Section 6.2.

Example 6.2. One-way layout with $n = 18$, $p = 3$, and $\lambda = 0.10$

In this example we simulated coverage probabilities of the confidence region (6.11) using a one-way layout. We choose nine values of $\hat{\lambda}$, four values of $\sigma^2(\lambda)$, and five coverage probabilities. We also select two different sets of means, one set of the means are very near to each other, the other more dispersed. The results are given in Table 6.2 and 6.3. As predicted in Section 6.2, the effect of a wrong transformation is not serious.

Table 6.1. Simulation results for 3^3 factorial design with additive effects

$\sigma^2(\lambda)$	$1-\alpha$=.75	.90	.95	.975	.99	$1-\alpha$=.75	.90	.95	.975	.99
	$\hat{\lambda} = -0.12$					$\hat{\lambda} = 0.001$				
.0001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.01	.7682	.9074	.9536	.9752	.9884	.7678	.9102	.9558	.9764	.9916
1.0	.7540	.8986	.9520	.9758	.9904	.7496	.8986	.9514	.9766	.9910
4.0	.7438	.8960	.9484	.9730	.9876	.7506	.8974	.9494	.9746	.9874
	$\hat{\lambda} = -0.09$					$\hat{\lambda} = -0.03$				
.0001	.9670	.9944	.9990	.9998	1.000	.9724	.9950	.9992	1.000	1.000
.01	.7574	.9056	.9566	.9790	.9906	.7526	.9062	.9532	.9780	.9922
1.0	.7602	.9016	.9526	.9788	.9910	.7428	.8876	.9410	.9682	.9880
4.0	.7514	.8974	.9506	.9744	.9914	.7444	.8952	.9490	.9742	.9868
	$\hat{\lambda} = -0.065$					$\hat{\lambda} = -0.055$				
.0001	.7606	.9024	.9544	.9782	.9894	.7488	.8992	.9512	.9772	.9884
.01	.7440	.9042	.9542	.9794	.9924	.7550	.9086	.9510	.9756	.9906
1.0	.7504	.8984	.9476	.9722	.9906	.7572	.9014	.9508	.9784	.9914
4.0	.7488	.9014	.9506	.9762	.9900	.7520	.8982	.9520	.9744	.9888
	$\hat{\lambda} = -0.06$									
.0001	.7468	.8998	.9488	.9734	.9888					
.01	.7552	.9042	.9492	.9742	.9894					
1.0	.7544	.9028	.9500	.9738	.9900					
4.0	.7504	.8968	.9518	.9762	.9906					

Table 6.2. Simulation results for one-way layout with $n = 18$, $p = 3$, and the three means 10, 8, and 6.

$\sigma^2(\lambda)$	$1-\alpha$=.75	.90	.95	.975	.99	$1-\alpha$=.75	.90	.95	.975	.99
	$\hat{\lambda} = -0.10$					$\hat{\lambda} = 0.30$				
.0001	.6918	.8532	.9108	.9480	.9734	.7464	.8864	.9418	.9676	.9858
.01	.7396	.8948	.9474	.9722	.9880	.7426	.8876	.9390	.9672	.9852
1.0	.7428	.8868	.9388	.9670	.9830	.7334	.8926	.9380	.9676	.9854
4.0	.7388	.8896	.9422	.9672	.9854	.7494	.8880	.9374	.9670	.9840
	$\hat{\lambda} = 0.001$					$\hat{\lambda} = 0.20$				
.0001	.7192	.8730	.9324	.9600	.9824	.7446	.8984	.9492	.9752	.9892
.01	.7428	.8936	.9426	.9706	.9898	.7514	.8992	.9500	.9740	.9902
1.0	.7462	.8944	.9448	.9724	.9876	.7424	.8966	.9458	.9704	.9900
4.0	.7506	.8994	.9510	.9762	.9912	.7518	.9000	.9516	.9758	.9902
	$\hat{\lambda} = 0.05$					$\hat{\lambda} = 0.15$				
.0001	.7170	.8776	.9338	.9644	.9834	.7290	.8926	.9406	.9700	.9848
.01	.7472	.8950	.9468	.9724	.9886	.7536	.8992	.9454	.9736	.9900
1.0	.7552	.9082	.9576	.9794	.9908	.7498	.8998	.9484	.9734	.9902
4.0	.7486	.9018	.9538	.9788	.9924	.7482	.8954	.9472	.9744	.9902
	$\hat{\lambda} = 0.09$					$\hat{\lambda} = 0.11$				
.0001	.7236	.8880	.9430	.9680	.9866	.7238	.8882	.9394	.9648	.9860
.01	.7452	.8930	.9460	.9706	.9870	.7546	.9016	.9538	.9788	.9910
1.0	.7530	.8956	.9468	.9726	.9888	.7482	.8992	.9488	.9748	.9900
4.0	.7370	.8966	.9496	.9726	.9892	.7490	.8960	.9438	.9734	.9896
	$\hat{\lambda} = 0.10$									
.0001	.7334	.8896	.9418	.9714	.9870					
.01	.7466	.8918	.9460	.9722	.9876					
1.0	.7532	.9052	.9522	.9750	.9896					
4.0	.7388	.9014	.9526	.9764	.9908					

Table 6.3. Simulation results for one-way layout with $n = 18$, $p = 3$, and the three means 8.5, 8.0, and 7.5.

$\sigma^2(\lambda)$	$1-\alpha$=.75	.90	.95	.975	.99	$1-\alpha$=.75	.90	.95	.975	.99
	$\hat{\lambda} = -0.10$					$\hat{\lambda} = 0.30$				
.0001	.6880	.8552	.9168	.9502	.9746	.7358	.8970	.9456	.9718	.9912
.01	.7520	.9062	.9506	.9758	.9900	.7508	.9024	.9522	.9756	.9918
1.0	.7462	.8968	.9468	.9728	.9890	.7470	.8972	.9484	.9758	.9892
4.0	.7496	.8976	.9504	.9732	.9882	.7472	.8994	.9512	.9758	.9896
	$\hat{\lambda} = 0.001$					$\hat{\lambda} = 0.20$				
.0001	.7226	.8766	.9284	.9628	.9848	.7450	.8874	.9400	.9710	.9886
.01	.7408	.8994	.9476	.9732	.9888	.7526	.8988	.9492	.9738	.9890
1.0	.7570	.8986	.9524	.9750	.9900	.7534	.9034	.9534	.9778	.9924
4.0	.7392	.8958	.9458	.9728	.9878	.7478	.9018	.9482	.9732	.9896
	$\hat{\lambda} = 0.05$					$\hat{\lambda} = 0.15$				
.0001	.7120	.8814	.9396	.9686	.9878	.7334	.8884	.9390	.9660	.9852
.01	.7572	.9068	.9582	.9786	.9916	.7452	.9028	.9522	.9764	.9904
1.0	.7482	.8968	.9482	.9748	.9902	.7566	.8926	.9472	.9756	.9908
4.0	.7426	.8984	.9488	.9744	.9914	.7476	.8940	.9450	.9728	.9896
	$\hat{\lambda} = 0.09$					$\hat{\lambda} = 0.11$				
.0001	.7318	.8894	.9412	.9692	.9884	.7340	.8932	.9440	.9696	.9896
.01	.7480	.9020	.9480	.9736	.9898	.7466	.9014	.9544	.9762	.9904
1.0	.7518	.9010	.9468	.9730	.9886	.7402	.8960	.9510	.9736	.9896
4.0	.7592	.9086	.9560	.9808	.9926	.7464	.9064	.9548	.9754	.9888
	$\hat{\lambda} = 0.10$									
.0001	.7384	.8948	.9464	.9742	.9882					
.01	.7388	.8944	.9452	.9722	.9878					
1.0	.7472	.9030	.9548	.9792	.9914					
4.0	.7578	.9046	.9520	.9766	.9900					

6.4. A Method for Approximating Coverage Probabilities

In this section we describe an analytic method for approximating the coverage probability $P\{C_\alpha(\hat{\lambda})\}$ instead of Monte Carlo simulations. The basic idea is to choose a random quantity with an known distribution to approximate the random quantity of interest with an unknown distribution by matching moments. Since a detailed derivation would be lengthy and since the results are not directly related to the main results of the thesis, only a sketch is given here.

Let

$$S_N(\hat{\lambda}) = \|P[g(z, \hat{\lambda}, \lambda) - Eg(z, \hat{\lambda}, \lambda)\|^2,$$

and

$$S_D(\hat{\lambda}) = \|Qg(z, \hat{\lambda}, \lambda)\|^2,$$

then

$$\begin{aligned} P\{C_\alpha(\hat{\lambda})\} &= P\{\mathcal{F}(\hat{\lambda}) \le F_{p,\, n-p}(1-\alpha)\} \\ &= P\left\{\frac{S_N(\hat{\lambda})}{S_D(\hat{\lambda})} \bullet \frac{(n-p)}{p} \le F_{p,\, n-p}(1-\alpha)\right\} \\ &= P\{S_N(\hat{\lambda}) + aS_D(\hat{\lambda}) \le 0\} \end{aligned} \tag{6.12}$$

where $a = -pF_{p,\, n-p}(1-\alpha)/(n-p)$.

Calculation of (6.12) requires the distribution of $S \equiv S_N(\hat{\lambda}) + aS_D(\hat{\lambda})$. The exact distribution of S is not available, but the moments of S may be written down exactly. We evaluate the first four moments of S and then approximate the distribution of S by a distribution which has the same first four moments. Calculation of the first four moments of S requires the first four moments and product-moments of $S_N(\hat{\lambda})$ and $S_D(\hat{\lambda})$ and the latter require the first eight moments of $g(z_i,\hat{\lambda},\lambda)-Eg(z_i,\hat{\lambda},\lambda)$, $i = 1, \ldots, n$.

Putting $e_i^* = g(z_i,\hat{\lambda},\lambda)-Eg(z_i,\hat{\lambda},\lambda)$, then by approximation (6.7) we obtain

$$e_i^* \approx \sigma(\lambda)g_\eta(\eta_i, \hat{\lambda}, \lambda)e_i + \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda)(e_i^2 - 1)$$

and

$$S_N(\hat{\lambda}) \approx \sum_{i=1}^{n}\sum_{i=1}^{n} p_{ij} e_i^* e_j^*, \; S_D(\hat{\lambda}) \approx \sum_{i=1}^{n}\sum_{i=1}^{n} q_{ij}(e_i^* + D_i)(e_j^* + D_j),$$

where

$$D_i = g(\eta_i, \hat{\lambda}, \lambda) + \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda), \; i = 1, \ldots, n,$$

p_{ij} is the ijth element of P, and q_{ij} is the ijth element of Q.

Using ***Mathematica*** (a computer language which does symbolic manipulations), we can easily obtain the moment generating function, the first eight moments, and the first eight cumulants of e_i^* :

Moment generating function of e_i^* ,

$$\begin{aligned} \mathrm{M}(t) &= [1 - \sigma^2(\lambda)g_{\eta\eta}(\eta_i, \lambda_1, \lambda)t]^{-1/2} \\ &\quad \bullet \mathrm{Exp}\left[\frac{[\sigma(\lambda)g_\eta(\eta_i, \hat{\lambda}, \lambda)\mathrm{t}]^2}{2[1 - \sigma^2(\lambda)g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda)t]} - \frac{\sigma^2(\lambda)}{2} g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda)t\right]; \end{aligned}$$

The first eight moments of e_i^* ,

$$\begin{aligned} \alpha_{1i} &= 0, \\ \alpha_{2i} &= \sigma^2(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda) + \frac{\sigma^4(\lambda)}{2} g_{\eta\eta}^2(\eta_i, \hat{\lambda}, \lambda), \\ \alpha_{3i} &= 3\sigma^4(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda) + \sigma^6(\lambda)g_{\eta\eta}^6(\eta_i, \hat{\lambda}, \lambda), \\ \alpha_{4i} &= 3\sigma^4(\lambda)g_\eta^4(\eta_i, \hat{\lambda}, \lambda) + 15\sigma^6(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^2(\eta_i, \hat{\lambda}, \lambda) \\ &\quad + \frac{15}{4}\sigma^8(\lambda)g_{\eta\eta}^4(\eta_i, \hat{\lambda}, \lambda), \\ \alpha_{5i} &= 30\sigma^6(\lambda)g_\eta^4(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda) + 85\sigma^8(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^3(\eta_i, \hat{\lambda}, \lambda) \\ &\quad + 17\sigma^{10}(\lambda)g_{\eta\eta}^5(\eta_i, \hat{\lambda}, \lambda), \\ \alpha_{6i} &= 15\sigma^6(\lambda)g_\eta^6(\eta_i, \hat{\lambda}, \lambda) + \frac{585}{2}\sigma^8(\lambda)g_\eta^4(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^2(\eta_i, \hat{\lambda}, \lambda) \\ &\quad + \frac{2265}{4}\sigma^{10}(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^4(\eta_i, \hat{\lambda}, \lambda) + \frac{755}{8}\sigma^{12}(\lambda)g_{\eta\eta}^6(\eta_i, \hat{\lambda}, \lambda), \end{aligned}$$

$$\alpha_{7i} = 315\sigma^8(\lambda)g_\eta^6(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda) + 2940\sigma^{10}(\lambda)g_\eta^4(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^3(\eta_i, \hat{\lambda}, \lambda) + \frac{17283}{4}\sigma^{12}(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^5(\eta_i, \hat{\lambda}, \lambda) + \frac{2469}{4}\sigma^{14}(\lambda)g_{\eta\eta}^7(\eta_i, \hat{\lambda}, \lambda),$$

$$\alpha_{8i} = 105\sigma^8(\lambda)g_\eta^8(\eta_i, \hat{\lambda}, \lambda) + 5250\,\sigma^{10}(\lambda)g_\eta^6(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^2(\eta_i, \hat{\lambda}, \lambda) + \frac{62895}{2}\sigma^{12}(\lambda)g_\eta^4(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^4(\eta_i, \hat{\lambda}, \lambda) + \frac{74417}{2}\sigma^{14}(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^6(\eta_i, \hat{\lambda}, \lambda) + \frac{74417}{16}\sigma^{16}(\lambda)g_{\eta\eta}^8(\eta_i, \hat{\lambda}, \lambda);$$

The first eight cumulants of e_i^* ,

$$\kappa_{1\alpha} = 0$$

$$\kappa_{2\alpha} = \sigma^2(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda) + \frac{\sigma^4(\lambda)}{2}g_{\eta\eta}^2(\eta_i, \hat{\lambda}, \lambda),$$

$$\kappa_{3\alpha} = 3\sigma^4(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}(\eta_i, \hat{\lambda}, \lambda) + \sigma^6(\lambda)g_{\eta\eta}^6(\eta_i, \hat{\lambda}, \lambda),$$

$$\kappa_{4\alpha} = 12\sigma^6(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^2(\eta_i, \hat{\lambda}, \lambda) + 3\sigma^8(\lambda)g_{\eta\eta}^4(\eta_i, \hat{\lambda}, \lambda),$$

$$\kappa_{5\alpha} = 60\sigma^8(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^3(\eta_i, \hat{\lambda}, \lambda) + 12\sigma^{10}(\lambda)g_{\eta\eta}^5(\eta_i, \hat{\lambda}, \lambda),$$

$$\kappa_{6\alpha} = 360\sigma^{10}(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^4(\eta_i, \hat{\lambda}, \lambda) + 60\sigma^{12}(\lambda)g_{\eta\eta}^6(\eta_i, \hat{\lambda}, \lambda),$$

$$\kappa_{7\alpha} = 2520\sigma^{12}(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^5(\eta_i, \hat{\lambda}, \lambda) + 360\sigma^{14}(\lambda)g_{\eta\eta}^7(\eta_i, \hat{\lambda}, \lambda)$$

$$\kappa_{8\alpha} = 20160\sigma^{14}(\lambda)g_\eta^2(\eta_i, \hat{\lambda}, \lambda)g_{\eta\eta}^6(\eta_i, \hat{\lambda}, \lambda) + 2520\sigma^{16}(\lambda)g_{\eta\eta}^8(\eta_i, \hat{\lambda}, \lambda).$$

Now using the formulas in Section 4 and 5 of David and Johnson (1951), we can calculate the first four cumulants and product-cumulants of $S_N(\hat{\lambda})$ and $S_D(\hat{\lambda})$ and then convert them into the first four cumulants of S. Having obtained the first four cumulants of S, we select a particular Pearson curve based on those four moments and calculate the corresponding probabilities to approximate $P(S \leq 0)$. For detailed procedure on selecting Pearson curve type, see Springer (1979, p. 255) and Stuart and Ord (1987, p. 210).

BIBLIOGRAPHY

BICKEL, P.J. (1984), Comments on Hinkley and Runger (1984), *Journal of the American Statistical Association*, 79, 315-316.

BICKEL, P.J., and DOKSUM, K.A. (1981), "An Analysis of Transformations Revisited," *Journal of the American Statistical Association*, 76, 296-311.

BOX, G.E.P., and COX, D.R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society*," Ser, 26, 211-252.

———— (1982), "An Analysis of Transformations Revisited, Rebutted" *Journal of the American Statistical Association*, 79, 302-320.

CARROLL, R.J. (1982), "Tests for Regression Parameters in Power Transformation Models," *Scandinavian Journal of Statistics*, 9, 217-222.

CARROLL, R.J., and RUPPERT, D. (1981), "On Prediction and the Power Transformation Family," *Biometrika*, 68, 609-616.

DAVID, F.N., and JOHNSON, N.L. (1951), "A Method of Investigating the Effect of Nonnormality and Heterogeneity of Variance on Tests of the General Linear Hypothesis," *Annals of Mathematical Statistics*, 22, 382-392.

DOKSUM, K.A. (1984), Comments on Hinkley and Runger (1984), *Journal of the American Statistical Association*, 79, 316-319.

DOKSUM, K.A., and WONG, C.W. (1983), "Statistical Tests Based on Transformed Data," *Journal of the American Statistical Association*, 78, 411-417.

HERNANDEZ, F. (1978), "The Large Sample Behavior of Transformations to Normal or Exponential Distributions," unpublished Ph D dissertation, University of Wisconsin—Madison.

HERNANDEZ, F., and JOHNSON, R.A. (1980), "The Large-Sample Behavior of Transformation to Normality,' *Journal of the American Statistical Association*, 75, 855-861.

HINKLEY, D.V. (1975), "On Power Transformations to Symmetry," *Biometrika*, 62, 101-111.

HINKLEY, D.V., and RUNGER, G. (1984), "The Analysis of Transformed Data," *Journal of the American Statistical Association*, 79, 302-320.

HOADLEY, B. (1971), "Asymptotic Properties of Maximum Likelihood Estimators for the Independent not Identically Distributed Case," *The Annals of Mathematical Statistics*, 42, 1977-1991.

RUBIN, D.B. (1984), Comments on Hinkley and Runger (1984), *Journal of the American Statistical Association*, 79, 309-312.

SPRINGER, M.D. (1979), *The Algebra of Random Variables*, New York: Wiley.

STUART, O. and ORD, J.K. (1987), *Kendall's Advanced Theory of Statistics*, Vol. 1, 5th Edition, London: Charles Griffin and Company Limited.

TAYLOR, J.M.G. (1985a), "Power transformations to Symmetry," *Biometrika*, 72, 145-152.

——— (1985b), "Measures of Location of Skew Distribution Obtained Through Box-Cox Transformations," *Journal of the American Statistical Association*, 80, 427-432.

——— (1986), "The Retransformed Mean After a Fitted Power Transformation," *Journal of the American Statistical Association*, 81, 114-118.

APPENDIX A

DETAILED DERIVATIONS AND PROOFS FOR CHAPTER 3

Derivation of (3.25). From (3.24) we have

$$S(\xi, y) =$$

$$\begin{cases} \dfrac{\partial \ell(\xi, y)}{\partial \beta_j} = \dfrac{1}{\sigma}\displaystyle\sum_{i=1}^{n}\left\{ x_{ij}\,[\varepsilon_i(y_i, \xi) - \sigma \dfrac{1}{c_i(\xi)}\dfrac{\partial c_i(\xi)}{\partial \beta_j}]\right\}, \quad j = 1, \ldots, p, \\[2ex] \dfrac{\partial \ell(\xi, y)}{\partial \lambda} = \displaystyle\sum_{i=1}^{n}\left\{ \dfrac{h_{y\lambda}(y_i, \lambda)}{h_y(y_i, \lambda)} - \dfrac{1}{\sigma}\varepsilon_i(y_i, \xi)h_\lambda(y_i, \lambda) - \dfrac{1}{c_i(\xi)}\dfrac{\partial c_i(\xi)}{\partial \lambda}\right\}, \\[2ex] \dfrac{\partial \ell(\xi, y)}{\partial \sigma} = \displaystyle\sum_{i=1}^{n}\left\{ \dfrac{1}{\sigma}\varepsilon_i^2(y_i, \xi) - \dfrac{1}{c_i(\xi)}\dfrac{\partial c_i(\xi)}{\partial \sigma}\right\}. \end{cases} \qquad \text{(A.1)}$$

Now

$$\begin{aligned} \frac{1}{c_i(\xi)}\frac{\partial c_i(\xi)}{\partial \beta_j} &= \frac{1}{c_i(\xi)}\int_{I_1}\frac{\partial}{\partial \beta_j}\{\exp[-\tfrac{1}{2}\varepsilon_i^2(y_i, \xi)]h_y(y_i, \lambda)\}\,dy_i \\ &= \frac{1}{c_i(\xi)}\int_{I_1}\frac{1}{\sigma}x_{ij}\,\varepsilon_i(y_i, \xi)\{\exp[-\tfrac{1}{2}\varepsilon_i^2(y_i, \xi)]h_y(y_i, \lambda)\}\,dy_i \\ &= \frac{1}{\sigma}x_{ij}\,\mathrm{E}_\xi[\varepsilon_i(y_i, \xi)]\ , j = 1, \ldots, p, \end{aligned} \qquad \text{(A.2)}$$

$$\begin{aligned} \frac{1}{c_i(\xi)}\frac{\partial c_i(\xi)}{\partial \lambda} &= \frac{1}{c_i(\xi)}\int_{I_1}\frac{\partial}{\partial \lambda}\{\exp[-\tfrac{1}{2}\varepsilon_i^2(y_i, \xi)]h_y(y_i, \lambda)\}\,dy_i \\ &= \frac{1}{c_i(\xi)}\int_{I_1}[\frac{h_{y\lambda}(y_i, \lambda)}{h_y(y_i, \lambda)} - \frac{1}{\sigma}\varepsilon_i(y_i, \xi)h_\lambda(y_i, \lambda)] \\ &\quad \bullet \{\exp[-\tfrac{1}{2}\varepsilon_i^2(y_i, \xi)]h_y(y_i, \lambda)\}\,dy_i \\ &= \mathrm{E}_\xi[v_i(y_i, \xi)], \end{aligned} \qquad \text{(A.3)}$$

and

$$\frac{1}{c_i(\xi)}\frac{\partial c_i(\xi)}{\partial \sigma} = \frac{1}{c_i(\xi)}\int_{I_1}\frac{\partial}{\partial \sigma}\{\exp[-\frac{1}{2}\varepsilon_i^2(y_i,\xi)]h_y(y_i,\lambda)\}dy_i$$

$$=\frac{1}{c_i(\xi)}\int_{I_1}\frac{1}{\sigma}\varepsilon_i^2(y_i,\xi)\exp[-\frac{1}{2}\varepsilon_i^2(y_i,\xi)]h_y(y_i,\lambda)dy_i$$

$$=\frac{1}{\sigma}\mathrm{E}_\xi[\varepsilon_i^2(y_i,\xi)]. \qquad \text{(A.4)}$$

Substituting (A.2)–(A.4) into (A1) yields (3.25a) and hence (3.25b).

Derivation of (3.26). By the definition of $I_n(\xi)$ we have

$$I_n(\xi) = \mathrm{E}_\xi[S(\xi,y)S^T(\xi,y)]$$

$$=E_\xi\begin{bmatrix} \left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]^T, & \left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \lambda}\right], & \left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \sigma}\right] \\ \left[\frac{\partial \ell(\xi,y)}{\partial \lambda}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]^T, & \left[\frac{\partial \ell(\xi,y)}{\partial \lambda}\right]^2, & \left[\frac{\partial \ell(\xi,y)}{\partial \lambda}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \sigma}\right] \\ \left[\frac{\partial \ell(\xi,y)}{\partial \sigma}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]^T, & \left[\frac{\partial \ell(\xi,y)}{\partial \sigma}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \lambda}\right], & \left[\frac{\partial \ell(\xi,y)}{\partial \sigma}\right]^2 \end{bmatrix}$$

The e_i's are independent, so are the ε_i's and the v_i's. Therefore the elements of $I_n(\xi)$ become

$$\mathrm{E}_\xi\left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]\left[\frac{\partial \ell(\xi,y)}{\partial \beta}\right]^T$$

$$=\frac{1}{\sigma^2}X^T\mathrm{E}_\xi\{[\varepsilon(y,\xi)-\mathrm{E}_\xi\varepsilon(y,\xi)][\varepsilon(y,\xi)-\mathrm{E}_\xi\varepsilon(y,\xi)]^T\}X$$

$$=\frac{1}{\sigma^2}X^TD(\xi)X,$$

$$\mathrm{E}_\xi\left[\frac{\partial \ell(\xi, y)}{\partial \beta}\right]\left[\frac{\partial \ell(\xi, y)}{\partial \lambda}\right]$$

$$= \frac{1}{\sigma} X^T \mathrm{E}_\xi\{[\varepsilon(y, \xi) - \mathrm{E}_\xi \varepsilon(y, \xi)] I_n^T [v(y, \xi) - \mathrm{E}_\xi v(y, \xi)]^T\}$$

$$= \frac{1}{\sigma} X^T b_1(\xi),$$

$$\mathrm{E}_\xi\left[\frac{\partial \ell(\xi, y)}{\partial \beta}\right]\left[\frac{\partial \ell(\xi, y)}{\partial \sigma}\right]$$

$$= \frac{1}{\sigma^2} X^T \mathrm{E}_\xi\left\{ [\varepsilon(y, \xi) - \mathrm{E}_\xi \varepsilon(y, \xi)][\varepsilon^T(y, \xi)\varepsilon(y, \xi) - \mathrm{E}_\xi(\varepsilon^T(y, \xi)\varepsilon(y, \xi))] \right\}$$

$$= \frac{1}{\sigma^2} X^T b_2(\xi),$$

$$\mathrm{E}_\xi\left[\frac{\partial \ell(\xi, y)}{\partial \lambda}\right]^2 = \mathrm{E}_\xi\{ I_n^T [v(y, \xi) - \mathrm{E}_\xi v(y, \xi)]\}^2$$

$$= \frac{1}{\sigma^2} k_1,$$

$$\mathrm{E}_\xi\left[\frac{\partial \ell(\xi, y)}{\partial \lambda}\right]\left[\frac{\partial \ell(\xi, y)}{\partial \sigma}\right]$$

$$= \frac{1}{\sigma} \mathrm{E}_\xi\{ I_n^T [v(y, \xi) - \mathrm{E}_\xi v(y, \xi)][\varepsilon^T(y, \xi)\varepsilon(y, \xi) - \mathrm{E}_\xi(\varepsilon^T(y, \xi)\varepsilon(y, \xi))]\}$$

$$= \frac{1}{\sigma^2} k_2,$$

and

$$\mathrm{E}_\xi\left[\frac{\partial \ell(\xi, y)}{\partial \sigma}\right] = \frac{1}{\sigma^2} \mathrm{E}_\xi[\varepsilon^T(y, \xi)\varepsilon(y, \xi) - \mathrm{E}_\xi(\varepsilon^T(y, \xi)\varepsilon(y, \xi))]^2$$

$$= \frac{1}{\sigma^2} k_3.$$

The other three elements in the lower triangle part of $I_n(\xi)$ can be obtained by symmetry.

Derivation of (3.27). From A.2 to A.4 we can easily have the following identities,

$$\frac{\partial}{\partial\beta_j}f(y_i\,|\xi) = \frac{1}{\sigma}x_{ij}\,[\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i(y_i,\,\xi)])f(y_i\,|\xi),\; j = 1,\ldots,p, \tag{A.5}$$

$$\frac{\partial}{\partial\lambda}f(y_i\,|\xi) = [v_i(y_i,\,\xi) - \mathrm{E}_\xi v_i(y_i,\,\xi)]\,f(y_i\,|\xi), \tag{A.6}$$

and

$$\frac{\partial}{\partial\sigma}\,f(y_i\,|\xi) = \frac{1}{\sigma}\,[\varepsilon_i^2(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i^2(y_i,\,\xi)]\,f(y_i\,|\xi). \tag{A.7}$$

Now from (3.25a) we have

$$\begin{aligned}
\frac{\partial^2\ell(\xi,\,y)}{\partial\beta_j\,\partial\beta_k} &= \frac{\partial}{\partial\beta_k}\left\{\frac{1}{\sigma}\sum_{i=1}^{n}x_{ij}\,[\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i(y_i,\,\xi)]\right\}\\
&= \frac{1}{\sigma}\sum_{i=1}^{n}x_{ij}\left\{\frac{\partial}{\partial\beta_k}\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi[\frac{\partial}{\partial\beta_k}\varepsilon_i(y_i,\,\xi)] - \int_{I_1}\varepsilon_i(y_i,\,\xi)\frac{\partial}{\partial\beta_k}f(y_i\,|\xi)dy_i\right\}\\
&= -\frac{1}{\sigma^2}\sum_{i=1}^{n}x_{ij}\,x_{ik}\int_{I_1}\varepsilon_i(y_i,\,\xi)[\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i(y_i,\,\xi)]\,f(y_i\,|\xi)dy_i\\
&= -\frac{1}{\sigma^2}\sum_{i=1}^{n}x_{ij}\,x_{ik}\,\mathrm{VAR}_\xi[\varepsilon_i(y_i,\,\xi)],\;\; j = 1,\ldots,p,\;\; k = 1,\ldots,p,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\ell(\xi,\,y)}{\partial\beta_j\,\partial\lambda} &= \frac{\partial}{\partial\lambda}\left\{\frac{1}{\sigma}\sum_{i=1}^{n}x_{ij}\,[\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i(y_i,\,\xi)]\right\}\\
&= \frac{1}{\sigma}\sum_{i=1}^{n}x_{ij}\left\{\frac{\partial}{\partial\lambda}\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi[\frac{\partial}{\partial\lambda}\varepsilon_i(y_i,\,\xi)] - \int_{I_1}\varepsilon_i(y_i,\,\xi)\frac{\partial}{\partial\lambda}f(y_i\,|\xi)dy_i\right\}\\
&= \frac{1}{\sigma}\sum_{i=1}^{n}x_{ij}\left\{\frac{1}{\sigma}\,[h_\lambda(y_i,\,\lambda) - \mathrm{E}_\xi\,h_\lambda(y_i,\,\lambda)]\right.\\
&\qquad \left. - \int_{I_1}\varepsilon_i(y_i,\,\xi)[v_i(y_i,\,\xi) - \mathrm{E}_\xi v_i(y_i,\,\xi)]\,f(y_i\,|\xi)dy_i\right\}
\end{aligned}$$

$$= -\frac{1}{\sigma}\sum_{i=1}^{n} x_{ij}\left\{\mathrm{COV}_{\xi}[\varepsilon_i(y_i,\,\xi)\,,\nu_i(y_i,\,\xi)] - \frac{1}{\sigma}[h_\lambda(y_i,\,\lambda) - \mathrm{E}_\xi\, h_\lambda(y_i,\,\lambda)]\right\}$$

$$= -\frac{1}{\sigma}\sum_{i=1}^{n} x_{ij}\left\{b_{1i}(\xi) - u_{1i}(\xi)\right\},$$

$$\frac{\partial^2 \ell(\xi,y)}{\partial\beta_j\,\partial\sigma} = \frac{\partial}{\partial\sigma}\left\{\frac{1}{\sigma}\sum_{i=1}^{n} x_{ij}\,[\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i(y_i,\,\xi)]\right\}$$

$$= \sum_{i=1}^{n} x_{ij}\left\{-\frac{2}{\sigma^2}[\varepsilon_i(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i(y_i,\,\xi)] - \frac{1}{\sigma}\int_{I_1}\varepsilon_i(y_i,\,\xi)\frac{\partial}{\partial\sigma}f(y_i\,|\xi)dy_i\right\}$$

$$= -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ij}\left\{\frac{u_{2i}(y_i,\,\xi)}{\sigma^2} + \frac{1}{\sigma^2}\int_{I_1}\varepsilon_i(y_i,\,\xi)[\varepsilon_i^2(y_i,\,\xi) - \mathrm{E}_\xi\varepsilon_i^2(y_i,\,\xi)]f(y_i\,|\xi)dy_i\right\}$$

$$= -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ij}\left\{\mathrm{COV}_{\xi}[\varepsilon_i(y_i,\,\xi),\,\varepsilon_i^2(y_i,\,\xi)] + u_{2i}(y_i,\xi)\right\}$$

$$= -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ij}\left\{b_{2i}\,(\xi) + u_{2i}(y_i,\,\xi)\right\},$$

$$\frac{\partial^2 \ell(\xi,y)}{\partial\lambda\,\partial\sigma} = \frac{\partial}{\partial\sigma}\left\{\sum_{i=1}^{n}[\nu_i(y_i,\,\xi) - \mathrm{E}_\xi\nu_i(y_i,\,\xi)]\right\}$$

$$= \sum_{i=1}^{n}\left\{\frac{\partial}{\partial\sigma}\nu_i(y_i,\,\xi) - \mathrm{E}_\xi[\frac{\partial}{\partial\sigma}\nu_i(y_i,\,\xi)] - \int_{I_1}\nu_i(y_i,\,\xi)\frac{\partial}{\partial\sigma}f(y_i\,|\xi)dy_i\right\}$$

$$= \sum_{i=1}^{n}\left\{\frac{2}{\sigma^2}\varepsilon_i(y_i,\,\xi)h_\lambda(y_i,\,\lambda) - \frac{2}{\sigma^2}\mathrm{E}_\xi[\varepsilon_i(y_i,\,\xi)h_\lambda(y_i,\,\lambda)]\right.$$

$$\left. - \frac{1}{\sigma}\int_{I_1}\nu_i(y_i,\,\xi)[\varepsilon_i^2(y_i,\,\xi) - \mathrm{E}_\xi\,\varepsilon_i^2(y_i,\,\xi)]\,f(y_i\,|\xi)\right\}$$

$$= -\frac{1}{\sigma}\sum_{i=1}^{n}\left\{\mathrm{COV}_{\xi}[\nu_i(y_i,\,\xi),\,\varepsilon_i^2(y_i,\,\xi)] - w_{2i}\right\}$$

$$= -\frac{1}{\sigma}(k_2 - w_2\,),$$

$$\frac{\partial^2 \ell(\xi, y)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\left\{\sum_{i=1}^{n}[\varepsilon_i^2(y_i, \xi) - \mathrm{E}_\xi \varepsilon_i^2(y_i, \xi)]\right\}$$

$$= \sum_{i=1}^{n}\left\{-\frac{3}{\sigma^2}[\varepsilon_i^2(y_i, \xi) - \mathrm{E}_\xi \varepsilon_i^2(y_i, \xi)] - \frac{1}{\sigma}\int_{I_1} \varepsilon_i^2(y_i, \xi)\frac{\partial}{\partial \sigma} f(y_i \mid \xi) dy_i\right\}$$

$$= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{\mathrm{VAR}_\xi[\varepsilon_i^2(y_i, \xi)] + w_{3i}\right\}$$

$$= -\frac{1}{\sigma^2}(k_3 + w_3),$$

and

$$\frac{\partial^2 \ell(\xi, y)}{\partial \lambda^2} = \frac{\partial}{\partial \lambda}\left\{\sum_{i=1}^{n}[\nu_i(y_i, \xi) - \mathrm{E}_\xi \nu_i(y_i, \xi)]\right\}$$

$$= \sum_{i=1}^{n}\left\{\frac{\partial}{\partial \lambda}\nu_i(y_i, \xi) - \mathrm{E}_\xi[\frac{\partial}{\partial \lambda}\nu_i(y_i, \xi)] - \int_{I_1} \nu_i(y_i, \xi)\frac{\partial}{\partial \lambda} f(y_i \mid \xi) dy_i\right\}.$$

Since

$$\frac{\partial}{\partial \lambda}\nu_i(y_i, \xi) = \sigma^2 \frac{h_{y\lambda\lambda}(y_i, \lambda)h_y(y_i, \lambda) - [h_{y\lambda}(y_i, \lambda)]^2}{[h_y(y_i, \lambda)]^2} - [h_\lambda(y_i, \lambda)]^2$$

$$- \sigma\varepsilon_i(y_i, \xi)h_{\lambda\lambda}(y_i, \lambda),$$

and

$$\int_{I_1} \nu_i(y_i, \xi)\frac{\partial}{\partial \lambda} f(y_i \mid \xi) dy_i = \mathrm{VAR}_\xi[\nu_i(y_i, \xi)],$$

we have

$$\frac{\partial^2 \ell(\xi, y)}{\partial \lambda^2} = \sum_{i=1}^{n}\left\{ w_{3i} - \mathrm{VAR}_\xi[\nu_i(y_i, \xi)]\right\}$$

$$= -\frac{1}{\sigma^2}(k_1 - w_1),$$

completing (3.27).

Derivation of (3.28).

Putting $F_{11} = \begin{bmatrix} X^TD(\xi)X, & \sigma X^Tb_1(\xi) \\ \sigma b_1^T(\xi)X, & k_1 \end{bmatrix}$, $F_{12} = \begin{bmatrix} X^Tb_2(\xi) \\ k_2 \end{bmatrix}$, $F_{21} = (F_{12})^T$, and

$F_{22} = k_3$, we have

$$\sigma^2 I_n(\xi) = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix},$$

and

$$[\sigma^2 I_n(\xi)]^{-1} = \begin{bmatrix} F_{11\cdot 2}^{-1}, & -F_{11\cdot 2}^{-1}F_{12}F_{22}^{-1} \\ -F_{22}^{-1}F_{21}F_{11\cdot 2}^{-1}, & F_{22}^{-1}F_{21}F_{11\cdot 2}^{-1}F_{12}F_{22}^{-1} + F_{22}^{-1} \end{bmatrix},$$

where $F_{11\cdot 2} = F_{11} - F_{12}F_{22}^{-1}F_{21}$.

Now

$$F_{11\cdot 2} = \begin{bmatrix} X^TD(\xi)X, & \sigma X^Tb_1(\xi) \\ \sigma b_1^T(\xi)X, & k_1 \end{bmatrix} - \frac{1}{k_3}\begin{bmatrix} X^Tb_2(\xi) \\ k_2 \end{bmatrix}\begin{bmatrix} b_2^T(\xi)X, & k_2 \end{bmatrix}$$

$$= \begin{bmatrix} X^TD_*(\xi)X, & X^Tb \\ b^TX, & k_1 - \frac{k_2^2}{k_3} \end{bmatrix}$$

$$\equiv \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

$$F_{11\cdot 2}^{-1} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} G_{11\cdot 2}^{-1}, & -G_{11\cdot 2}^{-1}G_{12}G_{22}^{-1} \\ -G_{22}^{-1}G_{21}G_{11\cdot 2}^{-1}, & G_{22}^{-1}G_{21}G_{11\cdot 2}^{-1}G_{12}G_{22}^{-1} + G_{22}^{-1} \end{bmatrix}.$$

To find $G_{11\cdot 2}^{-1}$, we have

$$G_{11\cdot 2} = G_{11} - G_{12}G_{22}^{-1}G_{21}$$
$$= X^T D_*(\xi)X - \frac{k_3}{k_1k_3 - k_2^2} X^T bb^T X.$$

Putting $A = X^T D_*(\xi)X,\ U = -\frac{k_3}{k_1k_3 - k_2^2} X^T,\ \ V = b^T X,$ and using the formula

$$(A + UV)^{-1} = A^{-1} - \frac{(A^{-1}U)(V^T A^{-1})}{1 + V^T A^{-1} U},$$

we have

$$G_{11\cdot 2}^{-1} = [X^T D_*(\xi)X]^{-1} + \frac{[X^T D_*(\xi)X]^{-1} X^T bb^T X[X^T D_*(\xi)X]^{-1}}{k_1 - \frac{k_2^2}{k_3} - b^T X[X^T D_*(\xi)X]^{-1} X^T b}$$
$$= [X^T D_*(\xi)X]^{-1} + \frac{1}{k_0} HH^T.$$

Putting $k^* = k_1 - \frac{k_2^2}{k_3}$, we have

$$k_0 = k^* - b^T XH$$
$$= k^* - H^T X^T b$$

$$-G_{11\cdot 2}^{-1} G_{12} G_{22}^{-1}$$
$$= -\frac{1}{k^*}\{([X^T D_*(\xi)X]^{-1} + \frac{1}{k_0} HH^T)X^T b\}$$
$$= -\frac{1}{k^* k_0}(k_0 H + HH^T X^T b)$$
$$= -\frac{1}{k^* k_0}\{k^* H - HH^T X^T b + HH^T X^T b\}$$
$$= -\frac{1}{k_0} H,$$

and

$$G_{22}^{-1} G_{21} G_{11\cdot 2}^{-1} G_{12} G_{22}^{-1} + G_{22}^{-1}$$
$$= \frac{1}{k^*} b^T XH \frac{1}{k_0} + \frac{1}{k^*}$$
$$= \frac{1}{k^* k_0}(b^T XH + k_0)$$

$$= \frac{1}{k^* k_0}(b^T XH + k^* - b^T XH) = \frac{1}{k_0}.$$

Hence

$$F_{11\cdot 2}^{-1} = \begin{bmatrix} [X^T D_*(\xi) X]^{-1} + \frac{1}{k_0} HH^T, & -\frac{1}{k_0} H \\ -\frac{1}{k_0} H^T, & \frac{1}{k_0} \end{bmatrix},$$

$$-F_{11\cdot 2}^{-1} F_{12} F_{22}^{-1} = -\frac{1}{k_3} \begin{bmatrix} \sigma X^T D_*(\xi) X]^{-1} X^T b_2 + \frac{\sigma}{k_0} HH^T X^T b_2 - \frac{k_2}{k_0} H \\ -\frac{\sigma}{k_0} H^T X^T b_2 + \frac{k_2}{k_0} \end{bmatrix},$$

$$\begin{aligned} F_{22}^{-1} F_{21} F_{11\cdot 2}^{-1} F_{12} F_{22}^{-1} + F_{22}^{-1} &= -\frac{1}{k_3^2} \{ \sigma^2 b_2^T X [X^T D_*(\xi) X]^{-1} X^T b_2 + \frac{\sigma^2}{k_0} b_2^T X H H^T X^T b_2 \\ &\quad - \sigma \frac{k_2}{k_0} b_2^T X H - \sigma \frac{k_2}{k_0} H^T X^T b_2 + \frac{k_2^2}{k_0} \} + \frac{1}{k_3} \\ &= k. \end{aligned}$$

The derivation of (3.28) is now complete.

Derivation of (3.10). Putting $G_{11} = X^T D(\xi) X$, $G_{12} = X^T b_1$, $G_{21} = b_1^T X$, and $G_{22} = k_3$,, we have

$$\sigma^2 I_n(\beta, \sigma) = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

and

$$[\sigma^2 I_n(\beta, \sigma)]^{-1} = \begin{bmatrix} G_{11\cdot 2}^{-1}, & -G_{11\cdot 2}^{-1} G_{12} G_{22}^{-1} \\ -G_{22}^{-1} G_{21} G_{11\cdot 2}^{-1}, & G_{22}^{-1} G_{21} G_{11\cdot 2}^{-1} G_{12} G_{22}^{-1} + G_{22}^{-1} \end{bmatrix}.$$

Now

$$
\begin{aligned}
G_{11\cdot 2} &= G_{11} - G_{12}G_{22}^{-1}G_{21} \\
&= X^T D(\xi)X - \frac{1}{k_3}X^T b_1 b_1^T X \\
&= X^T D_*(\xi)X,
\end{aligned}
$$

$$
G_{11\cdot 2}^{-1} = [X^T D_*(\xi)X],
$$

$$
\begin{aligned}
-G_{11\cdot 2}^{-1}G_{12}G_{22}^{-1} &= -\frac{1}{k_3}[X^T D_*(\xi)X]^{-1}X^T b_1 \\
&= -\frac{1}{k_3}H_2,
\end{aligned}
$$

and

$$
G_{22}^{-1}G_{21}G_{11\cdot 2}^{-1}G_{12}G_{22}^{-1} + G_{22}^{-1} = \frac{1}{k_3^2}b_2^T X H_2 + \frac{1}{k_3},
$$

this gives (3.31).

APPENDIX B
PROOFS FOR CHAPTER 4

Proof of Lemma 4.1. Since $\hat{\xi}(\lambda)$ is root-n consistent, a first-order Taylor expansion of $\psi_{1i}(y_i, \hat{\xi}(\lambda))$ around $\xi(\lambda)$ and a first-order Taylor expansion of $\psi_{3i}(y_i, \hat{\xi}(\lambda))$ around $\xi(\lambda)$ give (3.10) and (3.11).

Now by a second-order Taylor series expansion of $\psi_{1i}(y_i, \hat{\xi}(\lambda))$, we have

$$\begin{aligned}
0 \; = & \frac{1}{n}\sum_{i=1}^{n} \{ \psi_{1i} + \psi_{1i\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \psi_{1i\lambda}(\hat{\lambda} - \lambda) + \psi_{1i\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + \frac{1}{2}[I_p \otimes (\hat{\beta}(\hat{\lambda}) - \beta(\lambda))^T]\psi_{1i\beta\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \frac{1}{2}\psi_{1i\lambda\lambda}(\hat{\lambda} - \lambda)^2 \\
& + \frac{1}{2}\psi_{1i\sigma\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 + \psi_{1i\beta\lambda}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\lambda} - \lambda) \\
& + \psi_{1i\beta\sigma}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + \psi_{1i\lambda\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))\} + O_p(n^{-3/2}),
\end{aligned}$$

that is

$$\begin{aligned}
0 \; = & \; \psi_1 + \psi_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \psi_{1\lambda}(\hat{\lambda} - \lambda) + \psi_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + \frac{1}{2}[I_p \otimes (\hat{\beta}(\hat{\lambda}) - \beta(\lambda))^T]\psi_{1\beta\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \frac{1}{2}\psi_{1\lambda\lambda}(\hat{\lambda} - \lambda)^2 \\
& + \frac{1}{2}\psi_{1\sigma\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 + \psi_{1\beta\lambda}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\lambda} - \lambda) \\
& + \psi_{1\beta\sigma}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + \psi_{1\lambda\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-3/2}).
\end{aligned}$$

By assumption A2′ we have

$$\begin{aligned}
0 \; = & \; \psi_1 + A_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + A_{1\lambda}(\hat{\lambda} - \lambda) + A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + n^{-1/2}\chi_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + n^{-1/2}\chi_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1/2}\chi_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + \frac{1}{2}[I_p \otimes (\hat{\beta}(\hat{\lambda}) - \beta(\lambda))^T]\psi_{1\beta\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + \frac{1}{2}\psi_{1\lambda\lambda}(\hat{\lambda} - \lambda)^2 \\
& + \frac{1}{2}\psi_{1\sigma\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 + \psi_{1\beta\lambda}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\lambda} - \lambda) \\
& + \psi_{1\beta\sigma}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + \psi_{1\lambda\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(^{-3/2}).
\end{aligned} \tag{B1}$$

Substituting (3.10) into (B1) for the terms of order $O_p(n^{-1})$ gives

$$
\begin{aligned}
0 = {} & \psi_1 + A_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + A_{1\lambda}((\hat{\lambda} - \lambda) + A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + n^{-1/2}\chi_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1/2}\chi_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + n^{-1/2}\chi_{1\beta}[n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1})] \\
& + \frac{1}{2}\{I_p \otimes [n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1})]^T \psi_{1\beta\beta} \\
& \bullet [n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1})]\} \\
& + \frac{1}{2}\psi_{1\lambda\lambda}(\hat{\lambda} - \lambda)^2 + \frac{1}{2}\psi_{1\sigma\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 \\
& + \psi_{1\beta\lambda}[n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1})](\hat{\lambda} - \lambda) \\
& + \psi_{1\beta\sigma}[n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-1})] \\
& \bullet (\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + \psi_{1\lambda\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-3/2}) \\
= {} & \psi_1 + A_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + A_{1\lambda}(\hat{\lambda} - \lambda) + A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + n^{-1/2}\chi_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1/2}\chi_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + n^{-1}\chi_{1\beta}g_1 \\
& + n^{-1/2}\chi_{1\beta}B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1/2}\chi_{1\beta}B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + \frac{1}{2}n^{-1}(I_p \otimes g_1^T)\psi_{1\beta\beta}g_1 + \frac{1}{2}[I_p \otimes (B_{1\beta}A_{1\lambda})^T]\,\psi_{1\beta\beta}B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda)^2 \\
& + \frac{1}{2}[I_p \otimes (B_{1\beta}A_{1\sigma})^T]\psi_{1\beta\beta}B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 \\
& + n^{-1/2}(I_p \otimes g_1^T)\psi_{1\beta\beta}B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) \\
& + n^{-1/2}(I_p \otimes g_1^T)\psi_{1\beta\beta}B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + [I_p \otimes (B_{1\beta}A_{1\lambda})^T]\psi_{1\beta\beta}B_{1\beta}A_{1\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + \frac{1}{2}\psi_{1\lambda\lambda}(\hat{\lambda} - \lambda)^2 + \frac{1}{2}\psi_{1\sigma\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)))^2 + n^{-1/2}\psi_{1\beta\lambda}g_1(\hat{\lambda} - \lambda) \\
& + \psi_{1\beta\lambda}B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda)^2 + \psi_{1\beta\lambda}B_{1\beta}A_{1\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + n^{-1/2}\psi_{1\beta\sigma}g_1(\sigma(\hat{\lambda}) - \sigma(\lambda)) + \psi_{1\beta\sigma}B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda)\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
& + \psi_{1\beta\sigma}B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 + \psi_{1\lambda\sigma}(\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-3/2}).
\end{aligned}
$$

Now

$$
\begin{aligned}
& -A_{1\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) \\
& = \psi_1 + A_{1\lambda}(\hat{\lambda} - \lambda) + A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + n^{-1}[\chi_{1\beta}g_1 + \frac{1}{2}(I_p \otimes g_1^T)\psi_{1\beta\beta}g_1]
\end{aligned}
$$

$$
\begin{aligned}
&+ n^{-1/2}[\chi_{1\lambda} + \chi_{1\beta} B_{1\beta} A_{1\lambda} + (I_p \otimes g_1^T)\psi_{1\beta\beta} B_{1\beta} A_{1\lambda} + \psi_{1\beta\lambda} g_1](\hat{\lambda} - \lambda) \\
&+ n^{-1/2}[\chi_{1\sigma} + \chi_{1\beta} B_{1\beta} A_{1\sigma} + (I_p \otimes g_1^T)\psi_{1\beta\beta} B_{1\beta} A_{1\sigma} + \psi_{1\beta\sigma} g_1]((\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) \\
&+ \{\tfrac{1}{2}[I_p \otimes (B_{1\beta} A_{1\lambda})^T]\psi_{1\beta\beta} B_{1\beta} A_{1\lambda} + \tfrac{1}{2}\psi_{1\lambda\lambda} + \psi_{1\beta\lambda} B_{1\beta} A_{1\lambda}\}(\hat{\lambda} - \lambda)^2 \\
&+ \{\tfrac{1}{2}[I_p \otimes (B_{1\beta} A_{1\sigma})^T]\psi_{1\beta\beta} B_{1\beta} A_{1\sigma} + \tfrac{1}{2}\psi_{1\sigma\sigma} + \psi_{1\beta\lambda} B_{1\beta} A_{1\sigma}\}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))^2 \\
&+ \{[I_p \otimes (B_{1\beta} A_{1\lambda})^T]\psi_{1\beta\beta} B_{1\beta} A_{1\sigma} + \psi_{1\beta\lambda} B_{1\beta} A_{1\sigma} + \psi_{1\beta\sigma} B_{1\beta} A_{1\lambda} + \psi_{1\lambda\sigma}\} \\
&\bullet (\hat{\lambda} - \lambda)(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda)) + O_p(n^{-3/2}).
\end{aligned}
$$

Pre-multiplying each side by $B_{1\beta}$ gives (4.1).

The derivation of (4.2) is similar to that of (4.1), using (3.11) and the second-order Taylor expansion of $\psi_{3i}(y_i, \hat{\xi}(\hat{\lambda}))$ around $\xi(\lambda)$.

Proof of Theorem 4.1. First substituting (3.9) into (4.2) for terms of order $O_p(1)$, we have

$$
\begin{aligned}
&\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda) \\
&= n^{-1/2} g_3 + B_{3\sigma} A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + B_{3\sigma} A_{3\lambda}(\hat{\lambda} - \lambda) + n^{-1} g_3^* \\
&\quad + n^{-1/2} H_{3\beta}[n^{-1/2} B g_1 + B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda) \\
&\quad + n^{-1/2} B B_{1\beta} A_{1\sigma} g_3 + O_p(n^{-1})] + n^{-1/2} H_{3\lambda}(\hat{\lambda} - \lambda) \\
&\quad + [\, n^{-1/2} B g_1 + B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda) + n^{-1/2} B B_{1\beta} A_{1\sigma} g_3 \\
&\qquad + O_p(n^{-1})]^T H_{3\beta\beta}[n^{-1/2} B g_1 + B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda) \\
&\qquad + n^{-1/2} B B_{1\beta} A_{1\sigma} g_3 + O_p(n^{-1})] \\
&\quad + H_{3\lambda\lambda}(\hat{\lambda} - \lambda)^2 + H_{3\beta\lambda}[n^{-1/2} B g_1 + B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda) \\
&\qquad + n^{-1/2} B B_{1\beta} A_{1\sigma} g_3 + O_p(n^{-1})](\hat{\lambda} - \lambda) + O_p(n^{-3/2}) \\
&= n^{-1/2} g_3 + B_{3\sigma} A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + B_{3\sigma} A_{3\lambda}(\hat{\lambda} - \lambda) + n^{-1} g_3^* \\
&\quad + n^{-1} H_{3\beta} B(g_1 + B_{1\beta} A_{1\sigma} g_3) + n^{-1/2} H_{3\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})(\hat{\lambda} - \lambda) \\
&\quad + n^{-1/2} H_{3\lambda}(\hat{\lambda} - \lambda) + n^{-1}[B(g_1 + B_{1\beta} A_{1\sigma} g_3)]^T H_{3\beta\beta}[B(g_1 + B_{1\beta} A_{1\sigma} g_3)]
\end{aligned}
$$

$$+ [BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})]^T H_{3\beta\beta}[\cdots](\hat{\lambda} - \lambda)^2$$
$$+ 2n^{-1/2}[B(g_1 + B_{1\beta}A_{1\sigma}g_3)]^T H_{3\beta\beta}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})(\hat{\lambda} - \lambda)$$
$$+ H_{3\lambda\lambda}(\hat{\lambda} - \lambda)^2 + n^{-1/2}H_{3\beta\lambda}B(g_1 + B_{1\beta}A_{1\sigma}g_3)(\hat{\lambda} - \lambda)$$
$$+ H_{3\beta\lambda}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})(\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2})$$
$$= n^{-1/2}g_3 + B_{3\sigma}A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda)) + B_{3\sigma}A_{3\lambda}(\hat{\lambda} - \lambda)$$
$$+ n^{-1}\{g_3^* + H_{3\beta}B(g_1 + B_{1\beta}A_{1\sigma}g_3) + [B(g_1 + B_{1\beta}A_{1\sigma}g_3)]^T H_{3\beta\beta}[\cdots]\}$$
$$+ n^{-1/2}\{H_{3\beta}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda}) + H_{3\lambda} + H_{3\beta\lambda}B(g_1 + B_{1\beta}A_{1\sigma}g_3)$$
$$+ 2[B(g_1 + B_{1\beta}A_{1\sigma}g_3)]^T H_{3\beta\beta}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})\}(\hat{\lambda} - \lambda)$$
$$+ \Big\{H_{3\lambda\lambda} + [BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})]^T H_{3\beta\beta}[\cdots]$$
$$+ H_{3\beta\lambda}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})\Big\}(\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2}). \quad \text{(B2)}$$

Now substituting (B2) into (4.1) for the term $B_{1\beta}A_{1\sigma}(\hat{\sigma}(\hat{\lambda}) - \sigma(\lambda))$ and (4.4) into (4.1) for the terms of order $O_p(1)$, we obtain

$$\hat{\beta}(\hat{\lambda}) - \beta(\lambda)$$
$$= n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1}g_1^* + n^{-1/2}H_{1\lambda}(\hat{\lambda} - \lambda) + H_{1\lambda\lambda}(\hat{\lambda} - \lambda)^2$$
$$+ n^{-1/2}B_{1\beta}A_{1\sigma}g_3 + B_{1\beta}A_{1\sigma}B_{3\sigma}A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))$$
$$+ B_{1\beta}A_{1\sigma}B_{3\sigma}A_{3\lambda}(\hat{\lambda} - \lambda) + n^{-1}B_{1\beta}A_{1\sigma}\Big\{g_3^* + H_{3\beta}B(g_1 + B_{1\beta}A_{1\sigma}g_3)$$
$$+[B(g_1 + B_{1\beta}A_{1\sigma}g_3)]^T H_{3\beta\beta}[\cdots]\Big\}$$
$$+ n^{-1/2}B_{1\beta}A_{1\sigma}\Big\{H_{3\beta}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda}) + H_{3\lambda} + H_{3\beta\lambda}B(g_1 + B_{1\beta}A_{1\sigma}g_3)$$
$$+ 2[B(g_1 + B_{1\beta}A_{1\sigma}g_3)]^T H_{3\beta\beta}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})\Big\}(\hat{\lambda} - \lambda)$$
$$+ B_{1\beta}A_{1\sigma}\Big\{H_{3\lambda\lambda} + [BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})]^T H_{3\beta\beta}[\cdots]$$
$$+ H_{3\beta\lambda}BB_{1\beta}(A_{1\lambda} + A_{1\sigma}B_{3\sigma}A_{3\lambda})\Big\}(\hat{\lambda} - \lambda)^2$$
$$+ n^{-1/2}H_{1\sigma}D[n^{-1/2}(g_3 + B_{3\sigma}A_{3\beta}g_1) + B_{3\sigma}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda)]$$
$$+ H_{1\sigma\sigma}[n^{-1/2}D(g_3 + B_{3\sigma}A_{3\beta}g_1) + DB_{3\sigma}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda)]^2$$
$$+ H_{1\lambda\sigma}[n^{-1/2}D(g_3 + B_{3\sigma}A_{3\beta}g_1) + DB_{3\sigma}(A_{3\lambda} + A_{3\beta}B_{1\beta}A_{1\lambda})(\hat{\lambda} - \lambda)](\hat{\lambda} - \lambda)$$
$$= n^{-1/2}g_1 + B_{1\beta}A_{1\lambda}(\hat{\lambda} - \lambda) + n^{-1/2}B_{1\beta}A_{1\sigma}g_3 + B_{1\beta}A_{1\sigma}B_{3\sigma}A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))$$

$$+ B_{1\beta} A_{1\sigma} B_{3\sigma} A_{3\lambda}(\hat{\lambda} - \lambda)$$

$$+ n^{-1}\{ g_1^* + H_{1\sigma} D(g_3 + B_{3\sigma} A_{3\beta} g_1) + H_{1\sigma\sigma}[D(g_3 + B_{3\sigma} A_{3\beta} g_1)]^2 + B_{1\beta} A_{1\sigma} g_3^*$$

$$+ B_{1\beta} A_{1\sigma} H_{3\beta} B(g_1 + B_{1\beta} A_{1\sigma} g_3)$$

$$+ B_{1\beta} A_{1\sigma}[B(g_1 + B_{1\beta} A_{1\sigma} g_3)]^T H_{3\beta\beta} [\cdot \cdot \cdot] \}$$

$$+ n^{-1/2}\{ H_{1\lambda} + B_{1\beta} A_{1\sigma}[H_{3\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) + H_{3\lambda}]$$

$$+ B_{1\beta} A_{1\sigma} H_{3\beta\lambda} B(g_1 + B_{1\beta} A_{1\sigma} g_3) + 2B_{1\beta} A_{1\sigma}[B(g_1 + B_{1\beta} A_{1\sigma} g_3)]^T$$

$$\bullet H_{3\beta\beta} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda}) + H_{1\sigma} D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})$$

$$+ 2H_{1\sigma\sigma} D^2(g_3 + B_{3\sigma} A_{3\beta} g_1) B_{3\sigma} (A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})$$

$$+ H_{1\lambda\sigma} D(g_3 + B_{3\sigma} A_{3\beta} g_1) \} (\hat{\lambda} - \lambda)$$

$$+ \{ H_{1\lambda\lambda} + B_{1\beta} A_{1\sigma} H_{3\lambda\lambda} + B_{1\beta} A_{1\sigma}[B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})]^T H_{3\beta\beta} [\cdot \cdot \cdot]$$

$$+ B_{1\beta} A_{1\sigma} H_{3\beta\lambda} B B_{1\beta}(A_{1\lambda} + A_{1\sigma} B_{3\sigma} A_{3\lambda})$$

$$+ H_{1\sigma\sigma}[D B_{3\sigma} (A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda})]^2$$

$$+ H_{1\lambda\sigma} D B_{3\sigma}(A_{3\lambda} + A_{3\beta} B_{1\beta} A_{1\lambda}) \} (\hat{\lambda} - \lambda)^2 + O_p(n^{-3/2}).$$

Finally moving the term $B_{1\beta} A_{1\sigma} B_{3\sigma} A_{3\beta}(\hat{\beta}(\hat{\lambda}) - \beta(\lambda))$ to the left hand side and premultiplying each side by B yield ti.e result (4.3).

APPENDIX C

FORTRAN SOURCE CODE

Four FORTRAN PROGRAMS, five SUBROUTINES, and five EXTERNAL FUNCTIONS are listed in this appendix. The PROGRAM2 was not used in the thesis but is listed out for those who are interested in performing simulations by the exact method. The exact method is very expansive compared to the approximate method.

PROGRAM1 SimuApprox simulates the coverage probabilities using the approximate asymptotic expansion formula. SimuApprox calls for the subroutine **inverse** which calculate the inverse of a matrix, and subroutine **rvNormal** which generates the standard normal random numbers. The inverse subroutine calls for two other subroutines **LUDCMP** and **LUBKSB**, and rvNormal calls for **rvU(i)** function for generating the uniform random numbers.

PROGRAM2 SimuExact simulates the coverage probabilities using the exact formula. SimuExact calls for the subroutines **inverse**, **rvNormal**, and **zbrac**, and function **rtsafe**. The subroutine zbrac calls function **func,** and the function rtsafe calls for function **funcd.**

PROGRAM3 Checkea generates pairs of values of $\hat{\lambda}_B$ and $\mathcal{F}(\hat{\lambda}_B)$ using two different methods. Checkea calls for the same set subroutines and functions as Program2.

PROGRAM4 SimuFix simulates the coverage probabilities when the unknown transformation is specified a priori. SimuFix calls for the same set of subroutines and functions as Program1.

PROGRAM1: SimuApprox

```
c     Program SimuApprox is for simulating the coverage probability
c     using the asymptotic expansion method
c     ______________________________________________________________
c     Values in the parameter and data statements are to be modified
c     ______________________________________________________________
      program SimuApprox
      integer p,cm
      real lamda,la,la1
      parameter (n=27,p=4,lamda=-0.06,cm=5000)
      dimension x(n,p),z(n,p),pr(n,n),qr(n,n),a(p,p),y(p,p),b(p),
     *        par(n),indx(p),cr(5),prob(4,5),var(4),a00(n),a10(n),
     *        a20(n),a01(n),a11(n),a21(n),a111(n),a211(n),dr(n),ht(n)
      real id(n,n)
c     Input the values for model parameters, variances and critical c
      values
      data b/5.2523,0.569,-0.4312,-0.2682/
      data cr/1.45,2.21,2.80,3.41,4.26/
      data var/0.0001,0.01,1.0,10/
c     Input the design matrix
      data (x(i,1),i=1,n)/27*1.0/
      data (x(i,2),i=1,n)/9*-1.0,9*0.0,9*1.0/
      data (x(i,3),i=1,n)/3*-1.0,3*0.0,3*1.0,3*-1.0,3*0.0,
     *         3*1.0,3*-1.0,3*0.0,3*1.0/
      data (x(i,4),i=1,n-2,3)/9*-1.0/
      data (x(i,4),i=2,n-1,3)/9*0.0/
      data (x(i,4),i=3,n,3)/9*1.0/
c     compute x'x denoted as a
      do (i=1,p)
        do (j=1,p)
        w=0
          do (k=1,n)
           w=x(k,i)*x(k,j)+w
          enddo
          a(i,j)=w
        enddo
      enddo
c     calculate the inverse y of a
      call inverse(a,y,p,p,indx)
c     calculate the product z of x and y
      do (i=1,n)
        do(j=1,p)
          w=0
          do (k=1,p)
           w=x(i,k)*y(k,j)+w
          enddo
          z(i,j)=w
        enddo
      enddo
c     find the product pr of z and x'
      do (i=1,n)
        do (j=1,n)
           w=0
           do (k=1,p)
             w=z(i,k)*x(j,k)+w
```

```
            enddo
            pr(i,j)=w
         enddo
       enddo
c      set up an nxn identity matrix id
       do (i=1,n)
          do (j=1,n)
            id(i,j)=0.0
          enddo
          id(i,i)=1.0
       enddo
c      compute the matrix id-pr denoted by qr
       do (i=1,n)
         do (j=1,n)
          qr(i,j)=id(i,j)-pr(i,j)
         enddo
       enddo
c      Loop for changing variance
       do (j1=1,4)
         num1=0
         num2=0
         num3=0
         num4=0
         num5=0
c      compute parameter-related quantities
       tem1=0.0
       do (i=1,n)
         par(i)=0.0
          do (j=1,p)
           par(i)=x(i,j)*b(j)+par(i)
          enddo
            tem=1+lamda*par(i)
            a00(i)=log(tem)/lamda
            a10(i)=1/tem
            a20(i)=-lamda/tem**2
            a01(i)=(tem*log(tem)-lamda*par(i))/lamda**2
            a11(i)=log(tem)/lamda
            a111(i)=a11(i)**2
            a21(i)=1/tem
            a211(i)=2*log(tem)/(lamda*tem)
            tem1=a11(i)+tem1
        enddo
        c01=0.0
        c02=0.0
        do (i=1,n)
   .       c01=var(j1)*a21(i)**2/(2*n)+2*(a11(i)-tem1/n)**2/n+c01
          do (j=1,n)
           c02=(a01(i)+var(j1)*a21(i)/2)*qr(i,j)
     *             *(a01(j)+var(j1)*a21(j)/2)/(n*var(j1))+c02
          enddo
        enddo
        c0=c01+c02
        write(*,*) c0
        xxx=rvU(0)
c       main loop
        do (k=1,cm)
c       Compute (lamdahat-lamda) denoted by la
        call rvNormal(n,ht)
```

```
      tem2=0.0
      tem3=0.0
      tem4=0.0
      do (i=1,n)
        tem2=ht(i)**2+tem2
        do (j=1,n)
         tem3=a01(i)*qr(i,j)*ht(j)/sqrt(var(j1))+tem3
         tem4=sqrt(var(j1))*a21(i)*pr(i,j)*ht(j)/2+tem4
        enddo
      enddo
      la1=0.0
      do (i=1,n)
        la1=a11(i)*ht(i)**2+sqrt(var(j1))*a21(i)*ht(i)**3/2
     *           +a11(i)*(1-tem2/n)-a00(i)-sqrt(var(j1))*a10(i)*ht(i)
     *           -var(j1)*a20(i)*ht(i)**2/2+la1
      enddo
      la=(-la1-tem3+tem4)/(n*c0)
      do (i=1,n)
        dr(i)=la*(-2*tem1*ht(i)/n+2*a11(i)*ht(i)+sqrt(var(j1))
     *              *a21(i)*(ht(i)**2-1))+(a111(i)*ht(i)+sqrt(var(j1))
     *              *(ht(i)**2-1)/2)*la**2
      enddo
c     calculate the numerator and denominator of F
      xu=0.0
      de=0.0
      de1=0.0
        do (i=1,n)
          do (j=1,n)
            xu=ht(i)*pr(i,j)*ht(j)+xu
            de=ht(i)*qr(i,j)*ht(j)+de
            de1=(n-p)*ht(i)*pr(i,j)*dr(j)/(n*p)
          enddo
        enddo
        fr=(xu/de)*(n-p)/p+de1
        if (fr.le.cr(1)) then
          num1=num1+1
          num2=num2+1
          num3=num3+1
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(2)) then
          num2=num2+1
          num3=num3+1
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(3)) then
          num3=num3+1
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(4)) then
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(5)) then
          num5=num5+1
        end if
       write(*,*) j1,num3,k,la,de1
       enddo
       prob(j1,1)=num1*1.0/cm
```

```
      prob(j1,2)=num2*1.0/cm
      prob(j1,3)=num3*1.0/cm
      prob(j1,4)=num4*1.0/cm
      prob(j1,5)=num5*1.0/cm
      xxx=rvU(2)
      enddo
c     write the simulated coverage probabilities into file 'out prob'
      open (99, file='out prob')
      do (i=1,4)
      write (99,*) (prob(i,j),j=1,5)
      enddo
      end
```

PROGRAM2: SimuExact

```
c     Program SimuExact is used to simulate the coverage probability
c     using the numerical root-finding method to generate lamda-hat
      ___________________________________________________________________
c     Values in the parameter and data statements are to be modified
c     ___________________________________________________________________
      program SimuExact
      external func,funcd
      integer p,cm
      real lamda,la,mu
      parameter (n=27,p=4,lamda=-0.06,cm=5)
      dimension x(n,p),z(n,p),qr(n,n),pr(n,n),a(p,p),y(p,p),yd(n),
     *        par(n),e(n),ht(n),hh(n),indx(p),cr(5),prob(4,5),var(4),
     *        dr(n),ht(n),b(p)
      real id(n,n)
c     Input the values for model parameters, variances and critical
c     values
      data b/5.2523,0.569,-0.4312,-0.2682/
      data cr/1.45,2.21,2.80,3.41,4.26/
      data var/0.0001,0.01,1.0,10/
c     Input the design matrix
      data (x(i,1),i=1,n)/27*1.0/
      data (x(i,2),i=1,n)/9*-1.0,9*0.0,9*1.0/
      data (x(i,3),i=1,n)/3*-1.0,3*0.0,3*1.0,3*-1.0,3*0.0,
     *         3*1.0,3*-1.0,3*0.0,3*1.0/
      data (x(i,4),i=1,n-2,3)/9*-1.0/
      data (x(i,4),i=2,n-1,3)/9*0.0/
      data (x(i,4),i=3,n,3)/9*1.0/
c     compute x'x denoted as a
      do (i=1,p)
        do (j=1,p)
        w=0.0
          do (k=1,n)
           w=x(k,i)*x(k,j)+w
          enddo
          a(i,j)=w
        enddo
      enddo
c     calculate the inverse y of a
      call inverse(a,y,p,p,indx)
c     calculate the product z of x and y
      do (i=1,n)
        do(j=1,p)
          w=0.0
          do (k=1,p)
           w=x(i,k)*y(k,j)+w
          enddo
          z(i,j)=w
        enddo
      enddo
c     calculate the product pr of z and x'
      do (i=1,n)
        do (j=1,n)
          w=0.0
           do (k=1,p)
```

```
          w=z(i,k)*x(j,k)+w
         enddo
        pr(i,j)=w
      enddo
    enddo
c   set up an nxn identity matrix id
    do (i=1,n)
      do (j=1,n)
       id(i,j)=0.0
      enddo
     id(i,i)=1.0
    enddo
c   compute the matrix id-pr denoted by qr
    do (i=1,n)
      do (j=1,n)
       qr(i,j)=id(i,j)-pr(i,j)
      enddo
    enddo
c   compute the parameter-related quantities
    do (i=1,n)
      par(i)=0.0
      do (j=1,p)
        par(i)=x(i,j)*b(j)+par(i)
      enddo
    enddo
c   Loop for changing variance
    do (j1=1,4)
    num1=0
    num2=0
    num3=0
    num4=0
    num5=0
    xxx=rvU(0)
c   main loop
    do (k=1,cm)
     call rvNormal(n,ht)
      do (i=1,n)
        zi=1+lamda*(par(i)+sqrt(var(j1))*ht(i))
        if (zi.lt.0.0) then
        yd(i)=0.001
        else
        yd(i)=zi**(1/lamda)
        endif
      enddo
c    estimate lamda denoted by la
     call zbrac(func,-0.18,0.06,succes,r1,r2,yd,qr,n)
     la=rtsafe(funcd,r1,r2,0.001,yd,qr,n)
     write(*,*) la
     if (la.eq.0.0) then
      do (i=1,n)
       e(i) =(1/lamda)*log((1+lamda*par(i)))-lamda*(var(j1)/2)
    *           *(1+lamda*par(i))**2
       hh(i) =log(yd(i))
      enddo
     else
      do (i=1,n)
       e(i) = ((1+lamda*par(i))**(la/lamda)-1)/la+(var(j1)/2)
    *           *(la-lamda)*(1+lamda*par(i))**((la/lamda)-2)
```

```
          hh(i) = (yd(i)**la-1)/la
         enddo
        end if
c       calculate the numerator and denominator of F
        xu=0.0
        de=0.0
          do (i=1,n)
            do (j=1,n)
              xu=(hh(i)-e(i))*pr(i,j)*(hh(j)-e(j))+xu
              de=hh(i)*qr(i,j)*hh(j)+de
            enddo
          enddo
          if (de.eq.0.0) goto 22
          fr=(xu/de)*(n-p)/p
          if (fr.le.cr(1)) then
            num1=num1+1
            num2=num2+1
            num3=num3+1
            num4=num4+1
            num5=num5+1
          else if (fr.le.cr(2)) then
            num2=num2+1
            num3=num3+1
            num4=num4+1
            num5=num5+1
          else if (fr.le.cr(3)) then
            num3=num3+1
            num4=num4+1
            num5=num5+1
          else if (fr.le.cr(4)) then
            num4=num4+1
            num5=num5+1
          else if (fr.le.cr(5)) then
            num5=num5+1
          end if
 22     write(*,*) fr,num1,num2,num3,num4,num5,k,cm
        enddo
        prob(j1,1)=num1*1.0/cm
        prob(j1,2)=num2*1.0/cm
        prob(j1,3)=num3*1.0/cm
        prob(j1,4)=num4*1.0/cm
        prob(j1,5)=num5*1.0/cm
        xxx=rvU(2)
        enddo
c       write the simulated coverage probabilities into file 'out prob'
        open (99, file='out prob')
        do (i=1,4)
        write (99,*) (prob(i,j),j=1,5)
        enddo
        end
```

PROGRAM3: Checkea

```
c      Program Checkea is used for checking the accuracy of the
c      approximate method in simulating the coverage probability
c      The 3*3*3 factorial design with additive treatment effects
c      is used
c      ________________________________________________________________
c      Values in the parameter and data statements are to be modified
c      ________________________________________________________________
       program Checkea
       external func,funcd
       integer p,p1,cm
       real lamda,la,laa,la1
       parameter (n=27,p=4,p1=4,lamda=-0.06,cm=100)
       dimension x(n,p),z(n,p),qr(n,n),pr(n,n),a(p,p),y(p,p),yd(n),
     *      par(n),e(n),ht(n),hh(n),indx(p),cr(5),var(p1),out(p1*cm,4),
     *      a00(n),a10(n),a20(n),a01(n),a11(n),a21(n),a111(n),a211(n),
     *      dr(n),ht(n),b(p)
       real id(n,n)
c      Input the values for model parameters, variances and critical
c      values
       data b/5.2523,0.569,-0.4312,-0.2682/
       data cr/1.45,2.21,2.80,3.41,4.26/
       data var/0.0001,0.01,1.0,10/
c      Input the design matrix
       data (x(i,1),i=1,n)/27*1.0/
       data (x(i,2),i=1,n)/9*-1.0,9*0.0,9*1.0/
       data (x(i,3),i=1,n)/3*-1.0,3*0.0,3*1.0,3*-1.0,3*0.0,
     *          3*1.0,3*-1.0,3*0.0,3*1.0/
       data (x(i,4),i=1,n-2,3)/9*-1.0/
       data (x(i,4),i=2,n-1,3)/9*0.0/
       data (x(i,4),i=3,n,3)/9*1.0/
c      compute x'x denoted as a
       do (i=1,p)
         do (j=1,p)
         w=0
           do (k=1,n)
            w=x(k,i)*x(k,j)+w
           enddo
           a(i,j)=w
         enddo
       enddo
c      calculate the inverse y of a
       call inverse(a,y,p,p,indx)
c      calculate the product z of x and y
       do (i=1,n)
         do(j=1,p)
           w=0
           do (k=1,p)
            w=x(i,k)*y(k,j)+w
           enddo
           z(i,j)=w
         enddo
       enddo
c      find the product pr of z and x'
       do (i=1,n)
```

```
          do (j=1,n)
            w=0
            do (k=1,p)
             w=z(i,k)*x(j,k)+w
            enddo
            pr(i,j)=w
          enddo
        enddo
c       set up an nxn identity matrix id
        do (i=1,n)
          do (j=1,n)
           id(i,j)=0.0
          enddo
         id(i,i)=1.0
        enddo
c       compute the matrix id-pr denoted by qr
        do (i=1,n)
          do (j=1,n)
           qr(i,j)=id(i,j)-pr(i,j)
          enddo
        enddo
c       Loop for changing variance
        do (j1=1,p1)
c       compute parameter-related quantities
        tem1=0.0
        do (i=1,n)
         par(i)=0.0
         do (j=1,p)
          par(i)=x(i,j)*b(j)+par(i)
         enddo
          tem=1+lamda*par(i)
          a00(i)=log(tem)/lamda
          a10(i)=1/tem
          a20(i)=-lamda/tem**2
          a01(i)=(tem*log(tem)-lamda*par(i))/lamda**2
          a11(i)=log(tem)/lamda
          a111(i)=a11(i)**2
          a21(i)=1/tem
          a211(i)=2*log(tem)/(lamda*tem)
          tem1=a11(i)+tem1
         enddo
         c01=0.0
         c02=0.0
         do (i=1,n)
            c01=var(j1)*a21(i)**2/(2*n)+2*(a11(i)-tem1/n)**2/n+c01
           do (j=1,n)
            c02=(a01(i)+var(j1)*a21(i)/2)*qr(i,j)
     *              *(a01(j)+var(j1)*a21(j)/2)/(n*var(j1))+c02
           enddo
         enddo
         c0=c01+c02
         write(*,*) c0
         xxx=rvU(0)
c        main loop for simulating the F-ratio
         do (k=1,cm)
          call rvNormal(n,ht)
c        generate lamda-hat(la) by root-finding method, then calculate
c        F-ratio(fr)
```

```
      do (i=1,n)
        zi=1+lamda*(par(i)+sqrt(var(j1))*ht(i))
        if (zi.lt.0.0) then
        yd(i)=10000
        else
        yd(i)=zi**(1/lamda)
        endif
      enddo
c     estimate lamda denoted by la
      call zbrac(func,-0.18,0.06,succes,r1,r2,yd,qr,n)
      la=rtsafe(funcd,r1,r2,0.00001,yd,qr,n)
      if (la.eq.0.0) then
       do (i=1,n)
        e(i) =(1/lamda)*log((1+lamda*par(i)))-lamda*(var(j1)/2)
     *          *(1+lamda*par(i))**2
        hh(i) =log(yd(i))
       enddo
      else
       do (i=1,n)
        e(i) = ((1+lamda*par(i))**(la/lamda)-1)/la+(var(j1)/2)
     *          *(la-lamda)*(1+lamda*par(i))**((la/lamda)-2)
        hh(i) = (yd(i)**la-1)/la
       enddo
      end if
c     calculate F-ratio
      xu=0.0
      de=0.0
        do (i=1,n)
          do (j=1,n)
            xu=(hh(i)-e(i))*pr(i,j)*(hh(j)-e(j))+xu
            de=hh(i)*qr(i,j)*hh(j)+de
          enddo
        enddo
        if (de.eq.0.0) goto 22
        fr=(xu/de)*(n-p)/p
c     generate lamda-hat(hat) and F-ratio(fra) by asymptotic
c     expansion method
      tem2=0.0
      tem3=0.0
      tem4=0.0
      do (i=1,n)
        tem2=ht(i)**2+tem2
        do (j=1,n)
         tem3=a01(i)*qr(i,j)*ht(j)/sqrt(var(j1))+tem3
         tem4=sqrt(var(j1))*a21(i)*pr(i,j)*ht(j)/2+tem4
        enddo
      enddo
      la1=0.0
      do (i=1,n)
        la1=a11(i)*ht(i)**2+sqrt(var(j1))*a21(i)*ht(i)**3/2
     *          +a11(i)*(1-tem2/n)-a00(i)-sqrt(var(j1))*a10(i)*ht(i)
     *          -var(j1)*a20(i)*ht(i)**2/2+la1
      enddo
      laa=(-la1-tem3+tem4)/(n*c0)
      hat=lamda+laa
      do (i=1,n)
        dr(i)=laa*(-2*tem1*ht(i)/n+2*a11(i)*ht(i)+sqrt(var(j1))
     *          *a21(i)*(ht(i)**2-1))+(a111(i)*ht(i)+sqrt(var(j1))
```

```
     *          *(ht(i)**2-1)/2)*laa**2
      enddo
c     calculate F-ratio
      xu=0.0
      de=0.0
      de1=0.0
        do (i=1,n)
          do (j=1,n)
            xu=ht(i)*pr(i,j)*ht(j)+xu
            de=ht(i)*qr(i,j)*ht(j)+de
            de1=(n-p)*ht(i)*pr(i,j)*dr(j)/(n*p)
          enddo
        enddo
      fra=(xu/de)*(n-p)/p+de1
22    write(*,*) la,hat,fr,fra,k,cm
      j2=(j1-1)*cm+k
      out(j2,1)=la
      out(j2,2)=hat
      out(j2,3)=fr
      out(j2,4)=fra
      enddo
c     the end of the loop for cm
      xxx=rvU(2)
      enddo
c     the end of the loop for variance
c     write the results into file 'checkout'
      open (99, file='checkout')
      do (i=1,p1*cm)
      write(99,2) (out(i,j),j=1,4)
2     format(4f8.4)
      enddo
      end
```

PROGRAM4: SimuFix

```
c      SimuFix is for simulating the coverage probabilities when
c      the selected transformation is fixed
c      ____________________________________________________________
c      Values in the parameter and data statements are to be modified
c      ____________________________________________________________
       program SimuFix
       integer p,cm
       real lamda,lamda1
       parameter (n=18,p=3,lamda=0.10,cm=5000)
       dimension x(n,p),z(n,p),pr(n,n),qr(n,n),a(p,p),y(p,p),b(p),
      *        par(n),indx(p),cr(5),prob(36,5),var(4),e(n),ht(n),hh(n),
      *        trans(9)
       real id(n,n)
c      input the values for model parameters, variances, and
c      critical values
       data b/8.5,8.0,7.5/
       data cr/1.52,2.49,3.29,4.15,5.42/
       data var/0.0001,0.01,1.0,4.0/
       data trans/.30,.20,.15,.11,.10,.09,.05,.001,-.10/
c      input the design matrix
       data (x(i,1),i=1,n)/6*1.0,12*0.0/
       data (x(i,2),i=1,n)/6*0.0,6*1.0,6*0.0/
       data (x(i,3),i=1,n)/12*0.0,6*1.0/
c      compute x'x denoted as a
       do (i=1,p)
         do (j=1,p)
         w=0
           do (k=1,n)
            w=x(k,i)*x(k,j)+w
           enddo
           a(i,j)=w
         enddo
       enddo
c      calculate the inverse y of a
       call inverse(a,y,p,p,indx)
c      calculate the product z of x and y
       do (i=1,n)
         do(j=1,p)
           w=0
           do (k=1,p)
            w=x(i,k)*y(k,j)+w
           enddo
           z(i,j)=w
         enddo
       enddo
c      find the product pr of z and x'
       do (i=1,n)
         do (j=1,n)
           w=0
           do (k=1,p)
            w=z(i,k)*x(j,k)+w
           enddo
           pr(i,j)=w
         enddo
```

```
      enddo
c     set up an nxn identity matrix id
      do (i=1,n)
        do (j=1,n)
         id(i,j)=0.0
        enddo
       id(i,i)=1.0
      enddo
c     compute the matrix id-pr denoted by qr
      do (i=1,n)
        do (j=1,n)
         qr(i,j)=id(i,j)-pr(i,j)
        enddo
      enddo
c     Loop for Lamda1
      do (j2=1,9)
      lamda1=trans(j2)
c     Loop for changing variance
      do (j1=1,4)
      num1=0
      num2=0
      num3=0
      num4=0
      num5=0
c     compute parameter-related quantities
      do (i=1,n)
        par(i)=0.0
        do (j=1,p)
          par(i)=x(i,j)*b(j)+par(i)
        enddo
        e(i)=((1+lamda*par(i))**(lamda1/lamda)-1)/lamda1
     *    +(var(j1)/2)*(lamda1-lamda)*(1+lamda*par(i))**(lamda1/lamda-2)
      enddo
      xxx=rvU(0)
c     main loop
      do (k=1,cm)
      call rvNormal(n,ht)
c     calculate the numerator and denominator of F
      xu=0.0
      de=0.0
        do (i=1,n)
          ht(i)=par(i)+sqrt(var(j1))*ht(i)
          hh(i)=((1+lamda*ht(i))**(lamda1/lamda)-1)/lamda1
        enddo
c       calculate the numerator and denominator of F
        xu=0.0
        de=0.0
        do (i=1,n)
          do (j=1,n)
            xu=(hh(i)-e(i))*pr(i,j)*(hh(j)-e(j))+xu
            de=hh(i)*qr(i,j)*hh(j)+de
          enddo
        enddo
        fr=(xu/de)*(n-p)/p
        if (fr.le.cr(1)) then
          num1=num1+1
          num2=num2+1
          num3=num3+1
```

```
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(2)) then
          num2=num2+1
          num3=num3+1
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(3)) then
          num3=num3+1
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(4)) then
          num4=num4+1
          num5=num5+1
        else if (fr.le.cr(5)) then
          num5=num5+1
        end if
      enddo
c     The end of the loop cm
      j3=(j2-1)*4+j1
        prob(j3,1)=num1*1.0/cm
        prob(j3,2)=num2*1.0/cm
        prob(j3,3)=num3*1.0/cm
        prob(j3,4)=num4*1.0/cm
        prob(j3,5)=num5*1.0/cm
      xxx=rvU(2)
      enddo
c     The end of the loop for variance
      enddo
c     The end of the loop lamda1
c     write the results into file 'out prob'
      open (99, file='out prob')
      do (i=1,36)
        write (99,*) (prob(i,j),j=1,5)
      enddo
      end
```

SUBROUTINES and EXTERNAL FUNCTIONS

```
      subroutine inverse(a,y,n,np,indx)
      dimension a(np,np),y(np,np),indx(np)
      do (i=1,n)
        do (j=1,n)
         y(i,j)=0
        enddo
       y(i,i)=1
      enddo
      call LUDCMP(a,n,np,indx,d)
      do (j=1,n)
        call LUBKSB(a,n,np,indx,y(1,j))
      enddo
      return
      end

      SUBROUTINE LUDCMP(A,N,NP,INDX,D)
      PARAMETER (NMAX=100,TINY=1.0E-20)
      DIMENSION A(NP,NP),INDX(N),VV(NMAX)
      D=1.
      DO 12 I=1,N
        AAMAX=0.
        DO 11 J=1,N
          IF (ABS(A(I,J)).GT.AAMAX) AAMAX=ABS(A(I,J))
11      CONTINUE
        IF (AAMAX.EQ.0.) PAUSE 'Singular matrix.'
        VV(I)=1./AAMAX
12    CONTINUE
      DO 19 J=1,N
        IF (J.GT.1) THEN
          DO 14 I=1,J-1
            SUM=A(I,J)
            IF (I.GT.1)THEN
              DO 13 K=1,I-1
                SUM=SUM-A(I,K)*A(K,J)
13            CONTINUE
              A(I,J)=SUM
            ENDIF
14        CONTINUE
        ENDIF
        AAMAX=0.
        DO 16 I=J,N
          SUM=A(I,J)
          IF (J.GT.1)THEN
            DO 15 K=1,J-1
              SUM=SUM-A(I,K)*A(K,J)
15          CONTINUE
            A(I,J)=SUM
          ENDIF
          DUM=VV(I)*ABS(SUM)
          IF (DUM.GE.AAMAX) THEN
            IMAX=I
            AAMAX=DUM
          ENDIF
16      CONTINUE
```

```
          IF (J.NE.IMAX)THEN
            DO 17 K=1,N
              DUM=A(IMAX,K)
              A(IMAX,K)=A(J,K)
              A(J,K)=DUM
17          CONTINUE
            D=-D
            VV(IMAX)=VV(J)
          ENDIF
          INDX(J)=IMAX
          IF(J.NE.N)THEN
            IF(A(J,J).EQ.0.)A(J,J)=TINY
            DUM=1./A(J,J)
            DO 18 I=J+1,N
              A(I,J)=A(I,J)*DUM
18          CONTINUE
          ENDIF
19      CONTINUE
        IF(A(N,N).EQ.0.)A(N,N)=TINY
        RETURN
        END


        SUBROUTINE LUBKSB(A,N,NP,INDX,B)
        DIMENSION A(NP,NP),INDX(N),B(N)
        II=0
        DO 12 I=1,N
          LL=INDX(I)
          SUM=B(LL)
          B(LL)=B(I)
          IF (II.NE.0)THEN
            DO 11 J=II,I-1
              SUM=SUM-A(I,J)*B(J)
11          CONTINUE
          ELSE IF (SUM.NE.0.) THEN
            II=I
          ENDIF
          B(I)=SUM
12      CONTINUE
        DO 14 I=N,1,-1
          SUM=B(I)
          IF(I.LT.N)THEN
            DO 13 J=I+1,N
              SUM=SUM-A(I,J)*B(J)
13          CONTINUE
          ENDIF
          B(I)=SUM/A(I,I)
14      CONTINUE
        RETURN
        END

c.......rvNormal subroutine using Box-Muller algorithm...............
c
c     Usage:  call rvNormal(n,ht)
c
c     Computes n independent normal(0,1) random variables
C     using the Box-Muller method
```

```
c     (the second polar method on page 235 of Devroye)
c     and puts these in array ht.
c
c     Generates 100,000 x 1 numbers in 30 seconds
c                50,000 x 2         in 16
c               1,000 x 100       in 14
c               2,000 x 500       in 144
c
      subroutine rvNormal(n,ht)
      real *4 ht(500)
      do 10 i=1,n,2
        v=sqrt(-2.0*alog(rvU(1)))          !v= sqrt(2*exponential rv)
        w=6.283185307*rvU(1)          !w=2*pi*uniform
        ht(i)=v*cos(w)
        if (i.eq.n) go to 10
        ht(i+1)=v*sin(w)
10    continue
      return
      end


c................rvU using GFSR algorithm..............................
c
c     Usage:  x=rvU(i)
c
c           argument  i  = 0 to initialize
c                        = 1 to generate a random number
c                        = 2 to update the file 'rvU table'
c           table(p) = array of p previous random integers
c           k = place marker in table, initially 0
c           p,q = polynomial parameters
c           x**p + x**q + 1
c           p=532 q=37
c           32 = number of bits in Macintosh integer
c           2**31 -1 = 2 147 483 647 = maximum intege[illegible]
c
c     Result: the value of x is a pseudo random Uniform(0,1) variate
c
c     Implements the Generalized Feedback Shift Register generator of
c     Lewis and Payne (1973) described by Kennedy and Gentle (1980),
c     page 159.
c
c     Generates 1,000,000 rv's in 81 seconds on Mac II.
c
      function rvU(i)
      integer table(532),k,p,q,i,j
      save table,k,p,q
      if (i.eq.1) then
        k=k+1
        if (k.gt.p) k=1
        j=k+q
        if (j.gt.p) j=j-p
        table(k)=table(k).xor.table(j)
        rvU=float(table(k))/float(2 147 483 647)
        return
      else if (i.eq.0) then
        k=0
```

```
        p=532
        q=37
        open (99, file='rvU table')
        read (99,*) table
        close (99)
        rvU=float(table(1))/float(2 147 483 647)
        return
      else
        open (99, file='rvU table')
        write (99,1) (table(j),j=1,p)
1       format(i15)
        return
        rvU=float(table(k))/float(2 147 483 647)
      end if
      end


      FUNCTION RTSAFE(FUNCD,X1,X2,XACC,yd,qr,n)
      dimension yd(n),qr(n,n)
      PARAMETER (MAXIT=100)
      CALL FUNCD(X1,FL,DF,yd,qr,n)
      CALL FUNCD(X2,FH,DF,yd,qr,n)
      IF(FL*FH.GE.0.) PAUSE 'root must be bracketed'
      IF(FL.LT.0.)THEN
        XL=X1
        XH=X2
      ELSE
        XH=X1
        XL=X2
        SWAP=FL
        FL=FH
        FH=SWAP
      ENDIF
      RTSAFE=.5*(X1+X2)
      DXOLD=ABS(X2-X1)
      DX=DXOLD
      CALL FUNCD(RTSAFE,F,DF,yd,qr,n)
      DO 11 J=1,MAXIT
        IF(((RTSAFE-XH)*DF-F)*((RTSAFE-XL)*DF-F).GE.0.
     *      .OR. ABS(2.*F).GT.ABS(DXOLD*DF) ) THEN
        DXOLD=DX
        DX=0.5*(XH-XL)
        RTSAFE=XL+DX
        IF(XL.EQ.RTSAFE)RETURN
      ELSE
        DXOLD=DX
        DX=F/DF
        TEMP=RTSAFE
        RTSAFE=RTSAFE-DX
        IF(TEMP.EQ.RTSAFE)RETURN
      ENDIF
        IF(ABS(DX).LT.XACC) RETURN
        CALL FUNCD(RTSAFE,F,DF,yd,qr,n)
        IF(F.LT.0.) THEN
          XL=RTSAFE
          FL=F
        ELSE
          XH=RTSAFE
```

```
          FH=F
        ENDIF
11    CONTINUE
      PAUSE 'RTSAFE exceeding maximum iterations'
      RETURN
      END


      function funcd(x,f,df,yd,qr,n)
      dimension yd(n),qr(n,n)
      te=0.0
      te1=0.0
      te2=0.0
      te3=0.0
      te4=0.0
      do (i=1,n)
        te=log(yd(i))+te
        hi=(yd(i)**x-1)/x
        ui=yd(i)**x*log(yd(i))/x
        do (j=1,n)
          hj=(yd(j)**x-1)/x
          uj=yd(j)**x*log(yd(j))/x
          te1=hi*qr(i,j)*hj+te1
          te2=hi*qr(i,j)*uj+te2
          te3=ui*qr(i,j)*uj+te3
          te4=hi*qr(i,j)*uj*log(yd(j))+te4
        enddo
      enddo
      if (te1.eq.0.0) return
      f=-n*te2/te1+n/x+te
      df=n*(2*(te2/te1)**2-te3/te1-te4/te1-1/x**2)
      return
      end

      SUBROUTINE ZBRAC(FUNC,X1,X2,SUCCES,r1,r2,yd,qr,n)
      dimension yd(n),qr(n,n)
      PARAMETER (FACTOR=1.6,NTRY=50)
      LOGICAL SUCCES
      IF(X1.EQ.X2)PAUSE 'You have to guess an initial range'
      F1=FUNC(X1,yd,qr,n)
      F2=FUNC(X2,yd,qr,n)
      SUCCES=.TRUE.
      DO 11 J=1,NTRY
        r1=X1
        r2=X2
       IF(F1*F2.LT.0.)RETURN
        IF(ABS(F1).LT.ABS(F2))THEN
          X1=X1+FACTOR*(X1-X2)
          F1=FUNC(X1,yd,qr,n)
        ELSE
          X2=X2+FACTOR*(X2-X1)
          F2=FUNC(X2,yd,qr,n)
        ENDIF
11    CONTINUE
      SUCCES=.FALSE.
      RETURN
      END
```

```
function func(x,yd,qr,n)
dimension yd(n),qr(n,n)
te=0.0
te1=0.0
te2=0.0
do (i=1,n)
  te=log(yd(i))+te
  hi=(yd(i)**x-1)/x
  ui=yd(i)**x*log(yd(i))/x
  do (j=1,n)
    hj=(yd(j)**x-1)/x
    uj=yd(j)**x*log(yd(j))/x
    te1=hi*qr(i,j)*hj+te1
    te2=hi*qr(i,j)*uj+te2
  enddo
enddo
func=-n*te2/te1+n/x+te
return
end
```