STATISTICAL THEORY AND ROBUST METHODOLOGY FOR NONLINEAR MODELS WITH APPLICATION TO TOXICOLOGY

by
Changwon Lim

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2009

Approved by:

Pranab K. Sen, Advisor

Shyamal D. Peddada, Advisor

Douglas G. Kelly, Committee Member

Chuanshu Ji, Committee Member

J. S. Marron, Committee Member

ABSTRACT

CHANGWON LIM. Statistical Theory and Robust Methodology for Nonlinear Models with Application to Toxicology.
(Under the direction of Pranab K. Sen and Shyamal D. Peddada).

Nonlinear regression models are commonly used in dose-response studies, especially when researchers are interested in determining various toxicity characteristics of a chemical or a drug. There are several issues one needs to pay attention to when fitting nonlinear models for toxicology data, such as structure for the error variance in the model and the presence of potential influential and outlying observations. In this dissertation I developed robust statistical methods for analyzing nonlinear regression models, which are based on robust M-estimation and preliminary test estimation (PTE) procedures.

In the first part of this research the M-estimation methods in heteroscedastic nonlinear models are considered for two cases. In one case, the error variance is proportional to some known function of mean response, while in the other case the error variance is modeled as a polynomial function of dose. The asymptotic properties of the proposed M-procedures and the asymptotic efficiency of the proposed M-estimators are provided. In the second part I consider PTE-based methodology using M-methods for estimating the regression parameters. Based on the outcome of the preliminary test, the proposed methodology determines the appropriate error variance structure for the data and accordingly chooses the suitable estimation procedure. Since the resulting methodology uses M-estimators, it is expected to be robust to outliers and influential observations, although such issues have not been explored in this dissertation. Consequently, one does not have to pre-specify the error structure for the variances not does the user have to perform model diagnostics to choose a method of estimation. Some asymptotic results will be given to obtain the asymptotic covariance matrix of the PTE. Finally numerical studies are presented to illustrate the methodology. The results of the numerical studies suggest that the PTE using M-methods performs well and is robust to the error variance structure.

ACKNOWLEDGEMENTS

First of all, I would like to thank God for His constant help and guidance during my years as a student at the University of North Carolina at Chapel Hill. To Him be all the glory.

I would like to express my sincere gratitude to my advisors, Drs. Pranab K. Sen and Shyamal D. Peddada, for their invaluable advice and guidance during this important time of my career. Thanks also to the rest of my committee members, Drs. Douglas G. Kelly, Chuanshu Ji, and J. S. Marron for their valuable comments.

I am so grateful for the predoctoral visiting fellowship under Dr. Shyamal D. Peddada, in the Biostatistics Branch of the National Institute of Environmental Health Sciences, which provided a great opportunity for me to work in biostatistics.

I would also like to thank my past and present fellow graduate students in Chapel Hill, especially Dr. Myung Hee Lee, Dr. Chihoon Lee, Daniel Samarov, Xuanyao He, Fangfang Wang, Seo Young Park and Sung Kyu Jung for their help and friendship. Also, thanks to all the people at Emmanuel Antioch Presbyterian Church in Cary for their constant prayers.

I would like to thank to my parents, parents-in-law and family members for their prayers, encouragement and support over the years. Finally, I would like to express my deepest gratefulness for my wife, Jin Kyung Lee for her love, understanding, friendship and support during the years of my doctoral program.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

In 1978, the United States Congress established the National Toxicology Program (NTP) in order to broaden and strengthen scientific knowledge about potentially hazardous chemicals or compounds. The NTP's two-year cancer bioassay consists of exposing (both sexes of) rats and mice to various doses of chemical for two years and studying the tumor incidence. During the past thirty years, the NTP evaluated about 540 different chemicals that humans are exposed to. However, it is widely acknowledged that humans are exposed to several thousands of chemicals. Hence the NTP's cancer bioassay is slow and labor intensive. Furthermore, it is not always feasible to gain mechanistic understanding of the chemicals using this bioassay. As a consequence, in its roadmap for the 21st century, the NTP outlined a strategy to conduct high throughput screening (HTS) assays using cell-lines and lower order animals such as the *C. elegans* on thousands of chemicals at a time. These assays typically consist of dose-response studies involving several dose groups (as many as 10 or more) and fitting nonlinear regression models such as the Hill model for each chemical. Using these models, the toxicologists determine various parameters regarding the toxicity of a chemical.

Nonlinear regression models are commonly used in dose-response studies, especially when researchers are interested in determining various toxicity characteristics of a chemical or a drug. For example toxicologists are often interested in estimating parameters such as the dose corresponding to 50% of maximum toxic response (known as ED_{50}), the "slope" of the dose toxicity curve, the maximum tolerated dose (MTD), etc. (Velarde et al., 1999; Avalos et al., 2001; Pounds et al., 2004). The usual strategy is to fit a nonlinear regression model such as the Hill model (Gaylor and Aylward, 2004; Sand et al., 2004; Crofton et al., 2007) and estimate the parameters of the model using standard ordinary least squares estimation (OLSE). Using these parameter estimates, other estimates of interest can be obtained such as the benchmark doses (e.g., the relative ED_{01}, the dose based on a change in the mean response estimated to equal 1% of the estimated maximum change and BMD_{01}, the dose with an estimated excess risk of 1% of the animals in a low and/or high percentile among unexposed control animals) (Gaylor and Aylward, 2004; Sand et al., 2004). Eventually, using these point estimates one may construct a suitable confidence interval (e.g., LED_{01}, a lower confidence limit of ED_{01} and $BMDL_{01}$, the lower 95% confidence limit of BMD_{01}) or perform a test of hypothesis (Nitcheva et al., 2005; Piegorsch and West, 2005; Wu at al., 2006).

There are several issues one needs to pay attention to when fitting nonlinear models for toxicology data, such as structure for the error variance in the model and the presence of potential influential and outlying observations. However, since thousands nonlinear models need to be fitted, and is practically impossible to perform model diagnostics for each nonlinear model, in this dissertation research we are concerned with developing robust statistical methods for analyzing nonlinear regression models.

In the context of nonlinear models, statistical inference on unknown parameters, such as confidence interval estimation and testing of hypothesis, requires certain assumptions regarding the nonlinear model and the distributional properties of the random error associated with the model (Seber and Wild, 1989). One important assumption is that the variability in the response variable Y is constant across dose groups (known as homoscedasticity). If this assumption is violated, i.e., variance is not constant across dose groups (heteroscedasticity), then OLSE based inferential procedures may lead to confidence intervals that are subject to severe under or over coverage and thus not attain the desired confidence levels. Similarly, tests based on such procedures may not attain the desired nominal Type I error rate, and may be either too conservative or too liberal (Carroll and Ruppert, 1988; Kutner et al, 2005). Heteroscedasticity is rather common in many applications (Morris, Symanowicz, and Sarangapani, 2002; Gaylor and Aylward, 2004; Barata et al., 2006). In the context of dose-response modeling when variances were unequal, Morris, Symanowicz, and Sarangapani (2002) logtransformed their data to stabilize the variance, while Gaylor and Aylward (2004) estimated the variance to be proportional to a power of the mean response and Barata et al. (2006) used the weighted least squares estimation (WLSE) method with the weight proportional to the reciprocal of the variance.

The WLSE method is a procedure for estimating the regression parameters by minimizing the weighted error sum of squares where each weight determines how much each observation in the data set influences the final parameter estimates. However, since the weights are usually unknown in practice, they must be estimated. Thus, a typical strategy to deal with the heteroscedastic error variances is to use the iterated weighted least squares estimation (IWLSE) for estimating the parameters of the nonlinear model and then perform standard asymptotic tests and obtain confidence intervals for the desired parameters. The IWLSE method is an iterative procedure where for each iteration the updated weights are used for the WLSE. Carroll and

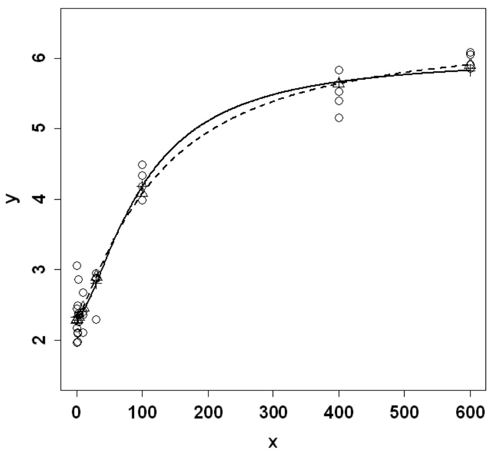
Ruppert (1982) proved that the IWLSE with unknown variances is asymptotically equivalent to the WLSE with known variances. Also, Carroll (1982) showed that even though the variances are unknown, an estimate of the regression parameter can be constructed which is asymptotically equivalent to the WLSE with known variances. Later, Davidian and Carroll (1987) developed a general theory for variance function estimation in heteroscedastic regression models. See Carroll and Ruppert (1988) for details. Carroll, Wu, and Ruppert (1988) investigated the effect of estimating weights for small-to-moderate sample sizes, and Shao (1992) developed asymptotic theory in heteroscedastic nonlinear models. Also, Hoferkamp and Peddada (2002) proposed an algorithm which iterates between the WLSE for the regression parameter and the isotonic regression estimator for the variances under order restrictions on the error variances, and Wilcox and Keselman (2004) considered some robust regression methods to achieve small standard errors when there is heteroscedasticity.

Recently, researchers have proposed tests for heteroscedasticity in regression models. Hoferkamp and Peddada (2001) studied the problem of testing for equality variances against ordered alternatives and proposed a test procedure which is unbiased under certain conditions on the design matrices. Carapeto and Holt (2003) developed a test based on the Goldfeld-Quandt methodology, which does not require further regression process and hence can be applied to all types of regression models. Lin and Wei (2003) also developed several score tests for nonlinear regression models based on either introducing a variance function for the model or randomizing regression coefficients and variance parameters. You and Chen (2005) considered partially linear regression models and developed a test procedure for heteroscedasticity based on an estimator for the best L^2-approximation of the variance function by a constant.

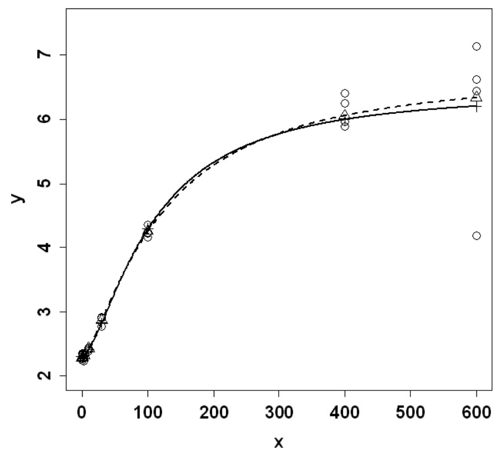
In practice one may not always know whether the data are homoscedastic or heteroscedastic. It is well-known that procedures that are suitable for heteroscedastic random errors are not likely to be robust to "nearly" homoscedastic data, where the error variances vary by less than 4-fold over the domain of the independent variable (Jacquez, Mather, and Crawford, 1968). To illustrate this point, we simulated two data sets from the Hill model with equal (Data 1) and unequal (Data 2) variances and calculated the OLSE and the WLSE from each of the two data sets. The Hill model, $f(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x^{\theta_2}/(\theta_3^{\theta_2} + x^{\theta_2})$, is a nonlinear regression model proposed by Hill (1910), usually used to study *in vivo* concentration response relationships. In the model:

- x: the dose.
- θ_0: the intercept parameter, that is, the value of y at $x = 0$.
- θ_1: the maximum effect of a drug ($\mathrm{E_{max}}$), i.e., the asymptotic value of $E(y) - \theta_0$ as x goes to ∞.
- θ_2: the slope parameter that reflects the steepness of the effect-concentration curve. Geometrically, it is the slope of the tangent line at $x = \theta_3$.
- θ_3: the sensitivity parameter, the drug concentration producing 50% of $\mathrm{E_{max}}$ ($\mathrm{ED_{50}}$).

Figure 1.1 shows the two data sets and the estimated values using the OLSE and the WLSE. In Table 1.1, when the data are generated from homoscedastic model (Data 1), the estimated standard errors of the WLSE for θ_1 ($\mathrm{E_{max}}$) and θ_3 ($\mathrm{ED_{50}}$) are much larger than the corresponding estimated standard errors of the OLSE. On the other hand, the estimated standard errors of the OLSE for θ_1 ($\mathrm{E_{max}}$) and θ_3 ($\mathrm{ED_{50}}$) are much larger than the corresponding estimated standard errors of the WLSE when the data are generated from heteroscedastic model (Data 2). Thus we see that a method suitable for homoscedastic data may not be robust for heteroscedastic data and vice versa.



(a) Data 1



(b) Data 2

Figure 1.1: Model predictions by OLSE and WLSE methods vs data: plus sign and solid line (OLSE); open triangles and dashed line (WLSE); open circles (data) for (a) Data 1 and (b) Data 2.

Table 1.1: Estimates and Standard Errors for parameters of the models for Data 1 and 2 using OLSE and WLSE methods.

	Parameter		*OLSE*		*WLSE*	
		True value	*Estimate*	*S.E.*	*Estimate*	*S.E.*
	θ_0	1	2.326	0.089	2.292	0.091
homoscedastic	θ_1	4	3.702	0.243	4.109	0.382
(Data 1)	θ_2	1.5	1.585	0.324	1.247	0.257
	θ_3	100	98.997	9.888	122.361	23.263
	θ_0	1	2.301	0.040	2.284	0.013
heteroscedastic	θ_1	4	4.175	0.864	4.496	0.240
(Data 2)	θ_2	1.5	1.541	0.442	1.377	0.062
	θ_3	100	106.152	33.676	120.161	10.154

The above observation motivates us to consider a methodology that selects between the OLSE and the WLSE for a given data set. Preliminary test estimation (PTE) procedures have been discussed extensively in the literature for such problems over the last few decades (cf. Judge and Bock, 1978; Sen, 1986). The PTE is a procedure where the result of a statistical test determines which of the two estimators, OLSE and WLSE, are used for estimating the regression parameters and the resulting estimator is called a preliminary test estimator. The PTE methodology has been used in the literature to analyze two situations, one for estimating variances and the other for estimating regression coefficients. In the first situation, it was assumed that two independent estimators s_1^2 and s_2^2 of variances σ_1^2 and σ_2^2 for a linear regression model are given together with prior information (in the form of a hypothesis) that $\sigma_1^2 \leq \sigma_2^2$. A preliminary $F-$test with s_1^2/s_2^2 as a test statistic then determines whether to use as an estimator for σ_1^2 the "pooled" estimator $\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$ or just s_1^2. The bias, variance and the Mean Square Error (MSE) of the preliminary test estimator were derived. Also, the expected value, variance and the MSE of the prelimiary test estimator under a likelihood ratio test criteria were derived. See Judge and Bock (1978) for details. Later, Sen (1986) has studied asymptotic distributional risks (i.e., the risk by reference to the asymptotic distribution of an estimator) for the pre-test versions of the maximum likelihood estimators.

Although researchers usually apply OLS and WLS methodologies to fit the data, these methodologies are not robust against outliers or influential observations. Therefore, the classical procedures are not very desirable and robust estimation procedures, such as M-procedures, are better in this respect. Sanhueza (2000) has proposed M-procedures in nonlinear regression models. The M-estimators were formulated along the lines of generalized least squares procedures and studied in univariate nonlinear models (see also Sanhueza and Sen, 2001, 2004) as well as nonlinear regression models for repeated measurements (see also Sanhueza, Sen, and Levia, 2009).

In this dissertation, we develop M-estimation and PTE-based methodologies for nonlinear regression models as the robust statistical methodologies. We propose M-estimation methods in heteroscedastic nonlinear models for two cases. In one case, the error variance is propotional to some function of mean response, while in the other case the error variance has a parametric linear form. The asymptotic properties of the proposed M-procedures and the asymptotic efficiency of the proposed M-estimators are studied. We then employ the proposed M-estimation methods together with the PTE methodology. The proposed methodology is robust to outliers and influential observations. Consequently, the user does not have to pre-specify the error structure for the variances not does the user have to perform model diagnostics to choose a method of estimation.

In the remaining sections of this chapter we present a review of the issues concerned with nonlinear models, M-procedures for nonlinear models, and PTE. The literature review in this chapter focuses mainly on three parts. In the first part, we review the nonlinear models. We present the classical methods, such as OLS, WLS for estimating parameters in these models. We consider the derivation of the covariance matrix and the consistency of the estimators. In the second part, we discuss the M-procedures for nonlinear models. We also mention the asymptotic properties of the M-procedures. In the last part, we present the PTE.

In the last section of this chapter we provide a synopsis of the new work presented in this dissertation.

1.2 Literature Review

1.2.1 Nonlinear Regression Models

Let $(\mathbf{x}_i, y_i)$, $i = 1, 2, \ldots, n$, be n observations from a fixed-effect nonlinear model with a known functional form f. Then

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \epsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where $E(\epsilon_i) = 0$, $\mathbf{x}_i$ is a $m \times 1$ vector, and $\boldsymbol{\theta}$ is known to belong to Θ, a subset of $\Re^p$. The OLSE of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, minimizes the error sum of squares

$$S_o(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\}^2 \tag{1.2}$$

over $\boldsymbol{\theta} \in \Theta$, assuming the ϵ_i to be independently and identically distributed (i.i.d.) with variance σ^2. Note that, unlike the least squares estimation for linear models, $S_o(\boldsymbol{\theta})$ may have several local minima in addition to the global minimum $\hat{\boldsymbol{\theta}}$.

For most nonlinear models the normal equations cannot be solved analytically, so that iterative methods are necessarily used. And most of the iterative methods attempt to either approximate derivatives numerically or approximate $S_o(\boldsymbol{\theta})$ in the neighborhood of $\boldsymbol{\theta}^{(k)}$, the value of $\boldsymbol{\theta}$ in the kth iteration by a smooth function (Seber and Wild, 1989). However, the normal equation may be a nonsmooth function for some nonlinear models, in which case we cannot obtain a good estimate by the iterative methods using those approximations. Instead, direct search methods do not use those approximations, rather only function evaluations. Swann (1972) has surveyed such methods well. Thus, we use the most successful of these algorithms, the "simplex" algorithm (Nelder and Mead, 1965), which is simple and easy to program.

The errors in the original model could be additive, proportional, or multiplicative.

Although it is usually assumed that the ϵ_i are i.i.d. $N(0,\, \sigma^2)$ in the model (1.1), there is often little understanding of the stochastic nature of the model, particularly with regard to the distribution and variance structure of the errors ϵ_i. Thus, the process to examine nonnormality and variance heterogeneity should be continued until a model is arrived at that appears adequate, and this final model is then used to make inferences from the data.

The method of least squares does not require the assumption of normality. However, this method may not give good estimates if there is variance heterogeneity. This heterogeneity can take two important forms: (a) $\mathrm{Var}(y)$ is proportional to some function of $E(y)$, and (b) $\mathrm{Var}(y)$ is proportional to some function of $\mathbf{x}$ (and possibly $\boldsymbol{\theta}$).

Consider the same model as in (1.1) where the ϵ_i are independently distributed as $N(0,\, \sigma^2(\mathbf{x}_i, \boldsymbol{\theta}))$ and $\sigma^2(\mathbf{x}_i, \boldsymbol{\theta}) = \sigma_0^2 w(\mathbf{x}_i, \boldsymbol{\theta})$. The weighted error sum of squares is then

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\{y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\}^2}{w(\mathbf{x}_i, \boldsymbol{\theta})}.$$

If we can determine a specific form of $w(\mathbf{x}_i, \boldsymbol{\theta})$, then we can estimate $\boldsymbol{\theta}$ by the following iterative procedure:

(i) Obtain a starting value of $\boldsymbol{\theta}$ by computing $\hat{\boldsymbol{\theta}}^{(0)}$, the OLSE of $\boldsymbol{\theta}$.

(ii) Calculate the $w\left(\mathbf{x}_i, \hat{\boldsymbol{\theta}}^{(0)}\right)$.

(iii) Compute $\hat{\boldsymbol{\theta}}$, the WLSE of $\boldsymbol{\theta}$.

(iv) Calculate a new value of w using the $\hat{\boldsymbol{\theta}}$ as a starting value.

(v) Continue the above steps until the iterations converge.

The standard errors of the parameter estimates are derived using the delta method. When each $f(\mathbf{x}_i;\, \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ is in the interior of Θ,

$\hat{\boldsymbol{\theta}}$ will satisfy

$$\left.\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = \mathbf{0}.$$

The above equation then leads to

$$\sum_i \frac{\{y_i - f(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\}}{w(\mathbf{x}_i, \hat{\boldsymbol{\theta}})} \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = \mathbf{0}.$$

Hence

$$\sum_i \frac{\{f(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) - f(\mathbf{x}_i, \boldsymbol{\theta})\}}{w(\mathbf{x}_i, \hat{\boldsymbol{\theta}})} \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = \sum_i \frac{\epsilon_i}{w(\mathbf{x}_i, \hat{\boldsymbol{\theta}})} \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}}$$

However, by the linear Taylor expansion

$$f(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) = f(\mathbf{x}_i, \boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} + o_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|),$$

and therefore,

$$\sum_i \frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t}{w(\mathbf{x}_i, \hat{\boldsymbol{\theta}})} \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} + o_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|) = \sum_i \frac{\epsilon_i}{w(\mathbf{x}_i, \hat{\boldsymbol{\theta}})} \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}},$$

where $\mathbf{a}^t$ is the transpose of a vector $\mathbf{a}$. If we use the notation $1_p = (1, \ldots, 1)^t$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^t$,

$$\widehat{\mathbf{V}} = \widehat{\mathrm{Var}}(\boldsymbol{\epsilon}) = \hat{\sigma}_0^2 \,\mathrm{Diag}\{w(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\} = \hat{\sigma}_0^2 \,\widehat{\mathbf{D}},$$

and

$$\mathbf{F} = \{F_{ij}\} = \left\{ \left.\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} \right\},$$

then,

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t \big(\mathbf{F}^t \,\widehat{\mathbf{D}}^{-1}\mathbf{F}\big) + o_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|)1_p = \boldsymbol{\epsilon}^t \widehat{\mathbf{D}}^{-1}\mathbf{F}.$$

Thus

$$\big(\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\mathbf{F}\big)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^t\big(\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\mathbf{F}\big)+o_p(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|)1_p{1_p}^t=\mathbf{F}^t\widehat{\mathbf{V}}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^t\widehat{\mathbf{V}}^{-1}\mathbf{F}$$

or

$$\big(\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\mathbf{F}\big)\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})\big(\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\mathbf{F}\big)+o(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|)1_p{1_p}^t=\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\widehat{\mathrm{Var}}(\boldsymbol{\epsilon})\widehat{\mathbf{V}}^{-1}\mathbf{F}=\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\mathbf{F},$$

and hence

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})\approx\big(\mathbf{F}^t\,\widehat{\mathbf{V}}^{-1}\mathbf{F}\big)^{-1}.$$

Now, from the model,

$$\sigma_0^2=\frac{\mathrm{Var}(\epsilon_i)}{w(\mathbf{x}_i,\boldsymbol{\theta})},\quad i=1,\ldots,n.$$

Therefore, we can obtain the estimate of σ_0^2 as follows:

$$\hat{\sigma}_0^2=\frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i-f(\mathbf{x}_i,\hat{\boldsymbol{\theta}}))^2}{w(\mathbf{x}_i,\hat{\boldsymbol{\theta}})}.$$

Suppose that we know the true value of $\boldsymbol{\theta}$, that is, $\boldsymbol{\theta}_0$ and use the notation $\mathbf{y}=(y_1,\ldots,y_n)^t$, $\mathbf{f}(\boldsymbol{\theta})=(f(\mathbf{x}_1,\boldsymbol{\theta}),\ldots,f(\mathbf{x}_n,\boldsymbol{\theta}))^t$, $\mathbf{D}_0=\mathrm{Diag}\big\{w(\mathbf{x}_i,\boldsymbol{\theta}_0)\big\}$, $s_n(\boldsymbol{\theta})=[\mathbf{y}-\mathbf{f}(\boldsymbol{\theta})]^t\mathbf{D}_0^{-1}[\mathbf{y}-\mathbf{f}(\boldsymbol{\theta})]$, and

$$\mathbf{F}(\boldsymbol{\theta})=\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{f}(\boldsymbol{\theta})=\left\{\frac{\partial f(\mathbf{x}_i,\boldsymbol{\theta})}{\partial\theta_j}\right\}.$$

Then, since $\mathbf{y}=\mathbf{f}(\boldsymbol{\theta}_0)+\boldsymbol{\epsilon}$, and hence $s_n(\boldsymbol{\theta})=[\boldsymbol{\epsilon}+\mathbf{f}(\boldsymbol{\theta}_0)-\mathbf{f}(\boldsymbol{\theta})]^t\mathbf{D}_0^{-1}[\boldsymbol{\epsilon}+\mathbf{f}(\boldsymbol{\theta}_0)-\mathbf{f}(\boldsymbol{\theta})]$,

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}}s_n(\boldsymbol{\theta})&=-2\mathbf{F}^t(\boldsymbol{\theta})\mathbf{D}_0^{-1}[\boldsymbol{\epsilon}+\mathbf{f}(\boldsymbol{\theta}_0)-\mathbf{f}(\boldsymbol{\theta})]\\\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}s_n(\boldsymbol{\theta})&=-2\left(\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{F}(\boldsymbol{\theta})\right)\mathbf{D}_0^{-1}[\boldsymbol{\epsilon}+\mathbf{f}(\boldsymbol{\theta}_0)-\mathbf{f}(\boldsymbol{\theta})]+2\mathbf{F}^t(\boldsymbol{\theta})\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}).\end{aligned}$$

Thus,

$$s_n(\hat{\boldsymbol{\theta}}) = (n-p)\hat{\sigma}_0^2; \qquad s_n(\boldsymbol{\theta}_0) = \boldsymbol{\epsilon}^t \mathbf{D}_0^{-1}\boldsymbol{\epsilon};$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} s_n(\boldsymbol{\theta}_0) = -2\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\boldsymbol{\epsilon}; \qquad \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^t} s_n(\boldsymbol{\theta}_0) = 2\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0);$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} s_n(\hat{\boldsymbol{\theta}}_n) = o_p(1); \qquad \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^t} s_n(\bar{\boldsymbol{\theta}}_n) = \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^t} s_n(\boldsymbol{\theta}_0) + o_p(1),$$

where $\bar{\boldsymbol{\theta}}_n = \alpha\boldsymbol{\theta}_0 + (1-\alpha)\hat{\boldsymbol{\theta}}_n$, $0 < \alpha < 1$. Note that $\lim_{n\to\infty} \bar{\boldsymbol{\theta}}_n = \lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$ a.s. Then,

$$\begin{aligned} s_n(\boldsymbol{\theta}_0) - s_n(\hat{\boldsymbol{\theta}}_n) &= \left(\frac{\partial}{\partial \boldsymbol{\theta}} s_n(\hat{\boldsymbol{\theta}}_n)\right)^t (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n) + \frac{1}{2}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n)^t \left(\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} s_n(\bar{\boldsymbol{\theta}}_n)\right)(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n) \\ &= o_p(1) + \frac{1}{2}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n)^t \big\{2\mathbf{F}'(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0) + o_p(1)\big\}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n) \\ &= (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \big\{\mathbf{F}'(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1), \end{aligned}$$

And

$$\frac{\partial}{\partial \boldsymbol{\theta}} s_n(\boldsymbol{\theta}_0) = \frac{\partial}{\partial \boldsymbol{\theta}} s_n(\hat{\boldsymbol{\theta}}_n) + \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^t} s_n(\bar{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n)$$

or

$$\begin{aligned} -2\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\boldsymbol{\epsilon} &= o_p(1) + \left(\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} s_n(\boldsymbol{\theta}_0) + o_p(1)\right)(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n) \\ &= 2\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n) + o_p(1), \end{aligned}$$

and hence,

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = \big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\boldsymbol{\epsilon} + o_p(1)$$

Therefore,

$$s_n(\boldsymbol{\theta}_0) - s_n(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\epsilon}^t\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\boldsymbol{\epsilon} + o_p(1)$$

or

$$s_n(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\epsilon}^t\big[\mathbf{I}_n - \mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\mathbf{F}^t(\boldsymbol{\theta}_0)\big]\mathbf{D}_0^{-1}\boldsymbol{\epsilon} + o_p(1).$$

Now,

$$
\begin{aligned}
& E\Big[\boldsymbol{\epsilon}^t\big[\mathbf{I}_n - \mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\big]\mathbf{D}_0^{-1}\boldsymbol{\epsilon}\Big] \\
&= E\Big[\mathrm{tr}\big(\boldsymbol{\epsilon}^t\big[\mathbf{I}_n - \mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\big]\mathbf{D}_0^{-1}\boldsymbol{\epsilon}\big)\Big] \\
&= E\Big[\mathrm{tr}\big(\big[\mathbf{I}_n - \mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\big]\mathbf{D}_0^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^t\big)\Big] \\
&= \mathrm{tr}\Big(\big[\mathbf{I}_n - \mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\big]\mathbf{D}_0^{-1}E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^t)\Big) \\
&= \mathrm{tr}\Big(\big[\mathbf{I}_n - \mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\big]\mathbf{D}_0^{-1}(\sigma_0^2\mathbf{D}_0)\Big) \\
&= \sigma_0^2\Big\{\mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}\big[\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\mathbf{F}^t(\boldsymbol{\theta}_0)\big]\Big\} \\
&= \sigma_0^2\Big\{n - \mathrm{tr}\big[\big\{\mathbf{F}^t(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big\}^{-1}\mathbf{F}^(\boldsymbol{\theta}_0)\mathbf{D}_0^{-1}\mathbf{F}(\boldsymbol{\theta}_0)\big]\Big\} \\
&= \sigma_0^2\big\{n - \mathrm{tr}(\mathbf{I}_p)\big\} = \sigma_0^2(n-p)
\end{aligned}
$$

And, if we let $R = o_p(1)$, $\lim_{n\to\infty} P(|R| > \epsilon) = 0$, $\forall \epsilon > 0$. Thus,

$$
\begin{aligned}
E|R| &= \int_0^\infty P(|R| > 0)dP \\
&= \int_0^\infty \big\{P(|R| > \epsilon) + P(|R| < \epsilon)\big\}dP \\
&= \int_0^\infty P(|R| > \epsilon)dP + \int_0^\infty P(|R| < \epsilon)dP \\
&\longrightarrow 0 \;\text{ as }\; \epsilon \to 0 \;\text{ and }\; n \to \infty,
\end{aligned}
$$

and hence, $0 < |E(R)| < E|R| \to 0$ or $E(R) = o(1)$. Therefore,

$$
E(\hat{\sigma}_0^2(n-p)) = E(s_n(\hat{\boldsymbol{\theta}}_n)) = \sigma_0^2(n-p) + o(1) \quad \text{(Gallant, 1987)}.
$$

1.2.2 M-estimation Procedures in Nonlinear Regression Models

Classical methods of estimation, such as Least Squares (LS) and Maximum Likelihood (ML) procedures in nonlinear regression usually give estimators which are usually nonrobust to outliers or departures from the specified distribution of the response variable. Thus, from consideration of robustness, the classical procedures are not very desirable and robust estimation procedures, such as M-procedures, are better in this context.

Sanhueza and Sen (2001) have proposed M-procedures for generalized nonlinear models. For the model in (1.1), an M-estimator of $\boldsymbol{\theta}$ are defined as the minimization:

$$\hat{\boldsymbol{\theta}}_n = Arg \cdot min \left\{ \sum_{i=1}^{n} \frac{1}{h\big[w(\mathbf{x}_i, \boldsymbol{\theta})\big]} h^2\big(y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\big) : \boldsymbol{\theta} \in \Theta \subseteq \Re^p \right\} \tag{1.3}$$

where $h(\cdot)$ is a real valued function. Under the assumption of fixed variance function in the above equation and $\psi(z) = (\partial/\partial z)h^2(z)$, the estimating equation for the minimization in the above equation is given by:

$$\sum_{i=1}^{n} \lambda(\mathbf{x}_i, y_i, \hat{\boldsymbol{\theta}}_n) = \mathbf{0}$$

where

$$\lambda(\mathbf{x}_i, y_i, \boldsymbol{\theta}) = \frac{1}{h\big[w(\mathbf{x}_i, \boldsymbol{\theta})\big]} \psi\big(y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\big) \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})$$

and $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta}) f(\mathbf{x}_i, \boldsymbol{\theta})$. It can be seen that choosing $h(z) = z$ produces the LS estimator and $h(z) = \sqrt{|z|}$ produces the least absolute deviation (LAD) estimator. In the coventional setup of robust methods (Huber, 1981; Hampel *et al.*, 1986; Jurečková and Sen, 1996), bounded and monotone functions $h(\cdot)$ are commonly used; the so called Huber-score function corresponds to

$$h(z) = \begin{cases} \frac{1}{\sqrt{2}} z, & \text{if } |z| \leq k \\ \{k(|z| - \frac{k}{2})\}^{\frac{1}{2}}, & \text{if } |z| > k \end{cases} \tag{1.4}$$

for suitable chosen $k(0 < k < \infty)$. Whenever the errors ϵ_i in (1.1) have a symmetric distribution, we may choose k as a suitable percentile point of this law, such as the 90th or 95th percentile, and let h as in (1.4).

Under appropriated regularity conditions, the M-estimator as defined in (1.3) is consistent and asymptotically normally distributed:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \longrightarrow N_p\big(\mathbf{0}, \gamma^{-2}\sigma_\psi^{-2}\boldsymbol{\Gamma}_1^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_2(\boldsymbol{\theta})\boldsymbol{\Gamma}_1^{-1}(\boldsymbol{\theta})\big),$$

where for $\epsilon = y - f(\mathbf{x}, \boldsymbol{\theta})$, $\gamma = E\psi'(\epsilon)$, $\sigma_\psi^2 = E\psi^2(\epsilon)/u(\mathbf{x})$, and

$$\boldsymbol{\Gamma}_1 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h[w(\mathbf{x}_i, \boldsymbol{\theta})]} \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta})$$

and

$$\boldsymbol{\Gamma}_2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{u(\mathbf{x}_i)}{h^2[w(\mathbf{x}_i, \boldsymbol{\theta})]} \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta})$$

are positive definite matrices.

Sanhueza and Sen (2001) have also proposed M-procedures for nonlinear regression models where no assumption on the underlying distribution for the response variable is made except the existence of the expected valude and variance of the response variable. If we consider the nonlinear model

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + a_i\epsilon_i, \quad i = 1, \ldots, n \tag{1.5}$$

where y_{ij} are the observable random variables (r.v.), $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{mi})^t$ are known regression constants, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^t$ is a vector of unknown parameters, $f(,)$ is a nonlinear function of $\boldsymbol{\theta}$ of specified form, the errors ϵ_{ij} are assumed to be i.i.d r.v.s with some distribution function which is defined on $\Re$, continuous and symmetric about 0 and the $a_i(> 0)$ are known constant, possibly dependent on the $\mathbf{x}_i$. Then, an M-estimator of $\boldsymbol{\theta}$ are defined as the minimization:

$$\hat{\boldsymbol{\theta}}_n = Arg \cdot min \left\{ \sum_{i=1}^{n} w_{ni} h^2\big(y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\big) : \boldsymbol{\theta} \in \Theta \subseteq \Re^p \right\} \tag{1.6}$$

where

$$w_{ni} = \big[E(\psi^2(y_i - f(\mathbf{x}_i, \boldsymbol{\theta})))\big]^{-1} / \sum_{j=1}^{n} \big[E(\psi^2(y_j - f(\mathbf{x}_j, \boldsymbol{\theta})))\big]^{-1}, \quad j = 1, \ldots, n.$$

Under certain regularity contions, they proved that the M-estimator is consistent and asymptotically normally distributed. The reader is referred to Sanhueza and Sen (2001, pp. 359–375).

1.2.3 Preliminary Test Estimation

In problems of statistical inference, the classical estimators of unknown parameters are based exclusively on the sample data. Such estimators ignore any other kind of non-sample prior information (NPI) in its definition. The notion of inclusion of NPI to the estimation of parameters has been introduced to 'improve' the estimation procedures. The natural expectation is that the use of additional information would result in a better estimator. In some cases, this may be true, but in many other cases the risk of worse consequences can not be ruled out.

The problem under consideration is then changed to the problems of statistical inference in the presence of uncertain NPI. It is usual in the literature to treat such uncertain NPI specified by a null hypothesis as a "nuisance parameter". Then the uncertainty in the form of the “nuisance parameter” is eliminated by a so-called preliminary test. Bancroft (1944, 1964, 1972) first addressed the problem, and proposed the well-known PTE. Later, Kitagawa (1963), Han and Bancroft (1968), Saleh and Han (1990), Ali and Saleh (1990), and Mahdi *et al.* (1998), contributed in the development of the PTE methods under the normal theory. Furthermore, Saleh and Sen (1978, 1985) published a series of articles in this area exploring the nonparametric as well as the asymptotic theory based on the least square estimators. Bhoj and Ahsanullah (1993, 1994) discussed the problem of estimation of conditional mean for simple regression model.

The PTE methodologies have studied and applied actively even recent years. Tabatabaey (1995), Kibria (1996), Khan and Saleh (1995, 1997, 2008), Khan (2000, 2005), Tabatabaey, Saleh and Kibria (2004a,b), and Arashi and Tabatabaey (2008) investigated the problem for a family of Student-t populations. Ahmed and Krzanowski (2004) proposed new PTEs for the parameter vectors in a simple multivariate normal regression model and Chaubey and Sen (2004) discussed the PTE for mean of an inverse Gaussian population. The PTE method was also employed to develop inference tools for an effect size parameter in a paired experiment by Al-Kandari, Buhamra and Ahmed (2005), and the relative performance of the PTE and shrinkage estimators of the intercept parameter of linear regression model has been investigated by Khan, Hoque and Saleh (2005). Furthermore, Kim and Saleh (2005) considered the PTE methodologies in the problem of simultaneous estimation of the regression parameters in a multiple regression model with measurement errors. Wan and Zou (2003), and Wan, Zou, Ohtani (2006) considered the choice of critial values for the PTE procedures based on the minimum risk criterion. Yunus and Khan (2007) proposed a robust test statistic based on M-statistic to formulate the asymptotic power functions for testing the intercept after pre-testing on slope and to carry out investigations on the asymptotic properties of this power function. And Baklizi (2008) considered the PTE based on the maximum likelihood estimator (MLE) of the parameter of the pareto distribution and obtained the optimal significance levels for the PT using the minmax regret criterion. Baklizi and Ahmed (2008) proposed the shrinkage PTE of the reliability function of the Weibull lifetime model to achieve improved estimation performance. Kabir and Khan (2008) proposed p-value based PTE based on the sample, NPI, p-value of an appropriate test for estimation of the slope parameter of a simple regression model. Khan (2008) considered the shrinkage PTE of the intercept parameters of two linear regression models with normal errors, when it is *a priori* suspected that the two regression lines are parallel, but in doubt. Menéndez, Pardo and Pardo (2008, 2009) considered the PTE of the parameters in the generalized linear models with binary data.

Let us consider the simple linear model with slope β, and intercept θ, given by

$$\mathbf{Y} = \theta \mathbf{1}_n + \beta \mathbf{x} + \boldsymbol{\epsilon}, \tag{1.7}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of the observable r.v.s, $\mathbf{x}$ is an $n \times 1$ vector of known constants, $\mathbf{1}_n$ is an $n \times 1$ vector of 1's, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of independent errors such that $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ with $\mathbf{I}_n$ the identity matrix of order n.

The maximum likelihood estimator (MLE) of $(\theta, \beta)^t$ in the model (1.7) is given as

$$\begin{pmatrix} \tilde{\theta}_n \\ \tilde{\beta}_n \end{pmatrix} = \begin{pmatrix} \bar{Y} - \tilde{\beta}_n \bar{x} \\ \frac{1}{Q}\left\{ \mathbf{x}^t \mathbf{Y} - \frac{1}{n}(\mathbf{1}_n^t \mathbf{x})(\mathbf{1}_n^t \mathbf{Y}) \right\} \end{pmatrix}, \tag{1.8}$$

where

$$Q = \mathbf{x}^t \mathbf{x} - \frac{1}{n}(\mathbf{1}_n^t \mathbf{x})^2. \tag{1.9}$$

When it is suspected that the slope parameter β may be β_0, since in practice the NPI is uncertain, it is used to test the hypothesis $H_0 : \beta = \beta_0$ against the alternative

$H_1 : \beta \neq \beta_0$ in order to eliminate the doubt on the prior information. Then we use the likelihood ration (LR) test statistic

$$T_n = \frac{(\tilde{\beta}_n - \beta_0)^2 Q}{s^2}, \tag{1.10}$$

which follows a noncentral F-distribution with $(1, m = n-2)$ degrees of freedom (d.f.) and noncentrality parameter $\Delta^2/2$ where

$$s^2 = \frac{1}{n-2}\left(\mathbf{Y} - \tilde{\theta}_n \mathbf{1}_n - \tilde{\beta}_n \mathbf{x}\right)^t \left(\mathbf{Y} - \tilde{\theta}_n \mathbf{1}_n - \tilde{\beta}_n \mathbf{x}\right) \tag{1.11}$$

is an unbiased estimator of the variance σ^2 and

$$\Delta^2 = \frac{(\beta - \beta_0)^2 Q}{\sigma^2}. \tag{1.12}$$

Under H_0, T_n follows a central F-distribution, so at the significance level α we obtain the critical value $F_{1,m}(\alpha)$ from this distribution and reject H_0 if $T_n \geq F_{1,m}(\alpha)$; otherwise, we accept H_0. As a result of this test, we choose the unrestricted estimator (UE) $\tilde{\theta}_n$ or the restricted estimator (RE) $\hat{\theta}_n$ based on the rejection or acceptance of H_0, where $\hat{\theta}_n$ is given by

$$\hat{\theta}_n = \bar{Y} - \beta_0 \bar{x}. \tag{1.13}$$

Accordingly, we define the PTE as

$$\hat{\theta}_n^{\mathrm{PT}} = \hat{\theta}_n I(T_n < F_{1,m}(\alpha)) + \tilde{\theta}_n I(T_n \geq F_{1,m}(\alpha)), \tag{1.14}$$

where $I(A)$ is the indicator function of the set A. For more details, see Judge and Bock (1978), Ahmed and Saleh (1988), Ahsanullah and Saleh (1972), and Saleh (2006).

Then, we compare the UE, RE, and PTE of θ with respect to the bias and the mean squared error (MSE), based on the following theorems (Saleh, 2006).

Theorem 1: The bias, $b(\cdot)$, and the quadratic bias $B(\cdot)$ expressions for $\tilde{\theta}_n$, $\hat{\theta}_n$, and $\hat{\theta}_n^{\rm PT}$ are given as follows

(i) $b(\tilde{\theta}_n) = 0, \quad B(\tilde{\theta}_n) = 0,$

(ii) $b(\hat{\theta}_n) = (\beta - \beta_0)\bar{x}, \quad B(\hat{\theta}_n) = (\bar{x}^2\Delta^2)/Q,$

(iii) $b(\hat{\theta}_n^{\rm PT}) = (\beta-\beta_0)\bar{x}G_{3,m}(\frac{1}{3}F_{1,m}(\alpha);\Delta^2),\ B(\hat{\theta}_n^{\rm PT}) = \big\{\bar{x}^2\Delta^2G_{3,m}^2(\frac{1}{3}F_{1,m}(\alpha);\Delta^2)\big\}/Q,$

where $G_{m_1,m_2}(\cdot;\Delta^2)$ is the cdf of a noncentral F-distribution with (m_1, m_2) d.f. and noncentrality parameter $\Delta^2/2$.

Theorem 2: The MSE expressions for $\tilde{\theta}_n$, $\hat{\theta}_n$, and $\hat{\theta}_n^{\rm PT}$ are given by

$$MSE(\tilde{\theta}_n) = \frac{\sigma^2}{n}\left(1 + \frac{n\bar{x}^2}{Q}\right), \quad MSE(\hat{\theta}_n) = \frac{\sigma^2}{n}\left(1 + \frac{n\bar{x}^2}{Q}\Delta^2\right)$$

and

$$\begin{aligned} MSE(\hat{\theta}_n^{\rm PT}) &= \frac{\sigma^2}{n}\left(1 + \frac{n\bar{x}^2}{Q}\right) - \sigma^2\frac{\bar{x}^2}{Q}G_{3,m}\left(\frac{1}{3}F_{1,m}(\alpha);\Delta^2\right) \\ &\quad +\sigma^2\frac{\bar{x}^2}{Q}\Delta^2\left\{2G_{3,m}\left(\frac{1}{3}F_{1,m}(\alpha);\Delta^2\right) - G_{5,m}\left(\frac{1}{5}F_{1,m}(\alpha);\Delta^2\right)\right\}, \end{aligned}$$

respectively.

1.3 Synopsis of Research

In this research, we are primarily interested in the study of M-procedures for estimating the parameters in heteroscedastic nonlinear models and the PTE-based methodology for nonlinear regression models using M-estimation methods, as the robust statistical methodologies.

In Chapter 2, we define M-estimators for the parameters in heteroscedastic nonlinear models, in which case the error variance is assumed to be proportional to the mean response. Under some regularity conditions, we derive the asymptotic distribution of the M-estimators. we also show that even though the variances are unknown, an M-estimate of the regression parameter can be constructed which is asymptotically equivalent to the M-estimate with known variances. Then, we iilustrate examples where we verify that the regularity conditions are satisfied for some situations.

In Chapter 3, we study M-procedures for estimating both the regression parameters and the parameters for variance in nonlinear models, where the log-variance of the random error is assumed to be linear in the explanatory variable. We formulate suitable M-estimators of the parameters in the model and study their asymptotic distribution, including consistency, uniform asymptotic linearity, and normality.

In Chapter 4, we consider the PTE-based methodology for nonlinear regression models using M-estimation methods. We propose the PTE procedures for estimation the regression parameter in heteroscedastic nonlinear regression models when it is suspected that the error variances are possibly homoscedastic. We derive some asymptotic results to obtain the asymptotic covariance matrix of the PTE.

Numerical results are presented in Chapter 5 and 6 to illustrate the methodology.

Finally in Chapter 7 we propose some topics for further research.

Chapter 2

M-METHODS IN HETEROSCEDASTIC NONLINEAR MODELS – I

2.1 Introduction

In this chapter, we discuss M-procedures in heteroscedastic nonlinear regression models. It may be seen as a different version of the M-procedures that Sanhueza and Sen (2001) studied in the generalized nonlinear models. Here, the methodology and asymptotic theory of the M-procedures are similar in nature to those in the generalized nonlinear model case. They used the function h, having the variance function from the generalized nonlinear models as the argument, in the denominator of the estimating equation. However, we generalize it so that instead of h we use a suitably chosen function w (assumed to be known) in the denominator which has the mean response as the argument, assuming that the variance is proportional to the mean response. And also we show that even though the variances are unknown, an M-estimate of the regression parameter can be constructed which is asymptotically equivalent to the M-estimate with known variances.

In Section 2.2 we define the M-estimator for the parameter of interest in the heteroscedastic nonlinear model. We also present the notation and regularity conditions necessary to derive the asymptotic results in this chapter. In Section 2.3 we develop the asymptotic properties of the M-estimates. In Section 2.4 we illustrate examples where we verify that the regularity conditions are satisfied for some situations.

Let

$$y_{ij} = f(\mathbf{x}_i, \boldsymbol{\theta}) + \epsilon_{ij}, \quad i = 1, \ldots, k, \; j = 1, \ldots, n_i \tag{2.1}$$

where y_{ij} are the observable r.v.s of size $n\,(= \sum n_i)$, $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{mi})^t$ are known regression constants, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^t$ is a vector of unknown parameters, $f(,)$ is a nonlinear function of $\boldsymbol{\theta}$ of specified form; and the errors ϵ_{ij} are assumed to be independent r.v.s such that $\epsilon_{ij} \sim N(0, \sigma_i^2)$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$. We assume that $\sigma_i^2 = \sigma^2(f(\mathbf{x}_i, \boldsymbol{\theta}))$ for $i = 1, \ldots, k$.

2.2 Definitions and Regularity Conditions

We define an M-estimator of $\boldsymbol{\theta}$ as the estimator that solves the following minimization problem:

$$\hat{\boldsymbol{\theta}}_n = Argmin \left\{ \sum_{i,j} \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))} h^2(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta})) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \Re^p \right\} \tag{2.2}$$

where the functional form of w is assumed to be known, $h(\cdot)$ is a real valued function, and $\boldsymbol{\Theta}$ is a compact subset of $\Re^p$. Let $\psi(z) = (\partial/\partial z)h^2(z)$, then the estimating equation for the minimization problem (2.2) is given by:

$$\sum_{i,j} \lambda(\mathbf{x}_i, y_{ij}, \hat{\boldsymbol{\theta}}_n) = \mathbf{0} \tag{2.3}$$

where

$$\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) = \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))}\psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \tag{2.4}$$

and $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta})f(\mathbf{x}_i, \boldsymbol{\theta})$.

We make the following sets of regularity assumptions regarding (A) the score function ψ, (B) the function $f(\cdot)$, and (C) the functions $k_1(\cdot)$ and $k_2(\cdot)$, where

$$k_1(\mathbf{x}_i, \boldsymbol{\theta}) = 1/w(f(\mathbf{x}_i, \boldsymbol{\theta})), \tag{2.5}$$

and

$$k_2(\mathbf{x}_i, \boldsymbol{\theta}) = \frac{w'(f(\mathbf{x}_i, \boldsymbol{\theta}))}{w^2(f(\mathbf{x}_i, \boldsymbol{\theta}))}. \tag{2.6}$$

[**A1**]:

ψ is nonconstant, absolutely continuous and differentiable with respect to $\boldsymbol{\theta}$.

[**A2**]: Let $\epsilon = Y - f(\mathbf{x}, \boldsymbol{\theta})$,

(i) $E\psi^2(\epsilon) = \sigma^2_\psi u(\mathbf{x}) < \infty$, and $E\psi(\epsilon) = 0$

(ii) $E|\psi'(\epsilon)|^{1+\delta} < \infty$ for some $0 < \delta \leq 1$, and $E\psi'(\epsilon) = \gamma\,(\neq 0)$

[**A3**]:

(i) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\|\,\leq\,\delta}\left|\psi(Y - f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})) - \psi(Y - f(\mathbf{x}, \boldsymbol{\theta}))\right|\right\} = 0$

(ii) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\|\,\leq\,\delta}\left|\psi'(Y - f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})) - \psi'(Y - f(\mathbf{x}, \boldsymbol{\theta}))\right|\right\} = 0$

[**B1**]:

$f(\mathbf{x}, \boldsymbol{\theta})$ is continuous and twice differentiable with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact subset of $\Re^p$.

[**B2**]:

(i) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_1(\boldsymbol{\theta})$, where

$$\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) = \sum_{i,j}\left\{\frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\mathbf{f}^t_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\right\},$$

and $\mathbf{\Gamma}_1(\boldsymbol{\theta})$ is a positive definite matrix.

(ii) $\lim_{n\to\infty}\frac{1}{n}\mathbf{\Gamma}_{2n}(\boldsymbol{\theta}) = \mathbf{\Gamma}_2(\boldsymbol{\theta})$, where

$$\mathbf{\Gamma}_{2n}(\boldsymbol{\theta}) = \sum_{i,j}\left\{\frac{u(\mathbf{x}_i)}{w^2(f(\mathbf{x}_i,\boldsymbol{\theta}))}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta})\mathbf{f}^t_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta})\right\},$$

and $\mathbf{\Gamma}_2(\boldsymbol{\theta})$ is a positive definite matrix.

(iii) $\max\left\{\frac{u(\mathbf{x}_i)}{w^2(f(\mathbf{x}_i,\boldsymbol{\theta}))}\mathbf{f}^t_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta})(\mathbf{\Gamma}_{2n}(\boldsymbol{\theta}))^{-1}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta})\right\} \longrightarrow 0$, as $n\to\infty$

[**B3**]:

(i) $\lim_{\delta\to 0}\sup_{\|\mathbf{\Delta}\|\le\delta}\Big|(\partial/\partial\theta_j)f(\mathbf{x},\boldsymbol{\theta}+\mathbf{\Delta})(\partial/\partial\theta_l)f(\mathbf{x},\boldsymbol{\theta}+\mathbf{\Delta})$

$\qquad - (\partial/\partial\theta_j)f(\mathbf{x},\boldsymbol{\theta})(\partial/\partial\theta_l)f(\mathbf{x},\boldsymbol{\theta})\Big| = 0$ for $j,l = 1,\ldots,p$

(ii) $\lim_{\delta\to 0}\sup_{\|\mathbf{\Delta}\|\le\delta}\Big|(\partial^2/\partial\theta_j\partial\theta_l)f(\mathbf{x},\boldsymbol{\theta}+\mathbf{\Delta}) - (\partial^2/\partial\theta_j\partial\theta_l)f(\mathbf{x},\boldsymbol{\theta})\Big| = 0$ for $j,l = 1,\ldots,p$

[**C**]:

(i) $\lim_{\delta\to 0}\sup_{\|\mathbf{\Delta}\|\le\delta}\big|k_1(\mathbf{x},\boldsymbol{\theta}+\mathbf{\Delta}) - k_1(\mathbf{x},\boldsymbol{\theta})\big| = 0$, uniformly in $\mathbf{x}$.

(ii) $\lim_{\delta\to 0}\sup_{\|\mathbf{\Delta}\|\le\delta}\big|k_2(\mathbf{x},\boldsymbol{\theta}+\mathbf{\Delta}) - k_2(\mathbf{x},\boldsymbol{\theta})\big| = 0$, uniformly in $\mathbf{x}$.

2.3 Asymptotic Results

First we present two lemmas.

Lemma 2.1. *Let the conditions* [**A1**]-[**A3**], [**B1**]-[**B3**], *and* [**C**] *hold and let* $\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ *be the rth element of the vector* $\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ *for* $r = 1,\ldots,p$. *Then for* $r = 1,\ldots,p$

$$\sup_{\|\mathbf{t}\|\le C}\left|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l\Big\{(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}+\frac{s\mathbf{t}}{\sqrt{n}}\Big)-(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\}\right| = o_p(1), \quad (2.7)$$

where

$$\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) = \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))}\psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}), \quad r = 1, \ldots, p, \tag{2.8}$$

and $f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) = (\partial/\partial\theta_r)f(\mathbf{x}_i, \boldsymbol{\theta})$.

Proof. For $l, r = 1, \ldots, p$, by differentiating we get,

$$\begin{aligned}(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) = k_1(\mathbf{x}_i, \boldsymbol{\theta})\,&\{\psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))(\partial^2/\partial\theta_r\partial\theta_l)f(\mathbf{x}_i, \boldsymbol{\theta}) \\ &-\psi'(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta})f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\} \\ &-k_2(\mathbf{x}_i, \boldsymbol{\theta})\psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta})f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta}),\end{aligned} \tag{2.9}$$

and $k_1(\mathbf{x}_i, \boldsymbol{\theta})$ and $k_2(\mathbf{x}_i, \boldsymbol{\theta})$ are given in (2.5) and (2.6), respectively. Then,

$$\begin{aligned}&\sup_{\|\mathbf{t}\|\le C}\left|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l\Big\{(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\}\right| \\ &\quad\le \frac{1}{n}C\sum_{i,j}\sum_{l=1}^{p}\sup_{\|\mathbf{t}\|\le C}\Big|(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big|,\end{aligned}$$

and

$$\begin{aligned}&\sup_{\|\mathbf{t}\|\le C}\Big|(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big| \\ &\le \sup_{\|\mathbf{t}\|\le C}\Big\{\Big|k_1\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - k_1(\mathbf{x}_i, \boldsymbol{\theta})\Big| \\ &\qquad\Big|\psi'\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big)f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\Big\} \\ &\quad + \sup_{\|\mathbf{t}\|\le C}\Big\{\Big|\psi'\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi'(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big| \\ &\qquad\Big|k_1(\mathbf{x}_i, \boldsymbol{\theta})f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\Big\} \\ &\quad + \sup_{\|\mathbf{t}\|\le C}\Big\{\Big|f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta})f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Big| \\ &\qquad\Big|k_1(\mathbf{x}_i, \boldsymbol{\theta})\psi'(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\Big\}\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{\|\mathbf{t}\|\le C} \Big\{ \Big| k_1\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - k_1(\mathbf{x}_i, \boldsymbol{\theta}) \Big| \\
&\qquad \Big| \psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) (\partial^2/\partial\theta_r\partial\theta_l) f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) \Big| \Big\} \\
&+ \sup_{\|\mathbf{t}\|\le C} \Big\{ \Big| \psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta})) \Big| \\
&\qquad \Big| k_1(\mathbf{x}_i, \boldsymbol{\theta}) (\partial^2/\partial\theta_r\partial\theta_l) f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) \Big| \Big\} \\
&+ \sup_{\|\mathbf{t}\|\le C} \Big\{ \Big| (\partial^2/\partial\theta_r\partial\theta_l) f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial^2/\partial\theta_r\partial\theta_l) f(\mathbf{x}_i, \boldsymbol{\theta}) \Big| \\
&\qquad \Big| k_1(\mathbf{x}_i, \boldsymbol{\theta}) \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta})) \Big| \Big\} \\
&+ \sup_{\|\mathbf{t}\|\le C} \Big\{ \Big| k_2\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - k_2(\mathbf{x}_i, \boldsymbol{\theta}) \Big| \\
&\qquad \Big| \psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) \Big| \Big\} \\
&+ \sup_{\|\mathbf{t}\|\le C} \Big\{ \Big| \psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta})) \Big| \\
&\qquad \Big| k_2(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) \Big| \Big\} \\
&+ \sup_{\|\mathbf{t}\|\le C} \Big\{ \Big| f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta}) \Big| \\
&\qquad \Big| k_2(\mathbf{x}_i, \boldsymbol{\theta}) \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta})) \Big| \Big\}.
\end{aligned}
$$

Then, by taking the expectation at both sides,

$$
\begin{aligned}
&E\Big\{ \sup_{\|\mathbf{t}\|\le C} \Big| (\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) \Big| \Big\} \\
&\le \sup_{\|\mathbf{t}\|\le C} \Big| k_1\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - k_1(\mathbf{x}_i, \boldsymbol{\theta}) \Big| E\Big\{ \sup_{\|\mathbf{t}\|\le C} \Big| \psi'\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) \Big| \Big\} \\
&\qquad \sup_{\|\mathbf{t}\|\le C} \Big| f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) \Big|
\end{aligned}
$$

$$
\begin{aligned}
&+ E\left\{\sup_{\|\mathbf{t}\|\le C}\Big|\psi'\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi'(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\right\}\\
&\qquad \sup_{\|\mathbf{t}\|\le C}\Big|k_1(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\\
&+ \sup_{\|\mathbf{t}\|\le C}\Big|f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Big|\\
&\qquad \Big|k_1(\mathbf{x}_i, \boldsymbol{\theta})\Big| E\Big\{\Big|\psi'(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\Big\}\\
&+ \sup_{\|\mathbf{t}\|\le C}\Big|k_1\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - k_1(\mathbf{x}_i, \boldsymbol{\theta})\Big| E\left\{\sup_{\|\mathbf{t}\|\le C}\Big|\psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big)\Big|\right\}\\
&\qquad \sup_{\|\mathbf{t}\|\le C}\Big|(\partial^2/\partial\theta_r\partial\theta_l) f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\\
&+ E\left\{\sup_{\|\mathbf{t}\|\le C}\Big|\psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\right\}\\
&\qquad \Big|k_1(\mathbf{x}_i, \boldsymbol{\theta})\Big| \sup_{\|\mathbf{t}\|\le C}\Big|(\partial^2/\partial\theta_r\partial\theta_l) f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\\
&+ \sup_{\|\mathbf{t}\|\le C}\Big|(\partial^2/\partial\theta_r\partial\theta_l) f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial^2/\partial\theta_r\partial\theta_l) f(\mathbf{x}_i, \boldsymbol{\theta})\Big|\\
&\qquad \Big|k_1(\mathbf{x}_i, \boldsymbol{\theta})\Big| E\Big\{\Big|\psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\Big\}\\
&+ \sup_{\|\mathbf{t}\|\le C}\Big|k_2\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - k_2(\mathbf{x}_i, \boldsymbol{\theta})\Big| E\left\{\sup_{\|\mathbf{t}\|\le C}\Big|\psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big)\Big|\right\}\\
&\qquad \sup_{\|\mathbf{t}\|\le C}\Big|f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\\
&+ E\left\{\sup_{\|\mathbf{t}\|\le C}\Big|\psi\Big(y_{ij} - f\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\right\} \Big|k_2(\mathbf{x}_i, \boldsymbol{\theta})\Big|\\
&\qquad \sup_{\|\mathbf{t}\|\le C}\Big|f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big)\Big|\\
&+ \sup_{\|\mathbf{t}\|\le C}\Big|f_{\theta_r}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_l}\Big(\mathbf{x}_i, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Big|\\
&\qquad \Big|k_2(\mathbf{x}_i, \boldsymbol{\theta})\Big| E\Big\{\Big|\psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))\Big|\Big\}.
\end{aligned}
$$

Thus, by conditions **[A3]** (i)-(ii), **[B3]** (i)-(ii), and **[C]** (i)-(ii), we have

$$E\left\{\sup_{\|\mathbf{t}\|\le C}\left|(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}+\frac{s\mathbf{t}}{\sqrt{n}}\Big)-(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\right|\right\}\longrightarrow 0,\quad \forall i, j,$$

and

$$E\left[\sup_{\|\mathbf{t}\|\le C}\left|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l\Big\{(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}+\frac{s\mathbf{t}}{\sqrt{n}}\Big)-(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\}\right|\right]\longrightarrow 0$$

Also,

$$\begin{aligned}
&\mathrm{Var}\left[\sup_{\|\mathbf{t}\|\le C}\left|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l\Big\{(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}+\frac{s\mathbf{t}}{\sqrt{n}}\Big)-(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\}\right|\right] \\
&\le \frac{C^2}{n^2}\sum_{i,j}\mathrm{Var}\left\{\sum_{l=1}^{p}\sup_{\|\mathbf{t}\|\le C}\left|(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}+\frac{s\mathbf{t}}{\sqrt{n}}\Big)-(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\right|\right\} \\
&\le C^2K/n\longrightarrow 0.
\end{aligned}$$

Therefore, we have the result in (2.7). □

Lemma 2.2. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ *be the rth element of the vector* $\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ *for* $r = 1, \ldots, p$. *Then for* $r = 1, \ldots, p$

$$\begin{aligned}
\sup_{\|\mathbf{t}\|\le C}\Bigg|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) \\
+\frac{\gamma}{n}\sum_{i,j}\sum_{l=1}^{p} t_l\frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))}f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta})f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Bigg| = o_p(1),
\end{aligned}\tag{2.10}$$

where $\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ *is defined in (2.8).*

Proof. From (2.9), we have

$$\sup_{\|\mathbf{t}\|\le C}\left|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) + \frac{\gamma}{n}\sum_{i,j}\sum_{l=1}^{p} t_l \frac{1}{w(f(\mathbf{x}_i,\boldsymbol{\theta}))} f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta})\right|$$

$$= \sup_{\|\mathbf{t}\|\le C}\left|\frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l\psi(y_{ij} - f(\mathbf{x}_i,\boldsymbol{\theta}))\big\{k_1(\mathbf{x}_i,\boldsymbol{\theta})(\partial^2/\partial\theta_r\partial\theta_l) f(\mathbf{x}_i,\boldsymbol{\theta}) - k_2(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta})\big\}\right.$$

$$\left. - \frac{1}{n}\sum_{i,j}\sum_{l=1}^{p} t_l k_1(\mathbf{x}_i,\boldsymbol{\theta})\big\{\psi'(y_{ij} - f(\mathbf{x}_i,\boldsymbol{\theta})) - \gamma\big\} f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta})\big\}\right|$$

$$\le C\sum_{l=1}^{p}\left|\frac{1}{n}\sum_{i,j}\psi(y_{ij} - f(\mathbf{x}_i,\boldsymbol{\theta}))\Big\{k_1(\mathbf{x}_i,\boldsymbol{\theta})(\partial^2/\partial\theta_r\partial\theta_l) f(\mathbf{x}_i,\boldsymbol{\theta}) - k_1(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta})\Big\}\right|$$

$$+ C\sum_{l=1}^{p}\left|\frac{1}{n}\sum_{i,j} k_1(\mathbf{x}_i,\boldsymbol{\theta})\big\{\psi'(y_{ij} - f(\mathbf{x}_i,\boldsymbol{\theta})) - \gamma\big\} f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta})\right|$$

which by using the Markov WLLN and conditions **[A2]** (i)-(ii) yields:

$$\frac{1}{n}\sum_{i,j}\psi(y_{ij} - f(\mathbf{x}_i,\boldsymbol{\theta}))\Big\{k_1(\mathbf{x}_i,\boldsymbol{\theta})(\partial^2/\partial\theta_r\partial\theta_l) f(\mathbf{x}_i,\boldsymbol{\theta}) - k_1(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta})\Big\} = o_p(1)$$

and

$$\frac{1}{n}\sum_{i,j} k_1(\mathbf{x}_i,\boldsymbol{\theta})\big\{\psi'(y_{ij} - f(\mathbf{x}_i,\boldsymbol{\theta})) - \gamma\big\} f_{\theta_r}(\mathbf{x}_i,\boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i,\boldsymbol{\theta}) = o_p(1).$$

Thus, we have the result in (2.10). □

Now we shall prove the uniform asymptotic linearity of M-statistics.

Theorem 2.3. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold. Then*

$$\sup_{\|\mathbf{t}\|\le C}\left\|\frac{1}{\sqrt{n}}\sum_{i,j}\{\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}) - \lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\} + \frac{\gamma}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\mathbf{t}\right\| = o_p(1) \qquad (2.11)$$

as $n \to \infty$, where $\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ was defined in (2.4).

Proof. We consider the rth element of the vector $\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})$ denoted as (2.8). Then, by using the first order term in the Taylor's expansion, we have, for $0 < s < 1$,

$$\begin{aligned}
&\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}) - \lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{l=1}^{p} t_l \{(\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\} \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{l=1}^{p} t_l \Big\{(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\}.
\end{aligned}$$

And for $r = 1, \ldots, p$ we have

$$\begin{aligned}
&\sup_{\|\mathbf{t}\| \le C} \Bigg| \frac{1}{\sqrt{n}} \sum_{i,j} \Big\{\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{\mathbf{t}}{\sqrt{n}}\Big) - \lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\} \\
&\qquad + \frac{\gamma}{n} \sum_{i,j} \sum_{l=1}^{p} \Big\{t_l \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))} f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Big\} \Bigg| \\
&\le \sup_{\|\mathbf{t}\| \le C} \Bigg| \frac{1}{n} \sum_{i,j} \sum_{l=1}^{p} t_l \Big\{(\partial/\partial\theta_l)\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{s\mathbf{t}}{\sqrt{n}}\Big) - (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\} \Bigg| \\
&\quad + \sup_{\|\mathbf{t}\| \le C} \Bigg| \frac{1}{n} \sum_{i,j} \sum_{l=1}^{p} t_l (\partial/\partial\theta_l)\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) \\
&\qquad + \frac{\gamma}{n} \sum_{i,j} \sum_{l=1}^{p} \Big\{t_l \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))} f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Big\} \Bigg|.
\end{aligned}$$

Therefore, from Lemma 2.1 and 2.2 we conclude that:

$$\begin{aligned}
&\sup_{\|\mathbf{t}\| \le C} \Bigg| \frac{1}{\sqrt{n}} \sum_{i,j} \Big\{\lambda_r\Big(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + \frac{\mathbf{t}}{\sqrt{n}}\Big) - \lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\Big\} \\
&\qquad + \frac{\gamma}{n} \sum_{i,j} \sum_{l=1}^{p} \Big\{t_l \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))} f_{\theta_r}(\mathbf{x}_i, \boldsymbol{\theta}) f_{\theta_l}(\mathbf{x}_i, \boldsymbol{\theta})\Big\} \Bigg| = o_p(1), \quad r = 1, \ldots, p.
\end{aligned}$$

□

We now prove the existence of a solution to (2.3) that is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}$ and admits an asymptotic representation.

Theorem 2.4. *Let the conditions* [**A1**]-[**A3**], [**B1**]-[**B3**], *and* [**C**] *hold. Then there exists a sequence* $\hat{\boldsymbol{\theta}}_n$ *of solutions of (2.3) such that:*

$$\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = O_p(1), \tag{2.12}$$

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + \frac{1}{n\gamma}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\sum_{i,j}\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) + o_p(n^{-\frac{1}{2}}) \tag{2.13}$$

or

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \frac{1}{\gamma}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i,j}\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) + o_p(1) \tag{2.14}$$

Proof. From Theorem 2.3 we know that the following system of equations:

$$\sum_{i,j}\lambda_r(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}) = 0$$

has a root $\mathbf{t}_n$ that lies in $\|\mathbf{t}\| \le C$ with probability exceeding $1-\epsilon$ for $n \ge n_0$. Then $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}_n$ is a solution of the equation (2.3) satisfying:

$$P(\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \le C) \ge 1-\epsilon \quad \text{for } n \ge n_0.$$

Inserting $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ into $\mathbf{t}$ in (2.7), we have the expression in (2.13) or (2.14). □

Theorem 2.5. *Let the conditions* [**A1**], [**A2**]*(i)-(ii),* [**B1**], [**B2**] *(i)-(ii) hold. Then*

$$\frac{1}{\sqrt{n}}\sum_{i,j}\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) \longrightarrow N_p\big(0, \sigma_\psi^2\boldsymbol{\Gamma}_2(\boldsymbol{\theta})\big) \quad \text{as } n\to\infty. \tag{2.15}$$

Proof. Let

$$Z_n^* = \boldsymbol{\eta}^t\frac{1}{\sqrt{n}}\sum_{i,j}\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}), \quad \boldsymbol{\eta} \in \Re^p,$$

then

$$
\begin{aligned}
Z_n^* &= \frac{1}{\sqrt{n}} \sum_{i,j} \frac{1}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))} \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta})) \boldsymbol{\eta}^t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \\
&= \sum_{i,j} c_{nij} Z_{ij},
\end{aligned}
$$

where

$$
c_{nij} = \frac{\sigma_\psi}{\sqrt{n}} \frac{\sqrt{u(\mathbf{x}_i)}}{w(f(\mathbf{x}_i, \boldsymbol{\theta}))} \boldsymbol{\eta}^t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}),
$$

and

$$
Z_{ij} = \psi(y_{ij} - f(\mathbf{x}_i, \boldsymbol{\theta}))/(\sigma_\psi \sqrt{u(\mathbf{x}_i)}).
$$

Then by using the Hájek-Šidak Central Limit Theorem, we show that Z_n^* converges in law to a normal distribution as $n \to \infty$. In order to use this theorem we need to verify that

$$
\max_{i,j} c_{nij}^2 / \sum_{i,j} c_{nij}^2 \longrightarrow 0,
$$

as $n \to \infty$, which can be reformulated as requiring

$$
\sup_{\boldsymbol{\eta} \in \Re^p} \left[\max_{i,j} \boldsymbol{\eta}^t \frac{u(\mathbf{x}_i)}{w^2(f(\mathbf{x}_i, \boldsymbol{\theta}))} \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta}) \boldsymbol{\eta} \ / \ \boldsymbol{\eta}^t \boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) \boldsymbol{\eta} \right] \longrightarrow 0.
$$

Now, in view of the Courant's Theorem, we have that:

$$
\begin{aligned}
&\sup_{\boldsymbol{\eta} \in \Re^p} \left[\boldsymbol{\eta}^t \frac{u(\mathbf{x}_i)}{w^2(f(\mathbf{x}_i, \boldsymbol{\theta}))} \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta}) \boldsymbol{\eta} \ / \ \boldsymbol{\eta}^t \boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) \boldsymbol{\eta} \right] \\
&= ch_1 \left\{ \frac{u(\mathbf{x}_i)}{w^2(f(\mathbf{x}_i, \boldsymbol{\theta}))} \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta}) \big(\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) \big)^{-1} \right\} \\
&= \frac{u(\mathbf{x}_i)}{w^2(f(\mathbf{x}_i, \boldsymbol{\theta}))} \mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta}) \big(\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) \big)^{-1} \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}),
\end{aligned}
$$

so this condition reduces to the condition **[B2]** (iii) (Noether's condition). Thus, we conclude that:

$$
Z_n^* / \big(\sum_{i,j} c_{nij}^2 \big)^{\frac{1}{2}} \longrightarrow N(0,1) \quad \text{as} \quad n \to \infty
$$

and by using Cramer-Wold Theorem and condition **[B2]** (ii) we prove the expression in (2.15). □

Corollary 2.6. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]** *hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \longrightarrow N_p\big(\mathbf{0}, \gamma^{-2}\sigma_\psi^2 \boldsymbol{\Gamma}_1^{-1}(\boldsymbol{\theta})\boldsymbol{\Gamma}_2(\boldsymbol{\theta})\boldsymbol{\Gamma}_1^{-1}(\boldsymbol{\theta})\big) \quad \textit{as } n \to \infty. \tag{2.16}$$

Proof. From Theorem 2.4 we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \frac{1}{\gamma}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i,j}\lambda(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta}) + o_p(1)$$

Then from Theorem 2.5 and Slutsky's Theorem we have the expression in (2.16). □

Corollary 2.7. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]** *hold. Then*

$$\hat{\boldsymbol{\Gamma}}^{-\frac{1}{2}}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \longrightarrow N_p(\mathbf{0}, \mathbf{I}_p) \quad \textit{as } n \to \infty, \tag{2.17}$$

where

$$\hat{\boldsymbol{\Gamma}} = \hat{\gamma}^{-2}\hat{\sigma}_\psi^2\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\hat{\boldsymbol{\theta}}_n)\Big)^{-1}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{2n}(\hat{\boldsymbol{\theta}}_n)\Big)\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\hat{\boldsymbol{\theta}}_n)\Big)^{-1}, \tag{2.18}$$

and $\hat{\gamma}$ *and* $\hat{\sigma}_\psi^2$ *are consistent estimators of* γ *and* σ_ψ^2, *respectively.*

Proof. Using (2.16) and Slutsky's Theorem we have the expression in (2.17). □

Corollary 2.8. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]** *hold. Then*

$$n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^t\hat{\boldsymbol{\Gamma}}^{-1}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \longrightarrow \chi_p^2 \quad \textit{as } n \to \infty. \tag{2.19}$$

Proof. Using (2.17) and Cochran's Theorem, we prove the expression in (2.19). □

Theorem 2.9. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \mathbf{0}$$

as $n \to \infty$, where $\hat{\boldsymbol{\theta}}_n^*$ is an M-estimator defined as (2.2) with no assumption of fixed variance function and obtained iteratively with updated variance function estimates.

Proof. From (2.13) we have

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + \frac{1}{n\gamma}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\sum_{i,j} k_1(\mathbf{x}_i,\boldsymbol{\theta})\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) + o_p(n^{-\frac{1}{2}}).$$

Also, $\hat{\boldsymbol{\theta}}_n^*$ can be expressed as

$$\hat{\boldsymbol{\theta}}_n^* = \boldsymbol{\theta} + \frac{1}{n\gamma}\Big(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{1n}(\boldsymbol{\theta})\Big)^{-1}\sum_{i,j} \hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) + o_p(n^{-\frac{1}{2}}),$$

where

$$\hat{\boldsymbol{\Gamma}}_{1n}(\boldsymbol{\theta}) = \sum_{i,j}\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta})\mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i,\boldsymbol{\theta})$$

and $\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})$ is an estimate of $k_1(\mathbf{x}_i,\boldsymbol{\theta})$, possibly $k_1(\mathbf{x}_i,\hat{\boldsymbol{\theta}}_n)$.

Therefore, we may write:

$$\begin{aligned}
&\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n \\
&= \frac{1}{n\gamma}\Big\{\Big(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{1n}(\boldsymbol{\theta})\Big)^{-1} - \Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\Big\}\sum_{i,j}\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) \\
&\quad + \frac{1}{n\gamma}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\sum_{i,j}\big\{\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta}) - k_1(\mathbf{x}_i,\boldsymbol{\theta})\big\}\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) + o_p(n^{-\frac{1}{2}}) \\
&= \frac{1}{n\gamma}\Big\{\Big(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{1n}(\boldsymbol{\theta})\Big)^{-1} - \Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\Big\}\sum_{i,j}k_1(\mathbf{x}_i,\boldsymbol{\theta})\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) \\
&\quad + \frac{1}{n\gamma}\Big\{\Big(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{1n}(\boldsymbol{\theta})\Big)^{-1} - \Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\Big\}\sum_{i,j}\big\{\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta}) - k_1(\mathbf{x}_i,\boldsymbol{\theta})\big\}\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) \\
&\quad + \frac{1}{n\gamma}\Big(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\Big)^{-1}\sum_{i,j}\big\{\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta}) - k_1(\mathbf{x}_i,\boldsymbol{\theta})\big\}\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta}) + o_p(n^{-\frac{1}{2}}). \qquad (2.20)
\end{aligned}$$

The first term on the right hand side (r.h.s.) of (2.20) may be expressed as

$$\frac{1}{n\gamma}\left\{\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}-\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}\right\}\left\{n\gamma\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})+o_p(n^{-\frac{1}{2}})\right\}$$

$$=\left\{\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)-\mathbf{I}_n\right\}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})+o_p(n^{-\frac{1}{2}}) \tag{2.21}$$

From (2.12), (2.21) is $o_p(n^{-\frac{1}{2}})$ whenever $\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)-\mathbf{I}_n=o_p(1)$, i.e., all the charateristic roots of $\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)-\mathbf{I}_n$ are $o_p(1)$. Similarly, the second term on the r.h.s. of (2.20) may be expressed as

$$\begin{aligned}&\left\{\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)-\mathbf{I}_n\right\}\frac{1}{n\gamma}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}\\&\times\sum_{i,j}k_1(\mathbf{x}_i,\boldsymbol{\theta})\left\{\left(k_1(\mathbf{x}_i,\boldsymbol{\theta})\right)^{-1}\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})-1\right\}\psi(\epsilon_{ij})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i,\boldsymbol{\theta})\end{aligned} \tag{2.22}$$

Thus, if $\left(k_1(\mathbf{x}_i,\boldsymbol{\theta})\right)^{-1}\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})-1$ are all $o_p(1)$ for $i=1,\dots,k$, whereas all the charateristic roots of $\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)-\mathbf{I}_n$ are $o_p(1)$, (2.22) is $o_p(n^{-\frac{1}{2}})$. A very similar treatment holds for the last term on the r.h.s. of (2.20). Thus, we conclude that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^*-\hat{\boldsymbol{\theta}}_n)\xrightarrow{p}\mathbf{0}$$

if

$$\text{ch}_1\left\{\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)\right\}-1=o_p(1)=\text{ch}_n\left\{\left(\frac{1}{n}\hat{\mathbf{\Gamma}}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\mathbf{\Gamma}_{1n}(\boldsymbol{\theta})\right)\right\}-1 \tag{2.23}$$

and

$$\left(k_1(\mathbf{x}_i,\boldsymbol{\theta})\right)^{-1}\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})-1=o_p(1)\ \text{ for } i=1,\dots,k. \tag{2.24}$$

However, (2.23) and (2.24) are equvalent to $\hat{k}_1(\mathbf{x}_i,\boldsymbol{\theta})-k_1(\mathbf{x}_i,\boldsymbol{\theta})=o_p(1)$ for $i=1,\dots,k$, which are guaranteed from the assumption **[C]** (i). □

2.4 Verification of the Conditions in Hill Model

In this section, we use L_1-norm and the Huber function as h, L_1-norm as w and f is defined as

$$f(x, \boldsymbol{\theta}) = \theta_0 + \frac{\theta_1 x^{\theta_2}}{\theta_3^{\theta_2} + x^{\theta_2}},$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)^t$. And we shall verify that those functions satisfy the regularity conditions in the previous section in each case.

2.4.1 h : $\mathbf{L_1}$-norm and w : $\mathbf{L_1}$-norm

In this case, we have

$$h(z) = |z|^{\frac{1}{2}} \quad \text{and} \quad w(z) = |z|.$$

Then, since $\psi(z) = (\partial/\partial z)h^2(z)$,

$$\psi(z) = \text{sign}(z) = \begin{cases} +1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases}$$

[A1]: Clearly, ψ is not absolutely continuous. However, since only at $z = 0$ ψ is not continuous and in our problem single point has probability zero, we may consider the condition is still satisfied.

[A2]: $\epsilon = Y - f(x, \boldsymbol{\theta}) \sim N(0, \sigma_x^2)$

(i) $E\psi^2(\epsilon) = E(\text{sign}(\epsilon))^2 = 1 < \infty$, so $\sigma_\psi^2 = 1$ and $u(x) = 1$.

$E\psi(\epsilon) = E\text{sign}(\epsilon) = (+1)P(\epsilon > 0) + (-1)P(\epsilon < 0) = 0.$

(ii) Since $\psi'(z) = 0$ if $z \neq 0$ and not defined if $z = 0$, we compute $E\psi'(\epsilon)$ indirectly.

For arbitrary small $h > 0$,

$$
\begin{aligned}
E\psi'(\epsilon) &\approx h^{-1}\int \Big\{\psi(x+h) - \psi(x)\Big\} f(x)dx \\
&= h^{-1}\left\{\int \psi(x+h)f(x)dx - \int \psi(x)f(x)dx\right\} \\
&= h^{-1}\left\{\int \psi(x)f(x-h)dx - \int \psi(x)f(x)dx\right\} \\
&= h^{-1}\int \psi(x)\Big\{f(x-h) - f(x)\Big\}dx \\
&\approx \int \psi(x)\left\{-\tfrac{f'(x)}{f(x)}\right\} f(x)dx,
\end{aligned}
$$

where f is a density function of ϵ. Thus, if we have the assumption of finite Fisher information, we can verify $E\psi'(\epsilon) = \gamma(\neq 0)$. We can do similar job for $E|\psi'(\epsilon)|^{1+\delta}$ for some $0 < \delta \leq 1$.

[A3]:

(i) $\lim_{\delta\to 0} E\Big\{\sup_{||\boldsymbol{\Delta}||\leq\delta}\Big|\psi(Y - f(x,\boldsymbol{\theta}+\boldsymbol{\Delta})) - \psi(Y - f(x,\boldsymbol{\theta}))\Big|\Big\}$

$= \lim_{\delta\to 0}\Big[\sup_{||\boldsymbol{\Delta}||\leq\delta}\Big|\text{sign}(Y - f(x,\boldsymbol{\theta}+\boldsymbol{\Delta})) - \text{sign}(0)\Big|P\big(Y = f(x,\boldsymbol{\theta})\big)$

$+ \sup_{||\boldsymbol{\Delta}||\leq\delta}\Big|\text{sign}(Y - f(x,\boldsymbol{\theta}+\boldsymbol{\Delta})) - \text{sign}(Y - f(x,\boldsymbol{\theta}))\Big|P\big(Y \neq f(x,\boldsymbol{\theta})\big)\Big]$

$= 0$

(ii) $\lim_{\delta\to 0} E\Big\{\sup_{||\boldsymbol{\Delta}||\leq\delta}\Big|\psi'(Y - f(x,\boldsymbol{\theta}+\boldsymbol{\Delta})) - \psi'(Y - f(x,\boldsymbol{\theta}))\Big|\Big\} = 0$

[B1]: $f(x,\boldsymbol{\theta}) = \theta_0 + \dfrac{\theta_1 x^{\theta_2}}{\theta_3^{\theta_2} + x^{\theta_2}}$ Obviously f is continous. Also,

$$
\frac{\partial f(x,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \left(1, \frac{x^{\theta_2}}{\theta_3^{\theta_2} + x^{\theta_2}}, \frac{\theta_1 x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^2}, -\frac{\theta_1 x^{\theta_2}\theta_2\theta_3^{\theta_2-1}}{(\theta_3^{\theta_2} + x^{\theta_2})^2}\right)^t
$$

$$
\frac{\partial^2 f}{\partial\theta_0^2} = \frac{\partial^2 f}{\partial\theta_0\partial\theta_1} = \frac{\partial^2 f}{\partial\theta_0\partial\theta_2} = \frac{\partial^2 f}{\partial\theta_0\partial\theta_3} = \frac{\partial^2 f}{\partial\theta_1^2} = 0
$$

$$
\frac{\partial^2 f}{\partial\theta_1\partial\theta_2} = \frac{x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^2} \qquad \frac{\partial^2 f}{\partial\theta_1\partial\theta_3} = -\frac{x^{\theta_2}\theta_2\theta_3^{\theta_2-1}}{(\theta_3^{\theta_2} + x^{\theta_2})^2}
$$

$$
\frac{\partial^2 f}{\partial\theta_2^2} = \frac{\theta_1 x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} - x^{\theta_2})}{(\theta_3^{\theta_2} + x^{\theta_2})^3}
$$

$$\frac{\partial^2 f}{\partial\theta_2\partial\theta_3} = \frac{\theta_1 x^{\theta_2}(\theta_2\theta_3^{\theta_2-1}\log(x/\theta_3) - \theta_3^{\theta_2-1})(\theta_3^{\theta_2}+x^{\theta_2}) - 2\theta_1\theta_2 x^{\theta_2}\theta_3^{2\theta_2-1}\log(x/\theta_3)}{(\theta_3^{\theta_2}+x^{\theta_2})^3}$$

$$\frac{\partial^2 f}{\partial\theta_3^2} = \frac{-\theta_1 x^{\theta_2}\theta_2(\theta_2-1)\theta_3^{\theta_2-2}(\theta_3^{\theta_2}+x^{\theta_2}) + 2\theta_1\theta_2^2 x^{\theta_2}\theta_3^{2\theta_2-2}}{(\theta_3^{\theta_2}+x^{\theta_2})^3}$$

Therefore, f is twice differentiable with respect to $\boldsymbol{\theta}$.

[B2]: It is natural in nonlinear regression models that we assume (i), (ii) and (iii) here. Recall that

(i) $\lim_{n\to\infty}\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_1(\boldsymbol{\theta})$, where

$$\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) = \sum_{i,j}\left\{\frac{1}{w(f(x_i,\boldsymbol{\theta}))}\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})\mathbf{f}_{\boldsymbol{\theta}}^t(x_i,\boldsymbol{\theta})\right\},$$

and $\boldsymbol{\Gamma}_1(\boldsymbol{\theta})$ is a positive definite matrix.

(ii) $\lim_{n\to\infty}\frac{1}{n}\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_2(\boldsymbol{\theta})$, where

$$\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) = \sum_{i,j}\left\{\frac{u(x_i)}{w^2(f(x_i,\boldsymbol{\theta}))}\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})\mathbf{f}_{\boldsymbol{\theta}}^t(x_i,\boldsymbol{\theta})\right\},$$

and $\boldsymbol{\Gamma}_2(\boldsymbol{\theta})$ is a positive definite matrix.

(iii) $\max\left\{\frac{u(x_i)}{w^2(f(x_i,\boldsymbol{\theta}))}\mathbf{f}_{\boldsymbol{\theta}}^t(x_i,\boldsymbol{\theta})(\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}))^{-1}\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})\right\} \longrightarrow 0$, as $n\to\infty$

Here,

$\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})\mathbf{f}_{\boldsymbol{\theta}}^t(x_i,\boldsymbol{\theta})$

$$= \begin{pmatrix} 1 & \frac{x_i^{\theta_2}}{\theta_3^{\theta_2}+x_i^{\theta_2}} & \frac{\theta_1 x_i^{\theta_2}\theta_3^{\theta_2}\log(x_i/\theta_3)}{(\theta_3^{\theta_2}+x_i^{\theta_2})^2} & -\frac{\theta_1 x_i^{\theta_2}\theta_2\theta_3^{\theta_2-1}}{(\theta_3^{\theta_2}+x_i^{\theta_2})^2} \\ & \frac{x_i^{2\theta_2}}{(\theta_3^{\theta_2}+x_i^{\theta_2})^2} & \frac{\theta_1 x_i^{2\theta_2}\theta_3^{\theta_2}\log(x_i/\theta_3)}{(\theta_3^{\theta_2}+x_i^{\theta_2})^3} & -\frac{\theta_1 x_i^{2\theta_2}\theta_2\theta_3^{\theta_2-1}}{(\theta_3^{\theta_2}+x_i^{\theta_2})^3} \\ \text{symm.} & & \frac{\theta_1^2 x_i^{2\theta_2}\theta_3^{2\theta_2}\log(x_i/\theta_3)^2}{(\theta_3^{\theta_2}+x_i^{\theta_2})^4} & -\frac{\theta_1^2 x_i^{2\theta_2}\theta_2\theta_3^{2\theta_2-1}\log(x_i/\theta_3)}{(\theta_3^{\theta_2}+x_i^{\theta_2})^4} \\ & & & \frac{\theta_1^2 x_i^{2\theta_2}\theta_2^2\theta_3^{2\theta_2-2}}{(\theta_3^{\theta_2}+x_i^{\theta_2})^4} \end{pmatrix}$$

Also, $w(f(x_i, \boldsymbol{\theta})) = |f(x_i, \boldsymbol{\theta})| = f(x_i, \boldsymbol{\theta})$ and $u(x_i) = 1$.

[B3]: (i) For $j, l = 0$,

$$\left(\frac{\partial}{\partial\theta_0} f(x, \boldsymbol{\theta} + \boldsymbol{\Delta})\right)^2 - \left(\frac{\partial}{\partial\theta_0} f(x, \boldsymbol{\theta})\right)^2 = 1 - 1 = 0$$

For $j = 0$ and $l = 1$,

$$\begin{aligned}&\left(\frac{\partial}{\partial\theta_0} f(x, \boldsymbol{\theta} + \boldsymbol{\Delta})\right)\left(\frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta} + \boldsymbol{\Delta})\right) - \left(\frac{\partial}{\partial\theta_0} f(x, \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta})\right)\\ &\quad = \frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta} + \boldsymbol{\Delta}) - \frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta})\end{aligned}$$

However, since f is twice differentiable, $(\partial/\partial\theta_1) f(x, \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$. Thus, we have

$$\lim_{\delta \to 0} \sup_{||\boldsymbol{\Delta}|| \leq \delta} \left| \frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta} + \boldsymbol{\Delta}) - \frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta}) \right| = 0.$$

Similarly, the assumptions are satisfied for $j = 0$; $l = 2$ and $j = 0$; $l = 3$. For $j = 1$ and $l = 1$, we define $g(x, \boldsymbol{\theta})$ as

$$g(x, \boldsymbol{\theta}) \equiv \left(\frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta})\right)^2 = \frac{x^{2\theta_2}}{(\theta_3^{\theta_2} + x^{\theta_2})^2}.$$

Then, similarly, since g is differentiable, the assumption is satisfied:

$$\frac{\partial g}{\partial\theta_0} = \frac{\partial g}{\partial\theta_1} = 0$$

$$\frac{\partial g}{\partial\theta_2} = 2x^{2\theta_2} \log x (\theta_3^{\theta_2} + x^{\theta_2})^{-2} - 2x^{2\theta_2} (\theta_3^{\theta_2} + x^{\theta_2})^{-3} (\theta_3^{\theta_2} \log \theta_3 + x^{\theta_2} \log x)$$

$$\frac{\partial g}{\partial\theta_3} = -\frac{2x^{2\theta_2} \theta_2 \theta_3^{\theta_2 - 1}}{(\theta_3^{\theta_2} + x^{\theta_2})^3}.$$

For $j = 1$ and $l = 2$, Similarly,

$$g(x, \boldsymbol{\theta}) \equiv \left(\frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta})\right) \left(\frac{\partial}{\partial\theta_2} f(x, \boldsymbol{\theta})\right) = \frac{\theta_1 x^{2\theta_2} \theta_3^{\theta_2} \log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

Then,

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = \frac{x^{2\theta_2} \theta_3^{\theta_2} \log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\begin{aligned}\frac{\partial g}{\partial\theta_2} &= \theta_1 x^{2\theta_2} \theta_3^{\theta_2} (\log(x/\theta_3)(2\log x + \log\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\ &\quad -3\theta_1 x^{2\theta_2} \theta_3^{\theta_2} \log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-4}(\theta_3^{\theta_2} \log\theta_3 + x^{\theta_2} \log x) \\ \frac{\partial g}{\partial\theta_3} &= \theta_1 x^{2\theta_2} (\theta_2 \theta_3^{\theta_2 - 1} \log(x/\theta_3) - \theta_3^{\theta_2 - 1})(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\ &\quad -3\theta_1 x^{2\theta_2} \theta_3^{\theta_2} \log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-4} \theta_2 \theta 3^{\theta_2 - 1}\end{aligned}$$

For $j = 1$ and $l = 3$,

$$g(x, \boldsymbol{\theta}) \equiv \left(\frac{\partial}{\partial\theta_1} f(x, \boldsymbol{\theta})\right) \left(\frac{\partial}{\partial\theta_3} f(x, \boldsymbol{\theta})\right) = -\frac{\theta_1 x^{2\theta_2} \theta_2 \theta_3^{\theta_2 - 1}}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = -\frac{x^{2\theta_2} \theta_2 \theta_3^{\theta_2 - 1}}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\begin{aligned}\frac{\partial g}{\partial\theta_2} &= -\theta_1 \theta_3^{-1} (x^{2\theta_2} \theta_3^{\theta_2} (2\log x + \log\theta_3)\theta_2 + x^{2\theta_2} \theta_3^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\ &\quad +3\theta_1 x^{2\theta_2} \theta_2 \theta_3^{\theta_2 - 1} (\theta_3^{\theta_2} + x^{\theta_2})^{-4} (\theta_3^{\theta_2} \log\theta_3 + x^{\theta_2} \log x) \\ \frac{\partial g}{\partial\theta_3} &= -\theta_1 x^{2\theta_2} \theta_2 (\theta_2 - 1) \theta_3^{\theta_2 - 2} (\theta_3^{\theta_2} + x^{\theta_3})^{-3} + 3\theta_1 x^{2\theta_2} \theta_2 \theta_3^{\theta_2 - 1} (\theta_3^{\theta_2} + x^{\theta_2})^{-4} \theta_2 \theta_3^{\theta_2 - 1}\end{aligned}$$

For $j, l = 2$,

$$g(x, \boldsymbol{\theta}) \equiv \left(\frac{\partial}{\partial\theta_2} f(x, \boldsymbol{\theta})\right)^2 = \frac{\theta_1^2 x^{2\theta_2} \theta_3^{2\theta_2} \log(x/\theta_3)^2}{(\theta_3^{\theta_2} + x^{\theta_2})^4}$$

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = \frac{2\theta_1 x^{2\theta_2} \theta_3^{2\theta_2} \log(x/\theta_3)^2}{(\theta_3^{\theta_2} + x^{\theta_2})^4}$$

$$\frac{\partial g}{\partial \theta_2} = \theta_1^2 \log(x/\theta_3)^2 x^{2\theta_2}\theta_3^{2\theta_2} 2(\log x + \log\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-4}$$
$$-4\theta_1^2 x^{2\theta_2}\theta_3^{2\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} + x^{\theta_2})^{-5}(\theta_3^{\theta_2}\log\theta_3 + x^{\theta_2}\log x)$$
$$\frac{\partial g}{\partial \theta_3} = \theta_1^2 x^{2\theta_2}(2\theta_2\theta_3^{2\theta_2-1}\log(x/\theta_3)^2 - 2\theta_3^{2\theta_2-1}\log(x/\theta_3))(\theta_3^{\theta_2} + x^{\theta_2})^{-4}$$
$$-4\theta_1^2 x^{2\theta_2}\theta_3^{2\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} + x^{\theta_2})^{-5}\theta_2\theta_3^{\theta_2-1}$$

For $j = 2$ and $l = 3$

$$g(x,\boldsymbol{\theta}) \equiv \left(\frac{\partial}{\partial\theta_2} f(x,\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_3} f(x,\boldsymbol{\theta})\right) = -\frac{\theta_1 x^{2\theta_2}\theta_2\theta_3^{2\theta_2-1}\log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^4}$$

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = -\frac{2\theta_1 x^{2\theta_2}\theta_2\theta_3^{2\theta_2-1}\log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^4}$$

$$\frac{\partial g}{\partial \theta_2} = -\theta_1^2\log(x/\theta_3)\theta_3^{-1}(x^{2\theta_2}\theta_2\theta_3^{2\theta_2}2(\log x + \log\theta_3) + x^{2\theta_2}\theta_3^{2\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-4}$$
$$+4\theta_1^2 x^{2\theta_2}\theta_2\theta_3^{2\theta_2-1}\log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-5}(\theta_3^{\theta_2}\log\theta_3 + x^{\theta_2}\log x)$$
$$\frac{\partial g}{\partial \theta_3} = -\theta_1^2 x^{2\theta_2}\theta_2((2\theta_2 - 1)\theta_3^{2\theta_2-2}\log(x/\theta_3) - \theta_3^{2\theta_2-2})(\theta_3^{\theta_2} + x^{\theta_2})^{-4}$$
$$+4\theta_1^2 x^{2\theta_2}\theta_2\theta_3^{2\theta_2-1}\log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-5}\theta_2\theta_3^{\theta_2-1}$$

For $j, l = 3$,

$$g(x,\boldsymbol{\theta}) \equiv \left(\frac{\partial}{\partial\theta_3} f(x,\boldsymbol{\theta})\right)^2 = \frac{\theta_1^2 x^{2\theta_2}\theta_2^2\theta_3^{2\theta_2-2}}{(\theta_3^{\theta_2} + x^{\theta_2})^4}$$

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = \frac{2\theta_1 x^{2\theta_2}\theta_2^2\theta_3^{2\theta_2-2}}{(\theta_3^{\theta_2} + x^{\theta_2})^4}$$

$$\frac{\partial g}{\partial \theta_2} = \theta_1^2\theta_3^{-2}(x^{2\theta_2}\theta_3^{2\theta_2}2(\log x + \log\theta_3)\theta_2^2 + x^{2\theta_2}\theta_3^{2\theta_2}2\theta_2)(\theta_3^{\theta_2} + x^{\theta_2})^{-4}$$
$$-4\theta_1^2 x^{2\theta_2}\theta_2^2\theta_3^{2\theta_2-2}(\theta_3^{\theta_2} + x^{\theta_2})^{-5}(\theta_3^{\theta_2}\log\theta_3 + x^{\theta_2}\log x)$$
$$\frac{\partial g}{\partial \theta_3} = \theta_1^2 x^{2\theta_2}\theta_2^2(2\theta_2 - 2)\theta_3^{2\theta_2-3}(\theta_3^{\theta_2} + x^{\theta_2})^{-4} - 4\theta_1^2 x^{2\theta_2}\theta_2^2\theta_3^{2\theta_2-2}(\theta_3^{\theta_2} + x^{\theta_2})^{-5}\theta_2\theta_3^{\theta_2-1}$$

(ii) For $(j,l) = (0,0), (0,1), (0,2), (0,3)$, and $(1,1)$, since the derivatives are all zero, the assumptions are satisfied. For $j = 1$ and $l = 2$, simialarly we define $g(x,\boldsymbol{\theta})$ as

$$g(x,\boldsymbol{\theta}) \equiv \frac{\partial^2}{\partial\theta_1\partial\theta_2} f(x,\boldsymbol{\theta}) = \frac{x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)}{(\theta_3^{\theta_2} + x^{\theta_2})^2}$$

Then, since we have

$$\frac{\partial g}{\partial \theta_0} = \frac{\partial g}{\partial \theta_1} = 0$$

$$\begin{aligned}
\frac{\partial g}{\partial \theta_2} &= x^{\theta_2}\theta_3^{\theta_2}(\log x + \log\theta_3)\log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-2} \\
&\quad -2x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-3}(\theta_3^{\theta_2}\log\theta_3 + x^{\theta_2}\log x) \\
\frac{\partial g}{\partial \theta_3} &= x^{\theta_2}(\theta_2\theta_3^{\theta_2-1}\log(x/\theta_3) - \theta_3^{\theta_2-1})(\theta_3^{\theta_2} + x^{\theta_2})^{-2} \\
&\quad -2x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)(\theta_3^{\theta_2} + x^{\theta_2})^{-3}\theta_2\theta_3^{\theta_2-1},
\end{aligned}$$

the assumption is satisfied. For $j = 1$ and $l = 3$,

$$g(x, \boldsymbol{\theta}) \equiv \frac{\partial^2}{\partial\theta_1\partial\theta_3} f(x, \boldsymbol{\theta}) = -\frac{x^{\theta_2}\theta_2\theta_3^{\theta_2-1}}{(\theta_3^{\theta_2} + x^{\theta_2})^2}$$

$$\frac{\partial g}{\partial \theta_0} = \frac{\partial g}{\partial \theta_1} = 0$$

$$\begin{aligned}
\frac{\partial g}{\partial \theta_2} &= -\theta_3^{\theta_2-1}(x^{\theta_2}\theta_2\log x + x^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-2} \\
&\quad +2x^{\theta_2}\theta_2\theta_3^{\theta_2-1}(\theta_3^{\theta_2} + x^{\theta_2})^{-3}(\theta_3^{\theta_2}\log\theta_3 + x^{\theta_2}\log x) \\
\frac{\partial g}{\partial \theta_3} &= -x^{\theta_2}\theta_2(\theta_2 - 1)\theta_3^{\theta_2-2}(\theta_3^{\theta_2} + x^{\theta_2})^{-2} + 2x^{\theta_2}\theta_2\theta_3^{\theta_2-1}(\theta_3^{\theta_2} + x^{\theta_2})^{-3}\theta_2\theta_3^{\theta_2-1}
\end{aligned}$$

For $j, l = 2$

$$g(x, \boldsymbol{\theta}) \equiv \frac{\partial^2}{\partial\theta_2^2} f(x, \boldsymbol{\theta}) = \frac{\theta_1 x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} - x^{\theta_2})}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\frac{\partial g}{\partial \theta_0} = 0 \qquad \frac{\partial g}{\partial \theta_1} = \frac{x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} - x^{\theta_2})}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\begin{aligned}
\frac{\partial g}{\partial \theta_2} &= \theta_1(\log x - \log\theta_3)^2(x^{\theta_2}\theta_3^{\theta_2}(\log x + \log\theta_3)(\theta_3^{\theta_2} - x^{\theta_2}) \\
&\qquad\qquad +x^{\theta_2}\theta_3^{\theta_2}(\theta_3^{\theta_2}\log\theta_3 - x^{\theta_2}\log x))(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad -3\theta_1 x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} - x^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-4}(\theta_3^{\theta_2}\log\theta_3 + x^{\theta_2}\log x) \\
\frac{\partial g}{\partial \theta_3} &= \theta_1 x^{\theta_2}(2\log(x/\theta_3)(-\theta_3^{-1})(\theta_3^{2\theta_2} - x^{\theta_2}\theta_3^{\theta_2}) \\
&\qquad + \log(x/\theta_3)^2(2\theta_2\theta_3^{2\theta_2-1} - x^{\theta_2}\theta_2\theta_3^{\theta_2-1}))(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad -3\theta_1 x^{\theta_2}\theta_3^{\theta_2}\log(x/\theta_3)^2(\theta_3^{\theta_2} - x^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-4}\theta_2\theta_3^{\theta_2-1}
\end{aligned}$$

For $j = 2$ and $l = 3$,

$$g(x, \boldsymbol{\theta}) \equiv \frac{\partial^2}{\partial\theta_2 \partial\theta_3} f(x, \boldsymbol{\theta}) = \frac{\theta_1 x^{\theta_2} \theta_3^{\theta_2 - 1}(\theta_2 \log(x/\theta_3)(\theta_3^{\theta_2} - x^{\theta_2}) - (\theta_3^{\theta_2} + x^{\theta_2}))}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = \frac{x^{\theta_2} \theta_3^{\theta_2 - 1}(\theta_2 \log(x/\theta_3)(\theta_3^{\theta_2} - x^{\theta_2}) - (\theta_3^{\theta_2} + x^{\theta_2}))}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\begin{aligned}
\frac{\partial g}{\partial\theta_2} &= \theta_1 \theta_3^{-1} x^{\theta_2} \theta_3^{\theta_2} (\log x + \log\theta_3)(\theta_2 \log(x/\theta_3)(\theta_3^{\theta_2} - x^{\theta_2}) - (\theta_3^{\theta_2} + x^{\theta_2})) \\
&\quad \times (\theta_3^{\theta_2} + x^{\theta_2})^{-3} + \theta_1 x^{\theta_2} \theta_3^{\theta_2 - 1}(\log(x/\theta_3)(x^{\theta_2} - \theta_3^{\theta_2}) + \theta_2 \log(x/\theta_3) \\
&\quad \times (x^{\theta_2} \log x - \theta_3^{\theta_2} \log\theta_3) - (\theta_3^{\theta_2} \log\theta_3 + x^{\theta_2} \log x))(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad - 3\theta_1 x^{\theta_2} \theta_3^{\theta_2 - 1}(\theta_2 \log(x/\theta_3)(x^{\theta_2} - \theta_3^{\theta_2}) - (\theta_3^{\theta_2} + x^{\theta_2})) \\
&\quad \times (\theta_3^{\theta_2} + x^{\theta_2})^{-4}(\theta_3^{\theta_2} \log\theta_3 + x^{\theta_2} \log x) \\
\frac{\partial g}{\partial\theta_3} &= \theta_1 x^{\theta_2}(\theta_2 - 1)\theta_3^{\theta_2 - 2}(\theta_2 \log(x/\theta_3)(x^{\theta_2} - \theta_3^{\theta_2}) - (\theta_3^{\theta_2} + x^{\theta_2}))(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad + \theta_1 x^{\theta_2} \theta_3^{\theta_2 - 1}(\theta_2(-\theta_3^{-1})(x^{\theta_2} - \theta_3^{\theta_2}) - \theta_2 \log(x/\theta_3)\theta_2 \theta_3^{\theta_2 - 1} - \theta_2 \theta_3^{\theta_2 - 1})(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad - 3\theta_1 x^{\theta_2} \theta_3^{\theta_2 - 1}(\theta_2 \log(x/\theta_3)(x^{\theta_2} - \theta_3^{\theta_2}) - (\theta_3^{\theta_2} + x^{\theta_2}))(\theta_3^{\theta_2} + x^{\theta_2})^{-4} \theta_2 \theta_3^{\theta_2 - 1}
\end{aligned}$$

For $j, l = 3$,

$$g(x, \boldsymbol{\theta}) \equiv \frac{\partial^2}{\partial\theta_3^2} f(x, \boldsymbol{\theta}) = \frac{\theta_1 x^{\theta_2} \theta_2 \theta_3^{\theta_2 - 2}((\theta_2 + 1)\theta_3^{\theta_2} - (\theta_2 - 1)x^{\theta_2})}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\frac{\partial g}{\partial\theta_0} = 0 \qquad \frac{\partial g}{\partial\theta_1} = \frac{x^{\theta_2} \theta_2 \theta_3^{\theta_2 - 2}((\theta_2 + 1)\theta_3^{\theta_2} - (\theta_2 - 1)x^{\theta_2})}{(\theta_3^{\theta_2} + x^{\theta_2})^3}$$

$$\begin{aligned}
\frac{\partial g}{\partial\theta_2} &= \theta_1 \theta_3^{-2} x^{\theta_2} \theta_3^{\theta_2}(\log x + \log\theta_3)\theta_2((\theta_2 + 1)\theta_3^{\theta_2} - (\theta_2 - 1)x^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-3} + \theta_1 x^{\theta_2} \theta_3^{\theta_2 - 2} \\
&\quad \times ((2\theta_2 + 1)\theta_3^{\theta_2} + (\theta_2^2 + \theta_2)\theta_3^{\theta_2} \log\theta_3 - (2\theta_2 - 1)x^{\theta_2} - (\theta_2^2 - \theta_2)x^{\theta_2} \log x)(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad - 3\theta_1 x^{\theta_2} \theta_2 \theta_3^{\theta_2 - 2}((\theta_2 + 1)\theta_3^{\theta_2} - (\theta_2 - 1)x^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-4}(\theta_3^{\theta_2} \log\theta_3 + x^{\theta_2} \log x) \\
\frac{\partial g}{\partial\theta_3} &= \theta_1 x^{\theta_2} \theta_2((\theta_2 + 1)(2\theta_2 - 2)\theta_3^{2\theta_2 - 3} - (\theta_2 - 1)(\theta_2 - 2)x^{\theta_2} \theta_3^{\theta_2 - 3})(\theta_3^{\theta_2} + x^{\theta_2})^{-3} \\
&\quad - 3\theta_1 x^{\theta_2} \theta_2 \theta_3^{\theta_2 - 2}((\theta_2 + 1)\theta_3^{\theta_2} - (\theta_2 - 1)x^{\theta_2})(\theta_3^{\theta_2} + x^{\theta_2})^{-4} \theta_2 \theta_3^{\theta_2 - 1}
\end{aligned}$$

[C]: (i) For $k_1(x,\boldsymbol{\theta}) = \dfrac{1}{w(f(x,\boldsymbol{\theta}))}$ where $w(z) = |z|$, since we have

$$\frac{\partial}{\partial\boldsymbol{\theta}} k_1(x,\boldsymbol{\theta}) = -\frac{w'(f(x,\boldsymbol{\theta}))}{w^2(f(x,\boldsymbol{\theta}))}\frac{\partial}{\partial\boldsymbol{\theta}} f(x,\boldsymbol{\theta})$$

where $w'(z) = \text{sign}(z)$, the assumption

$$\lim_{\delta\to 0}\sup_{\|\boldsymbol{\Delta}\|\le\delta}\left|k_1(\mathbf{x},\boldsymbol{\theta}+\boldsymbol{\Delta}) - k_1(\mathbf{x},\boldsymbol{\theta})\right| = 0, \text{ uniformly in } \mathbf{x}$$

is satisfied.

(ii) Similarly, for $k_2(x,\boldsymbol{\theta}) = \dfrac{w'(f(x,\boldsymbol{\theta}))}{w^2(f(x,\boldsymbol{\theta}))}$ we have

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}} k_2(x,\boldsymbol{\theta}) &= \left\{w''(f(x,\boldsymbol{\theta}))w^{-2}(f(x,\boldsymbol{\theta})) - 2w'(f(x,\boldsymbol{\theta}))w^{-3}(f(x,\boldsymbol{\theta}))\right\}\frac{\partial}{\partial\boldsymbol{\theta}} f(x,\boldsymbol{\theta}) \\ &= -2w'(f(x,\boldsymbol{\theta}))w^{-3}(f(x,\boldsymbol{\theta}))\frac{\partial}{\partial\boldsymbol{\theta}} f(x,\boldsymbol{\theta}).\end{aligned}$$

where $w''(z) = 0$ if $z \ne 0$, not defined if $z = 0$. Thus, the assumption

$$\lim_{\delta\to 0}\sup_{\|\boldsymbol{\Delta}\|\le\delta}\left|k_2(\mathbf{x},\boldsymbol{\theta}+\boldsymbol{\Delta}) - k_2(\mathbf{x},\boldsymbol{\theta})\right| = 0, \text{ uniformly in } \mathbf{x}$$

is satisfied.

2.4.2 h : Huber function and w : $\mathbf{L_1}$-norm

In this case, we have

$$h(z) = \begin{cases} \dfrac{1}{\sqrt{2}} z & |z| \le k_0 \\ \left[k_0\left(|z| - \dfrac{1}{2}k_0\right)\right]^{\frac{1}{2}} & |z| > k_0 \end{cases}$$

Then, since $\psi(z) = (\partial/\partial z)h^2(z)$,

$$\psi(z) = \begin{cases} z & |z| \le k_0 \\ \\ k_0 \text{sign}(z) & |z| > k_0 \end{cases}$$

[A1]: Clearly, ψ is absolutely continuous.

[A2]: $\epsilon = Y - f(x, \boldsymbol{\theta}) \sim N(0, \sigma_x^2)$

(i) $E\psi^2(\epsilon) = E\{\epsilon^2 I(|\epsilon| \le k_0)\} + k_0^2 P(\epsilon > k_0) + (-k_0)^2 P(\epsilon < -k_0)$

$= E\{\epsilon^2 I(|\epsilon| \le k_0)\} + 2k_0^2 P(\epsilon > k_0) \equiv \sigma_\psi^2 u(x) < \infty$ where $\sigma_\psi^2 = 1$.

$E\psi(\epsilon) = E\{\epsilon I(|\epsilon| \le k_0)\} + k_0 P(\epsilon > k_0) - k_0 P(\epsilon < -k_0) = 0$

(ii) $\psi'(z) = 1$ if $|z| < k_0$, 0 if $|z| > k_0$, and not defined if $|z| = k_0$

$E|\psi'(\epsilon)|^{1+\delta} = P(|\epsilon| < k_0) < \infty$

$E\psi'(\epsilon) = P(|\epsilon| < k_0) = \gamma (\neq 0)$

[A3]:

(i) $\lim_{\delta \to 0} E\Big\{ \sup_{||\boldsymbol{\Delta}|| \le \delta} \big|\psi(Y - f(x, \boldsymbol{\theta} + \boldsymbol{\Delta})) - \psi(Y - f(x, \boldsymbol{\theta}))\big| \Big\}$

$= \lim_{\delta \to 0} E\Big[\sup_{||\boldsymbol{\Delta}|| \le \delta} \big|(Y - f(x, \boldsymbol{\theta} + \boldsymbol{\Delta})) - (Y - f(x, \boldsymbol{\theta}))\big| I\big(|Y - f(x, \boldsymbol{\theta})| \le k_0\big)$

$+ \sup_{||\boldsymbol{\Delta}|| \le \delta} |k_0 - k_0| I\big(Y - f(x, \boldsymbol{\theta}) > k_0\big)$

$+ \sup_{||\boldsymbol{\Delta}|| \le \delta} |-k_0 - (-k_0)| I\big(Y - f(x, \boldsymbol{\theta}) < -k_0\big)\Big] = 0$

(ii) $\lim_{\delta \to 0} E\Big\{ \sup_{||\boldsymbol{\Delta}|| \le \delta} |1 - 1| I\big(|Y - f(x, \boldsymbol{\theta})| \le k_0\big)\Big\} = 0$

The verification of the assumptions [**B1**]-[**B3**] and [**C**] is the same as in the previous example.

Chapter 3

M-METHODS IN HETEROSCEDASTIC NONLINEAR MODELS – II

3.1 Introduction

In many applications it is reasonable to assume that the log-variance of the random error of a nonlinear model is linear in dose. In this chapter we derive the M-estimator of the regression parameter in a nonlinear model under this special heteroscedasticity structure. In addition to being reasonable in practice this assumption keeps the number of unknown parameters to a small number. In this chapter we obtain both the ordinary M-estimator (OME) as well the weighted M-estimator (WME).

In Section 3.2 we define the M-estimator of both the regression parameters and the parameters for variance. We also present the notation and regularity conditions. In Section 3.3 we develop the asymptotic properties of the M-estimators. We introduce first the uniform asymptotic linearity on M-statistics.

The following nonlinear regression model will be studied in this chapter:

$$y_{ij} = f(x_i, \boldsymbol{\theta}) + \epsilon_{ij},\ \epsilon_{ij} \sim N(0, \sigma_i^2),\ i = 1, \ldots, k,\ j = 1, \ldots, n_i,\ \textstyle\sum_{i=1}^k n_i = n, \quad (3.1)$$

where ϵ_{ij} are assumed to be independent and have the parametric variance, that is $\log \sigma_i = \mathbf{z}_i^t \boldsymbol{\tau}$, y_{ij} are the observable r.v.s, x_i are known regression constants, $\boldsymbol{\theta}$ is a p-vector of unknown regression parameters, $f(,)$ is a nonlinear function of $\boldsymbol{\theta}$ of specified form, $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq})^t$ are known vectors, possibly dependent on the x_i, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)^t$ is a vector of unknown parameters.

3.2 Definitions and Regularity Conditions

The M-estimator of $(\boldsymbol{\theta}^t, \boldsymbol{\tau}^t)^t$ is obtained by the following minimization problem:

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}_n \\ \hat{\boldsymbol{\tau}}_n \end{pmatrix} = Argmin \left\{ \sum_{i,j} \frac{1}{w(e^{2\mathbf{z}_i^t \boldsymbol{\tau}})} h^2\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1 \subseteq \Re^p, \boldsymbol{\tau} \in \boldsymbol{\Theta}_2 \subseteq \Re^q \right\} \quad (3.2)$$

If $\mathbf{z}_i = 1$ and $\boldsymbol{\tau} = \tau_0$ $(q = 1)$, $i = 1, \ldots, k$, then we get the OME of $\boldsymbol{\theta}$, while we have the WME of $\boldsymbol{\theta}$ by letting $\mathbf{z}_i = (1, x_i)^t$ and $\boldsymbol{\tau} = (\tau_0, \tau_1)^t$ $(q = 2)$, $i = 1, \ldots, k$. If we let $\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta}) f(x_i, \boldsymbol{\theta})$, and $\psi(z) = (\partial/\partial z) h^2(z)$, then the estimating equation for the minimization in (3.2) is given by:

$$\sum_{i,j} \lambda(x_i, y_{ij}, \hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\tau}}_n) = \mathbf{0} \quad (3.3)$$

where

$$\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) = \begin{pmatrix} w_1(\mathbf{z}_i, \boldsymbol{\tau}) \psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \\ w_2(\mathbf{z}_i, \boldsymbol{\tau}) h^2\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) \mathbf{z}_i \end{pmatrix}, \quad (3.4)$$

$$w_1(\mathbf{z}_i, \boldsymbol{\tau}) = \frac{1}{w(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})}, \quad \text{and} \quad w_2(\mathbf{z}_i, \boldsymbol{\tau}) = \frac{2w'(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})e^{2\mathbf{z}_i^t\boldsymbol{\tau}}}{w^2(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})}.$$

We make the following sets of regularity assumptions concenrning (A) the function h^2; the score function ψ, (B) the function $f(\cdot)$, and (C) the functions generated from $w(,)$.

[**A1**]:

ψ is a nonconstant, odd function which is absolutely continuous and differentiable with respect to $\boldsymbol{\theta}$.

[**A2**]: For $i = 1, \ldots, k$,

(i) $E\psi^2(\epsilon_{i1}) = \sigma^2_{\psi i} < \infty$, $Eh^2(\epsilon_{i1}) = \gamma_{3i}\,(\neq 0)$, and $\text{Var}(h^2(\epsilon_{i1})) = \sigma^2_{hi} < \infty$

(ii) $E|\psi'(\epsilon_{i1})|^{1+\delta} < \infty$ for some $0 < \delta \leq 1$, and $E\psi'(\epsilon_{i1}) = \gamma_{1i}\,(\neq 0)$

[**A3**]: Let $\epsilon(\boldsymbol{\theta}) = \big(y - f(\mathbf{x}, \boldsymbol{\theta})\big)$

(i) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\| \,\leq\, \delta} \left|\psi(\epsilon(\boldsymbol{\theta}+\boldsymbol{\Delta})) - \psi(\epsilon(\boldsymbol{\theta}))\right|\right\} = 0$

(ii) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\| \,\leq\, \delta} \left|\psi'(\epsilon(\boldsymbol{\theta}+\boldsymbol{\Delta})) - \psi'(\epsilon(\boldsymbol{\theta}))\right|\right\} = 0$

(iii) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\| \,\leq\, \delta} \left|h^2(\epsilon(\boldsymbol{\theta}+\boldsymbol{\Delta})) - h^2(\epsilon(\boldsymbol{\theta}))\right|\right\} = 0$

[**B1**]:

$f(\mathbf{x}, \boldsymbol{\theta})$ is continuous and twice differentiable with respect to $\boldsymbol{\theta} \in \Theta$, where Θ is a compact subset of $\Re^p$.

[**B2**]:

(i) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \boldsymbol{\Gamma}_1(\boldsymbol{\theta}, \boldsymbol{\tau})$, where

$$\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \sum_{i=1}^{k} n_i \gamma_{1i} w_1(\mathbf{z}_i, \boldsymbol{\tau}) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}^t_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}).$$

(ii) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \boldsymbol{\Gamma}_2(\boldsymbol{\theta}, \boldsymbol{\tau})$, where

$$\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \sum_{i=1}^{k} n_i \gamma_{3i} w_3(\mathbf{z}_i, \boldsymbol{\tau}) \mathbf{z}_i \mathbf{z}_i^t,$$

and

$$w_3(\mathbf{z}_i, \boldsymbol{\tau}) = -\frac{4[\{w''(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})e^{2\mathbf{z}_i^t\boldsymbol{\tau}} + w'(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})\}w(e^{2\mathbf{z}_i^t\boldsymbol{\tau}}) - 2\{w'(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})\}^2]e^{2\mathbf{z}_i^t\boldsymbol{\tau}}}{w^3(e^{2\mathbf{z}_i^t\boldsymbol{\tau}})}.$$

(iii) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{31n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \boldsymbol{\Gamma}_{31}(\boldsymbol{\theta}, \boldsymbol{\tau})$, where

$$\boldsymbol{\Gamma}_{31n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \sum_{i=1}^{k} n_i \sigma_{\psi i}^2 w_1^2(\mathbf{z}_i, \boldsymbol{\tau})\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})\mathbf{f}_{\boldsymbol{\theta}}^t(x_i, \boldsymbol{\theta}),$$

and $\boldsymbol{\Gamma}_{31}(\boldsymbol{\theta}, \boldsymbol{\tau})$ a postive definite matrix.

(iv) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{32n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \boldsymbol{\Gamma}_{32}(\boldsymbol{\theta}, \boldsymbol{\tau})$, where

$$\boldsymbol{\Gamma}_{32n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \sum_{i=1}^{k} n_i \sigma_{hi}^2 w_2^2(\mathbf{z}_i, \boldsymbol{\tau})\mathbf{z}_i\mathbf{z}_i^t,$$

and $\boldsymbol{\Gamma}_{32}(\boldsymbol{\theta}, \boldsymbol{\tau})$ a postive definite matrix.

(v) $\max_i \left\{ n_i \sigma_{\psi i}^2 w_1^2(\mathbf{z}_i, \boldsymbol{\tau})\mathbf{f}_{\boldsymbol{\theta}}^t(\mathbf{x}_i, \boldsymbol{\theta})(\boldsymbol{\Gamma}_{31n}(\boldsymbol{\theta}, \boldsymbol{\tau}))^{-1}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta}) \right\} \longrightarrow 0$, as $n \to \infty$.

(vi) $\max_i \{ n_i \sigma_{hi}^2 w_2^2(\mathbf{z}_i, \boldsymbol{\tau})\mathbf{z}_i^t(\boldsymbol{\Gamma}_{32n}(\boldsymbol{\theta}, \boldsymbol{\tau}))^{-1}\mathbf{z}_i \} \longrightarrow 0$, as $n \to \infty$.

[**B3**]:

(i) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big| (\partial/\partial\theta_j) f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})(\partial/\partial\theta_k) f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})$

$$- (\partial/\partial\theta_j) f(\mathbf{x}, \boldsymbol{\theta})(\partial/\partial\theta_k) f(\mathbf{x}, \boldsymbol{\theta}) \Big| = 0 \text{ for } j, l = 1, \ldots, p$$

(ii) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big| (\partial^2/\partial\theta_j\partial\theta_k) f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta}) - (\partial^2/\partial\theta_j\partial\theta_k) f(\mathbf{x}, \boldsymbol{\theta}) \Big| = 0$ for $j, l = 1, \ldots, p$

[**C**]:

(i) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \left| w_1(\mathbf{z}, \boldsymbol{\tau} + \boldsymbol{\Delta}) - w_1(\mathbf{z}, \boldsymbol{\tau}) \right| = 0$, uniformly in $\mathbf{z}$.

(ii) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \left| w_2(\mathbf{z}, \boldsymbol{\tau} + \boldsymbol{\Delta}) - w_2(\mathbf{z}, \boldsymbol{\tau}) \right| = 0$, uniformly in $\mathbf{z}$.

(iii) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \left| w_3(\mathbf{z}, \boldsymbol{\tau} + \boldsymbol{\Delta}) - w_3(\mathbf{z}, \boldsymbol{\tau}) \right| = 0$, uniformly in $\mathbf{z}$.

3.3 Asymptotic Results

3.3.1 The Lemmas

The following lemmas are needed for proving the main results discussed in this chapter.

Lemma 3.1. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *for* $l = 1, \ldots, p+q$. *Then for* $l = 1, \ldots, p$

$$\sup_{\|t\|\le C_1, \|s\|\le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r \left\{ (\partial/\partial\theta_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right| = o_p(1). \tag{3.5}$$

where

$$\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) = w_1(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_l}(x_i, \boldsymbol{\theta}), \quad l = 1, \ldots, p, \tag{3.6}$$

and $f_{\theta_l}(x_i, \boldsymbol{\theta}) = (\partial/\partial\theta_l) f(x_i, \boldsymbol{\theta})$.

Proof. By the definition of derivative, we may write for $r, l = 1, \ldots, p$

$$\begin{aligned}
&(\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \\
&= w_1(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) (\partial^2/\partial\theta_l\partial\theta_r) f(x_i, \boldsymbol{\theta}) \\
&\qquad - w_1(\mathbf{z}_i, \boldsymbol{\tau})\psi'\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta})
\end{aligned} \tag{3.7}$$

Then,

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r\left\{(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.$$
$$\left.\left. -(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|$$
$$\le \frac{1}{n}C_1\sum_{i,j}\sum_{r=1}^{p}\sup_{\|t\|\le C1,\|s\|\le C_2}\left|(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.$$
$$\left. -(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|$$

and

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)-(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|$$
$$\le \sup_{\|t\|\le C_1,\|s\|\le C_2}\left|w_1\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\psi\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)(\partial^2/\partial\theta_l\partial\theta_r)f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\right.$$
$$\left. -w_1(\mathbf{z}_i,\boldsymbol{\tau})\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)(\partial^2/\partial\theta_l\partial\theta_r)f(x_i,\boldsymbol{\theta})\right|$$
$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|w_1\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\psi'\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_r}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\right.$$
$$\left. -w_1(\mathbf{z}_i,\boldsymbol{\tau})\psi'\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)f_{\theta_l}(x_i,\boldsymbol{\theta})f_{\theta_r}(x_i,\boldsymbol{\theta})\right|$$
$$\le \sup_{\|t\|\le C_1,\|s\|\le C_2}\left\{\left|\psi\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)-\psi\big(y_{ij}-f(x_i,\boldsymbol{\theta})\big)\right|\right.$$
$$\left.\times\left|w_1\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)(\partial^2/\partial\theta_l\partial\theta_r)f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\right|\right\}$$
$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left\{\left|w_1\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)-w_1(\mathbf{z}_i,\boldsymbol{\tau})\right|\right.$$
$$\left.\times\left|(\partial^2/\partial\theta_l\partial\theta_r)f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\right|\left|\psi\big(y_{ij}-f(x_i,\boldsymbol{\theta})\big)\right|\right\}$$

$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left\{ \left| (\partial^2/\partial\theta_l\partial\theta_r) f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) - (\partial^2/\partial\theta_l\partial\theta_r) f(x_i, \boldsymbol{\theta}) \right| \right.$$

$$\left. \times \Big| w_1(\mathbf{z}_i, \boldsymbol{\tau}) \Big| \Big| \psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) \Big| \right\}$$

$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left\{ \left| \psi'\Big(y_{ij} - f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi'\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big) \right| \right.$$

$$\left. \times \left| w_1\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) f_{\theta_l}\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_r}\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) \right| \right\}$$

$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left\{ \left| w_1\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - w_1(\mathbf{z}_i, \boldsymbol{\tau}) \right| \right.$$

$$\left. \times \left| f_{\theta_l}\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_r}\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) \right| \Big| \psi'(y_{ij} - f(x_i, \boldsymbol{\theta})) \Big| \right\}$$

$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left\{ \left| f_{\theta_l}\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) f_{\theta_r}\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) - f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta}) \right| \right.$$

$$\left. \times \Big| w_1(\mathbf{z}_i, \boldsymbol{\tau}) \Big| \Big| \psi'\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) \Big| \right\}.$$

Then by taking expectations on both sides we get

$$E\left\{ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| (\partial/\partial\theta_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right| \right\}$$

$$\le E\left\{ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \psi\Big(y_{ij} - f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big) - \psi\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big) \right| \right\}$$

$$\times \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| w_1\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) (\partial^2/\partial\theta_l\partial\theta_r) f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) \right|$$

$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| w_1\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - w_1(\mathbf{z}_i, \boldsymbol{\tau}) \right|$$

$$\times \left| (\partial^2/\partial\theta_l\partial\theta_r) f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big) \right| E\left\{ \Big| \psi\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big) \Big| \right\}$$

$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial^2/\partial\theta_l\partial\theta_r)f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)-(\partial^2/\partial\theta_l\partial\theta_r)f(x_i,\boldsymbol{\theta})\right|$$

$$\times\Big|w_1(\mathbf{z}_i,\boldsymbol{\tau})\Big|\Big|E\Big\{\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)\Big|\Big\}$$

$$+E\left\{\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\psi'\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)-\psi'\big(y_{ij}-f(x_i,\boldsymbol{\theta})\big)\right|\right\}$$

$$\times\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|w_1\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_r}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\right|$$

$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|w_1\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)-w_1(\mathbf{z}_i,\boldsymbol{\tau})\right|$$

$$\times\left|f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_r}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\right|E\Big\{\Big|\psi'(y_{ij}-f(x_i,\boldsymbol{\theta}))\Big|\Big\}$$

$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)f_{\theta_r}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)-f_{\theta_l}(x_i,\boldsymbol{\theta})f_{\theta_r}(x_i,\boldsymbol{\theta})\right|$$

$$\times\Big|w_1(\mathbf{z}_i,\boldsymbol{\tau})\Big|\Big|E\Big\{\psi'\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)\Big|\Big\}.$$

Thus, by conditions **[A3]** (i)-(iv), **[B3]** (i)-(ii), and **[C]** (i)-(ii), we have that:

$$E\left\{\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)-(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\}$$
$$\longrightarrow 0,\quad \forall i,j$$

and

$$E\left[\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p}t_r\Big\{(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.$$
$$\left.\left.-(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\Big\}\right|\right]\longrightarrow 0.$$

Also,

$$
\begin{aligned}
&\operatorname{Var}\left[\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r\left\{(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.\left.\left.-(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|\right]\\
&\le \frac{C_1^2}{n^2}\sum_{i,j}\operatorname{Var}\left\{\sum_{r=1}^{p}\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.\left.-(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\}\\
&\le C_1^2K_1/n \longrightarrow 0.
\end{aligned}
$$

Therefore, we have (3.5). □

Lemma 3.2. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$ *for* $l=1,\ldots,p+q$. *Then for* $l=1,\ldots,p$

$$
\begin{aligned}
&\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right.\\
&\qquad\qquad\left.+\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r\gamma_{1i}w_1(\mathbf{z}_i,\boldsymbol{\tau})f_{\theta_l}(x_i,\boldsymbol{\theta})f_{\theta_r}(x_i,\boldsymbol{\theta})\right| = o_p(1),
\end{aligned}
\tag{3.8}
$$

where $\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$ *is defined in (3.6).*

Proof. From (3.7), we have

$$
\begin{aligned}
&\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})+\frac{\gamma_1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r\frac{f_{\theta_l}(x_i,\boldsymbol{\theta})f_{\theta_r}(x_i,\boldsymbol{\theta})}{(\mathbf{z}_i^t\boldsymbol{\tau})^2}\right|\\
&=\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r w_1(\mathbf{z}_i,\boldsymbol{\tau})\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)(\partial^2/\partial\theta_l\partial\theta_r)f(x_i,\boldsymbol{\theta})\right.\\
&\qquad\left.-\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r w_1(\mathbf{z}_i,\boldsymbol{\tau})\left\{\psi'\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)-\gamma_{1i}\right\}f_{\theta_l}(x_i,\boldsymbol{\theta})f_{\theta_r}(x_i,\boldsymbol{\theta})\right|
\end{aligned}
$$

$$\leq C_1 \sum_{r=1}^{p} \left| \frac{1}{n} \sum_{i,j} w_1(\mathbf{z}_i, \boldsymbol{\tau}) \psi \left(y_{ij} - f(x_i, \boldsymbol{\theta}) \right) (\partial^2/\partial\theta_l \partial\theta_r) f(x_i, \boldsymbol{\theta}) \right|$$
$$+ C_1 \sum_{r=1}^{p} \left| \frac{1}{n} \sum_{i,j} w_1(\mathbf{z}_i, \boldsymbol{\tau}) \left\{ \psi' \left(y_{ij} - f(x_i, \boldsymbol{\theta}) \right) - \gamma_{1i} \right\} f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta}) \right|$$

which by using the Markov WLLN and conditions [**A2**] (i)-(ii) yields

$$\frac{1}{n} \sum_{i,j} w_1(\mathbf{z}_i, \boldsymbol{\tau}) \psi \left(y_{ij} - f(x_i, \boldsymbol{\theta}) \right) (\partial^2/\partial\theta_l \partial\theta_r) f(x_i, \boldsymbol{\theta}) = o_p(1)$$

and

$$\frac{1}{n} \sum_{i,j} w_1(\mathbf{z}_i, \boldsymbol{\tau}) \left\{ \psi' \left(y_{ij} - f(x_i, \boldsymbol{\theta}) \right) - \gamma_{1i} \right\} f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta}) = o_p(1).$$

Therefore, we have (3.8). □

Lemma 3.3. *Let the conditions* [**A1**]-[**A3**], [**B1**]-[**B3**], *and* [**C**] *hold and let* $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *for* $l = 1, \ldots, p+q$. *Then for* $l = 1, \ldots, p$

$$\sup_{\|t\| \leq C_1, \|s\| \leq C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \left\{ (\partial/\partial\tau_r) \lambda_l \Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}} \Big) - (\partial/\partial\tau_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right| = o_p(1). \tag{3.9}$$

where $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *is defined in (3.6).*

Proof. By the definition of derivative, we may write for $l = 1, \ldots, p$, $r = 1, \ldots, q$,

$$(\partial/\partial\tau_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) = -w_2(\mathbf{z}_i, \boldsymbol{\tau}) \psi(y_{ij} - f(x_i, \boldsymbol{\tau})) f_{\theta_l}(x_i, \boldsymbol{\theta}) z_{ir} \tag{3.10}$$

Then,

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q}s_r\left\{(\partial/\partial\tau_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.$$
$$\left.\left.-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|$$
$$\le\frac{1}{n}C_2\sum_{i,j}\sum_{r=1}^{q}\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right)\right.$$
$$\left.-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|$$

and

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right)-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|$$
$$\le\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|w_2\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\psi\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)z_{ir}\right.$$
$$\left.-w_2(\mathbf{z}_i,\boldsymbol{\tau})\psi\big(y_{ij}-f(x_i,\boldsymbol{\theta})\big)f_{\theta_l}(x_i,\boldsymbol{\theta})z_{ir}\right|$$
$$\le\sup_{\|t\|\le C_1,\|s\|\le C_2}\left\{\left|\psi\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)-\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)\right|\right.$$
$$\left.\times\left|w_2\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)z_{ir}\right|\right\}$$
$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left\{\left|w_2\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)-w_2(\mathbf{z}_i,\boldsymbol{\tau})\right|\right.$$
$$\left.\times\left|f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)z_{ir}\right|\left|\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))\right|\right\}$$
$$+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left\{\left|f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)-f_{\theta_l}(x_i,\boldsymbol{\theta})\right|\left|w_2(\mathbf{z}_i,\boldsymbol{\tau})\right|\left|\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)\right|\right\}$$

Then by taking expectations on both sides we get

$$
\begin{aligned}
&E\left\{\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right)-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\}\\
&\le E\left\{\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\psi\Big(y_{ij}-f\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big)-\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)\right|\right\}\\
&\qquad\qquad\times\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|w_2\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)z_{ir}\right|\\
&+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left\{\left|w_2\Big(\mathbf{z}_i,\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)-w_2(\mathbf{z}_i,\boldsymbol{\tau})\right|\right.\\
&\qquad\qquad\left.\times\left|f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)z_{ir}\right|\right\}E\left\{\left|\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))\right|\right\}\\
&+\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|f_{\theta_l}\Big(x_i,\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}}\Big)-f_{\theta_l}(x_i,\boldsymbol{\theta})\right|\left|w_2(\mathbf{z}_i,\boldsymbol{\tau})\right|E\left\{\left|\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)\right|\right\}
\end{aligned}
$$

Then, by conditions **[A3]** (i)-(iv), **[B3]** (i)-(ii), and **[C]** (i)-(ii), we have that

$$
\begin{aligned}
&E\left\{\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right)-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\}\\
&\hfill\longrightarrow 0,\ \forall i,j
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q}s_r\left\{(\partial/\partial\tau_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.\right.&\\
\left.\left.\left.-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|\right]&\longrightarrow 0.
\end{aligned}
$$

Also,

$$\begin{aligned}
&\operatorname{Var}\left[\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q} s_r\left\{(\partial/\partial\tau_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big)\right.\right.\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.\left.\left.-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|\right]\\
&\le \frac{C_2^2}{n^2}\sum_{i,j}\operatorname{Var}\left\{\sum_{r=1}^{q}\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right)\right.\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.\left.-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\}\\
&\le C_2^2K_2/n \longrightarrow 0.
\end{aligned}$$

Therefore, we have (3.9). □

Lemma 3.4. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *for* $l = 1, \ldots, p+q$. *Then for* $l = 1, \ldots, p$

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q} s_r(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right| = o_p(1). \tag{3.11}$$

where $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *is defined in (3.6).*

Proof. From (3.10), we have

$$\begin{aligned}
&\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q} s_r(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\\
=\;&\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q} s_r w_2(\mathbf{z}_i,\boldsymbol{\tau})\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))f_{\theta_l}(x_i,\boldsymbol{\theta})z_{ir}\right|\\
=\;& C_2\sum_{r=1}^{q}\left|\frac{1}{n}\sum_{i,j} w_2(\mathbf{z}_i,\boldsymbol{\tau})\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))f_{\theta_l}(x_i,\boldsymbol{\theta})z_{ir}\right|
\end{aligned}$$

which by using the Markov WLLN and conditions **[A2]** (i)-(ii) yields

$$\frac{1}{n}\sum_{i,j} w_2(\mathbf{z}_i,\boldsymbol{\tau})\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))f_{\theta_l}(x_i,\boldsymbol{\theta})z_{ir}=o_p(1).$$

Therefore, we have (3.11). □

Lemma 3.5. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$ *for* $l=1,\ldots,p+q$. *Then for* $l=p+1,\ldots,p+q$

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r\left\{(\partial/\partial\theta_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|=o_p(1). \tag{3.12}$$

where

$$\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})=w_2(\mathbf{z}_i,\boldsymbol{\tau})h^2(y_{ij}-f(x_i,\boldsymbol{\theta}))z_{i,(l-p)},\quad l=p+1,\ldots,p+q. \tag{3.13}$$

Proof. By the definition of derivative, we may write for $r=1,\ldots,p$, $l=p+1,\ldots,p+q$,

$$(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})=-w_2(\mathbf{z}_i,\boldsymbol{\tau})\psi\left(y_{ij}-f(x_i,\boldsymbol{\theta})\right)f_{\theta_r}(x_i,\boldsymbol{\theta})z_{i,(l-p)} \tag{3.14}$$

Then since (3.14) is the same as (3.10), we can follow the proof of Lemma 3.3 to obtain the result of (3.12). □

Lemma 3.6. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$ *for* $l=1,\ldots,p+q$. *Then for* $l=p+1,\ldots,p+q$

$$\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r(\partial/\partial\theta_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|=o_p(1). \tag{3.15}$$

where $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *is defined in (3.13).*

Proof. From (3.14), we have

$$\begin{aligned}
&\sup_{\|t\|\le C_1, \|s\|\le C_2} \left| \frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r(\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right| \\
= &\sup_{\|t\|\le C_1, \|s\|\le C_2} \left| \frac{1}{n}\sum_{i,j}\sum_{r=1}^{p} t_r w_2(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_r}(x_i, \boldsymbol{\theta}) z_{i,(l-p)} \right| \\
\le & \; C_1 \sum_{r=1}^{p} \left| \frac{1}{n}\sum_{i,j} w_2(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_r}(x_i, \boldsymbol{\theta}) z_{i,(l-p)} \right|
\end{aligned}$$

which by using the Markov WLLN and conditions **[A2]** (i)-(ii) yields

$$\frac{1}{n}\sum_{i,j} w_2(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_r}(x_i, \boldsymbol{\theta}) z_{i,(l-p)} = o_p(1)$$

Therefore, we have the result in (3.15). □

Lemma 3.7. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold and let* $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *for* $l = 1, \ldots, p+q$. *Then for* $l = p+1, \ldots, p+q$

$$\begin{aligned}
\sup_{\|t\|\le C_1, \|s\|\le C_2} \Bigg| \frac{1}{n}\sum_{i,j}\sum_{r=1}^{q} s_r \Bigg\{ &(\partial/\partial\tau_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) \\
&- (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \Bigg\} \Bigg| = o_p(1).
\end{aligned} \tag{3.16}$$

where $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *is defined in (3.13).*

Proof. By the definition of derivative, we may write for $l = p+1, \ldots, p+q$, $r = 1, \ldots, q$,

$$\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) = w_3(\mathbf{z}_i, \boldsymbol{\tau})h^2(y_{ij} - f(x_i, \boldsymbol{\theta}))z_{i,(l-p)}z_{ir}. \tag{3.17}$$

Then,

$$\sup_{\|t\|\le C_1, \|s\|\le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \left\{ (\partial/\partial\tau_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) \right.\right.$$
$$\left.\left. - (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right|$$
$$\le \frac{1}{n} C_2 \sum_{i,j} \sum_{r=1}^{q} \sup_{\|t\|\le C_1, \|s\|\le C_2} \left| (\partial/\partial\tau_r)\lambda_l \left(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}} \right) \right.$$
$$\left. - (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right|$$

and

$$\sup_{\|t\|\le C_1, \|s\|\le C_2} \left| (\partial/\partial\tau_r)\lambda_l \left(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}} \right) - (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right|$$
$$\le \sup_{\|t\|\le C_1, \|s\|\le C_2} \left\{ \left| h^2\Big(y_{ij} - f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big) - h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) \right| \right.$$
$$\left. \times \left| w_3\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) z_{i,(l-p)} z_{ir} \right| \right\}$$
$$+ \sup_{\|t\|\le C_1, \|s\|\le C_2} \left| w_3\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - w_3(\mathbf{z}_i, \boldsymbol{\tau}) \right| \left| z_{i,(l-p)} z_{ir} \right| \left| h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) \right|$$

Then by taking expectations on both sides we get

$$E\left\{ \sup_{\|t\|\le C_1, \|s\|\le C_2} \left| (\partial/\partial\tau_r)\lambda_l \left(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}} \right) - (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right| \right\}$$
$$\le E\left\{ \sup_{\|t\|\le C_1, \|s\|\le C_2} \left| h^2\Big(y_{ij} - f\Big(x_i, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}\Big)\Big) - h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) \right| \right\}$$
$$\times \sup_{\|t\|\le C_1, \|s\|\le C_2} \left| w_3\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) z_{i,(l-p)} z_{ir} \right|$$
$$+ \sup_{\|t\|\le C_1, \|s\|\le C_2} \left| w_3\Big(\mathbf{z}_i, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - w_3(\mathbf{z}_i, \boldsymbol{\tau}) \right| \left| z_{i,(l-p)} z_{ir} \right| E\left\{ \left| h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) \right| \right\}$$

Then, by conditions [**A3**] (i)-(iv), [**B3**] (i)-(ii), and [**C**] (i)-(ii), we have that

$$E\left\{\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right)-(\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\} \longrightarrow 0,\ \forall i,j$$

and

$$E\left[\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q}s_r\left\{(\partial/\partial\tau_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|\right]\longrightarrow 0.$$

Also,

$$\begin{aligned}
&\mathrm{Var}\left[\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|\frac{1}{n}\sum_{i,j}\sum_{r=1}^{q}s_r\left\{(\partial/\partial\tau_r)\lambda_l\Big(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right\}\right|\right]\\
&\le \frac{C_2^2}{n^2}\sum_{i,j}\mathrm{Var}\left\{\sum_{r=1}^{q}\sup_{\|t\|\le C_1,\|s\|\le C_2}\left|(\partial/\partial\tau_r)\lambda_l\left(x_i,y_{ij},\boldsymbol{\theta}+\frac{u\mathbf{t}}{\sqrt{n}},\boldsymbol{\tau}+\frac{v\mathbf{s}}{\sqrt{n}}\right) - (\partial/\partial\tau_r)\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})\right|\right\}\\
&\le C_2^2K_2/n\longrightarrow 0.
\end{aligned}$$

Thus we have (3.16). □

Lemma 3.8. *Let the conditions* [**A1**]-[**A3**], [**B1**]-[**B3**], *and* [**C**] *hold and let* $\lambda_l(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$, *defined in (3.4), be the lth element of the vector* $\lambda(x_i,y_{ij},\boldsymbol{\theta},\boldsymbol{\tau})$ *for*

$l = 1, \ldots, p+q$. *Then for* $l = p+1, \ldots, p+q$

$$\sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r (\partial/\partial \tau_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \gamma_{3i} w_3(\mathbf{z}_i, \boldsymbol{\tau}) z_{i,(l-p)} z_{ir} \right| = o_p(1). \tag{3.18}$$

where $\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ *is defined in (3.13).*

Proof. From (3.17), we have

$$\begin{aligned} &\sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r (\partial/\partial \tau_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \gamma_{3i} w_3(\mathbf{z}_i, \boldsymbol{\tau}) z_{i,(l-p)} z_{ir} \right| \\ &= \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r w_3(\mathbf{z}_i, \boldsymbol{\theta}) \left\{ h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) - \gamma_{3i} \right\} z_{i,(l-p)} z_{ir} \right| \\ &\le 2C_2 \sum_{r=1}^{q} \left| \frac{1}{n} \sum_{i,j} w_3(\mathbf{z}_i, \boldsymbol{\theta}) \left\{ h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) - \gamma_{3i} \right\} z_{i,(l-p)} z_{ir} \right| \end{aligned}$$

which by using the Markov WLLN and conditions [**A2**] (i)-(ii) yields

$$\frac{1}{n} \sum_{i,j} w_3(\mathbf{z}_i, \boldsymbol{\theta}) \left\{ h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) - \gamma_{3i} \right\} z_{i,(l-p)} z_{ir} = o_p(1).$$

Therefore, we have the result in (3.18). □

3.3.2 The Main Results

In the following theorem we establish the asymptotic linearity of the M-statistics.

Theorem 3.9. *Let the conditions* [**A1**]-[**A3**], [**B1**]-[**B3**], *and* [**C**] *hold. Then*

$$\sup_{\|\mathbf{t}\| \le C_1, \|\mathbf{s}\| \le C_2} \left\| \frac{1}{\sqrt{n}} \sum_{i,j} \left\{ \lambda(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) - \lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} + \frac{1}{n} \left(\boldsymbol{\Gamma}_{1n}^t(\boldsymbol{\theta}, \boldsymbol{\tau}), \mathbf{0}^t \right)^t \mathbf{t} + \frac{1}{n} \left(\mathbf{0}^t, \boldsymbol{\Gamma}_{2n}^t(\boldsymbol{\theta}, \boldsymbol{\tau}) \right)^t \mathbf{s} \right\| = o_p(1) \tag{3.19}$$

as $n \to \infty$, where $\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ was defined in (3.4).

Proof. We consider the lth element of the vector $\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})$ denoted for

$$\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) = \begin{cases} w_1(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_l}(x_i, \boldsymbol{\theta}), & l = 1, \ldots, p \\ w_2(\mathbf{z}_i, \boldsymbol{\tau})h^2\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) z_{i,(l-p)}, & l = p+1, \ldots, p+q, \end{cases}$$

where $f_{\theta_l}(x_i, \boldsymbol{\theta}) = (\partial/\partial\theta_l)f(x_i, \boldsymbol{\theta})$. Using the first order term in the Taylor's expansion, we have for $0 < u, v < 1$,

$$\begin{aligned}
&\lambda_l(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) - \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \\
&= \frac{1}{\sqrt{n}} \sum_{r=1}^{p} t_r \big\{(\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})\big\} + \frac{1}{\sqrt{n}} \sum_{r=1}^{q} s_r \big\{(\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})\big\} \\
&+ \frac{1}{\sqrt{n}} \sum_{r=1}^{p} t_r \left\{(\partial/\partial\theta_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})\right\} \\
&+ \frac{1}{\sqrt{n}} \sum_{r=1}^{q} s_r \left\{(\partial/\partial\tau_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) - (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau})\right\},
\end{aligned}$$

where for $r = 1, \ldots, p$,

$$\begin{aligned}
&(\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \\
&= \begin{cases} w_1(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) (\partial^2/\partial\theta_l\partial\theta_r)f(x_i, \boldsymbol{\theta}) \\ \qquad -w_1(\mathbf{z}_i, \boldsymbol{\tau})\psi'\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_l}(x_i, \boldsymbol{\theta})f_{\theta_r}(x_i, \boldsymbol{\theta}), & l = 1, \ldots, p \\ -w_2(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_r}(x_i, \boldsymbol{\theta})z_{i,(l-p)} & l = p+1, \ldots, p+q, \end{cases}
\end{aligned}$$

and for $r = 1, \ldots, q$,

$$\begin{aligned}
&(\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \\
&= \begin{cases} -w_2(\mathbf{z}_i, \boldsymbol{\tau})\psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) f_{\theta_l}(x_i, \boldsymbol{\theta})z_{ir} & l = 1, \ldots, p \\ -w_3(\mathbf{z}_i, \boldsymbol{\tau})h^2\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) z_{i,(l-p)}z_{ir} & l = p+1, \ldots, p+q. \end{cases}
\end{aligned}$$

Then, we have

(i) for $l = 1, \dots, p$,

$$\sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{\sqrt{n}} \sum_{i,j} \left\{ \lambda_l(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) - \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right.$$
$$\left. + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r \gamma_{1i} w_1(\mathbf{x}_i, \boldsymbol{\tau}) f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta}) \right|$$

$$\le \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r \left\{ (\partial/\partial\theta_r) \lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) \right.\right.$$
$$\left.\left. - (\partial/\partial\theta_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right|$$
$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \left\{ (\partial/\partial\tau_r) \lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) \right.\right.$$
$$\left.\left. - (\partial/\partial\tau_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right|$$
$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r (\partial/\partial\theta_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right.$$
$$\left. + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r \gamma_{1i} w_1(\mathbf{x}_i, \boldsymbol{\tau}) f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta}) \right|$$
$$+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r (\partial/\partial\tau_r) \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right|.$$

Using Lemma 3.1 through 3.4 we deduce that for $l = 1, \dots, p$

$$\sup_{\|t\| \le C_1, \|s\| \le C_2} \left| \frac{1}{\sqrt{n}} \sum_{i,j} \left\{ \lambda_l(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) - \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \right.$$
$$\left. + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r \gamma_{1i} w_1(\mathbf{x}_i, \boldsymbol{\tau}) f_{\theta_l}(x_i, \boldsymbol{\theta}) f_{\theta_r}(x_i, \boldsymbol{\theta}) \right| = o_p(1). \tag{3.20}$$

(ii) for $l = p+1, \dots, p+q$,

$$\begin{aligned}
&\sup_{\|t\| \le C_1, \|s\| \le C_2} \Bigg| \frac{1}{\sqrt{n}} \sum_{i,j} \left\{ \lambda_l(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) - \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \\
&\qquad + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \gamma_{3i} w_3(\mathbf{z}_i, \boldsymbol{\tau}) z_{i,(l-p)} z_{ir} \Bigg| \\
&\le \sup_{\|t\| \le C_1, \|s\| \le C_2} \Bigg| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r \Big\{ (\partial/\partial\theta_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) \\
&\qquad\qquad - (\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \Big\} \Bigg| \\
&+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \Bigg| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \Big\{ (\partial/\partial\tau_r)\lambda_l\Big(x_i, y_{ij}, \boldsymbol{\theta} + \frac{u\mathbf{t}}{\sqrt{n}}, \boldsymbol{\tau} + \frac{v\mathbf{s}}{\sqrt{n}}\Big) \\
&\qquad\qquad - (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \Big\} \Bigg| \\
&+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \Bigg| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{p} t_r (\partial/\partial\theta_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \Bigg| \\
&+ \sup_{\|t\| \le C_1, \|s\| \le C_2} \Bigg| \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r (\partial/\partial\tau_r)\lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \gamma_{3i} w_3(\mathbf{z}_i, \boldsymbol{\tau}) z_{i,(l-p)} z_{ir} \Bigg| .
\end{aligned}$$

And, from Lemma 3.5 through 3.8 we conclude that for $l = p+1, \dots, p+q$

$$\begin{aligned}
&\sup_{\|t\| \le C_1, \|s\| \le C_2} \Bigg| \frac{1}{\sqrt{n}} \sum_{i,j} \left\{ \lambda_l(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) - \lambda_l(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) \right\} \\
&\qquad + \frac{1}{n} \sum_{i,j} \sum_{r=1}^{q} s_r \gamma_{3i} w_3(\mathbf{z}_i, \boldsymbol{\tau}) z_{i,(l-p)} z_{ir} \Bigg| = o_p(1).
\end{aligned} \tag{3.21}$$

Therefore, the result in (3.19) follows from both (3.20) and (3.21). □

We now consider the following theorem which states an existence of a solution of (3.3) that is a $\sqrt{n}$-consistent estimator of $(\boldsymbol{\theta}^t, \boldsymbol{\tau}^t)^t$ and admits an asymptotic representation.

Theorem 3.10. *Let the conditions* **[A1]**-**[A3]**, **[B1]**-**[B3]**, *and* **[C]** *hold. Then there* *exists a sequence* $(\hat{\boldsymbol{\theta}}_n^t, \hat{\boldsymbol{\tau}}_n^t)^t$ *of solutions of (3.3) such that:*

$$\sqrt{n}\left\|\begin{pmatrix} \hat{\boldsymbol{\theta}}_n \\ \hat{\boldsymbol{\tau}}_n \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\tau} \end{pmatrix}\right\| = O_p(1), \tag{3.22}$$

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}_n \\ \hat{\boldsymbol{\tau}}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\tau} \end{pmatrix} + \frac{1}{n}\left(\frac{1}{n}\boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta}, \boldsymbol{\tau})\right)^{-1} \sum_{i,j} \lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) + o_p(n^{-\frac{1}{2}}) \tag{3.23}$$

or

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \\ \hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau} \end{pmatrix} = \left(\frac{1}{n}\boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta}, \boldsymbol{\tau})\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i,j} \lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) + o_p(1), \tag{3.24}$$

where

$$\boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \begin{pmatrix} \boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}, \boldsymbol{\tau}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}, \boldsymbol{\tau}) \end{pmatrix}.$$

Proof. From Theorem 3.9 the following system of equations:

$$\sum_{i,j} \lambda_l(x_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}, \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}) = 0$$

has a root $(\mathbf{t}_n^t, \mathbf{s}_n^t)^t$ that lies in $\|\mathbf{t}\| \le C_1$, $\|\mathbf{s}\| \le C_2$ with probability exceeding $1 - \epsilon$ for $n \ge n_0$. Then $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}_n$, $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau} + n^{-\frac{1}{2}}\mathbf{s}_n$ is a solution of (3.3) satisfying:

$$P\left(\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \le C_1, \sqrt{n}\|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}\| \le C_2\right) \ge 1 - \epsilon \quad \text{for} \quad n \ge n_0.$$

Substituting $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ and $\sqrt{n}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau})$ into $\mathbf{t}$ and $\mathbf{s}$, respectively in (3.19), we have the expression in (3.23) and (3.24). □

In the following theorem and corollaries we prove the asymptotic normality of the M-estimator in (3.3).

Theorem 3.11. *Let the conditions* [**A1**], [**A2**], [**B1**], *and* [**B2**] *hold. Then*

$$\frac{1}{\sqrt{n}}\sum_{i,j}\left\{\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) - \mu(\mathbf{z}_i, \boldsymbol{\tau})\right\} \longrightarrow N_{p+q}\left(\mathbf{0}, \boldsymbol{\Gamma}_3(\boldsymbol{\theta}, \boldsymbol{\tau})\right) \quad as \quad n \to \infty. \tag{3.25}$$

where $\mu(\mathbf{z}_i, \boldsymbol{\tau}) = (\mathbf{0}^t, \gamma_{3i} w_2(\mathbf{z}_i, \boldsymbol{\tau})\mathbf{z}_i^t)^t$, *and*

$$\boldsymbol{\Gamma}_3(\boldsymbol{\theta}, \boldsymbol{\tau}) = \begin{pmatrix} \boldsymbol{\Gamma}_{31}(\boldsymbol{\theta}, \boldsymbol{\tau}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{32}(\boldsymbol{\theta}, \boldsymbol{\tau}) \end{pmatrix}.$$

Proof. Consider the following linear combination

$$Z_n^* = \boldsymbol{\eta}^t \frac{1}{\sqrt{n}} \sum_{i,j}\left\{\lambda(x_i, y_{ij}, \boldsymbol{\theta}, \boldsymbol{\tau}) - \mu(\mathbf{z}_i, \boldsymbol{\tau})\right\},$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^t, \boldsymbol{\eta}_2^t)^t$ and $\boldsymbol{\eta}_1 \in \Re^p$; $\boldsymbol{\eta}_2 \in \Re^q$. and have that:

$$\begin{aligned} Z_n^* &= \sum_{i,j} \frac{1}{\sqrt{n}} \Big[w_1(\mathbf{z}_i, \boldsymbol{\tau}) \boldsymbol{\eta}_1^t \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \psi\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) \\ &\qquad + w_2(\mathbf{z}_i, \boldsymbol{\tau}) \boldsymbol{\eta}_2^t \mathbf{z}_i \left\{ h^2\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right) - \gamma_{3i} \right\} \Big] \\ &= \sum_{i,j} \frac{1}{\sqrt{n}} (c_{i1} Z_{ij1} + c_{i2} Z_{ij2}), \end{aligned}$$

where $c_{i1} = w_1(\mathbf{z}_i, \boldsymbol{\tau})\boldsymbol{\eta}_1^t \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})$, $c_{i2} = w_2(\mathbf{z}_i, \boldsymbol{\tau})\boldsymbol{\eta}_2^t \mathbf{z}_i$, $Z_{ij1} = \psi(y_{ij} - f(x_i, \boldsymbol{\theta}))$, and $Z_{ij2} = h^2(y_{ij} - f(x_i, \boldsymbol{\theta})) - \gamma_{3i}$. Then,

$$EZ_{ij1} = EZ_{ij2} = 0,$$

and

$$EZ_{ij1}^2 = \sigma_{\psi i}^2; \quad EZ_{ij2}^2 = \sigma_{hi}^2; \quad EZ_{ij1}Z_{ij2} = 0.$$

Therefore, if we let

$$Z_n^* = \sum_{i,j} c_{ni} Z_{nij},$$

where

$$c_{ni} = \frac{1}{\sqrt{n}} \left\{ \boldsymbol{\eta}^t \boldsymbol{\Gamma}_4(x_i, \boldsymbol{\theta}, \boldsymbol{\tau}) \boldsymbol{\eta} \right\}^{\frac{1}{2}};$$

$$Z_{nij} = (c_{i1} Z_{ij1} + c_{i2} Z_{ij2}) / \left\{ \boldsymbol{\eta}^t \boldsymbol{\Gamma}_4(x_i, \boldsymbol{\theta}, \boldsymbol{\tau}) \boldsymbol{\eta} \right\}^{\frac{1}{2}},$$

and

$$\boldsymbol{\Gamma}_4(x_i, \boldsymbol{\theta}, \boldsymbol{\tau}) = \begin{pmatrix} \sigma_{\psi i}^2 w_1^2(\mathbf{z}_i, \boldsymbol{\tau}) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(x_i, \boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \sigma_{hi}^2 w_2^2(\mathbf{z}_i, \boldsymbol{\tau}) \mathbf{z}_i \mathbf{z}_i^t \end{pmatrix},$$

then by using the Hájek-Šidak Central Limit Theorem, we can show that Z_n^* converges in law to a normal distribution as $n \to \infty$. In order to use this theorem we need to verify the regularity condition about c_{ni}, which is given by

$$\max_i n_i c_{ni}^2 / \sum_{i=1}^k n_i c_{ni}^2 \longrightarrow 0$$

as $n \to \infty$, and it can be reformulated by requiring that as $n \to \infty$,

$$\sup_{\boldsymbol{\eta} \in \Re^{p+q}} \left[\max_i n_i \boldsymbol{\eta}^t \boldsymbol{\Gamma}_4(x_i, \boldsymbol{\theta}, \boldsymbol{\tau}) \boldsymbol{\eta} \ / \ \boldsymbol{\eta}^t \boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta}, \boldsymbol{\tau}) \boldsymbol{\eta} \right] \longrightarrow 0,$$

where

$$\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \begin{pmatrix} \boldsymbol{\Gamma}_{31n}(\boldsymbol{\theta}, \boldsymbol{\tau}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{32n}(\boldsymbol{\theta}, \boldsymbol{\tau}) \end{pmatrix}.$$

However, in view of the Courant's Theorem, we have that:

$$\begin{aligned}
\sup_{\boldsymbol{\eta}\in\Re^{p+q}} \left[n_i\boldsymbol{\eta}^t\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\eta} \;/\; \boldsymbol{\eta}^t\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\eta}\right] &= ch_1\left(n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau}),\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta},\boldsymbol{\tau})\right) \\
&= \delta_1 \\
&= ch_1(n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_{3n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau}))
\end{aligned}$$

Here, $\delta_1 \in \Re$ is the largest value of δ satisfying

$$\begin{aligned}
\det\left(n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau}) - \delta\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta},\boldsymbol{\tau})\right) &= 0 \\
\det\left(n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_{3n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau}) - \delta\mathbf{I}\right)\det\left(\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta},\boldsymbol{\tau})\right) &= 0 \\
\det\left(n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_{3n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau}) - \delta\mathbf{I}\right) &= 0.
\end{aligned}$$

Thus, δ_1 is the same as the largest eigen value of $n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_{3n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau})$. Also, since we have

$$\begin{aligned}
&ch_1(n_i\boldsymbol{\Gamma}_4(x_i,\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_{3n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau})) \\
&= \max_i\left\{n_i\sigma_{\psi i}^2 w_1^2(\mathbf{z}_i,\boldsymbol{\tau})\mathbf{f}_{\boldsymbol{\theta}}^t(x_i,\boldsymbol{\theta})\boldsymbol{\Gamma}_{31n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau})\mathbf{f}_{\boldsymbol{\theta}}^t(x_i,\boldsymbol{\theta}), n_i\sigma_{hi}^2 w_2^2(\mathbf{z}_i,\boldsymbol{\tau})\mathbf{z}_i^t\boldsymbol{\Gamma}_{32n}^{-1}(\boldsymbol{\theta},\boldsymbol{\tau})\mathbf{z}_i\right\},
\end{aligned}$$

the regularity condition is reduced to the condition **[B2]** (v)-(vi) (Noether's condition). Hence, we conclude that:

$$Z_n^*/\Big(\sum_{i,j} c_{ni}^2\Big)^{\frac{1}{2}} \longrightarrow N(0,1) \quad \text{as} \quad n \to \infty$$

and by using the Cramer-Wold Theorem and condition **[B2]** (v)-(vi) we prove the expression (3.25). □

Corollary 3.12. *Let the conditions* **[A1]**-**[A2]**, **[B1]**-**[B3]** *hold. Then*

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \\ \hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau} - \boldsymbol{\nu}_n(\boldsymbol{\theta},\boldsymbol{\tau}) \end{pmatrix} \longrightarrow N_{p+q}\big(\mathbf{0}, \boldsymbol{\Gamma}_5^{-1}(\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_3(\boldsymbol{\theta},\boldsymbol{\tau})\boldsymbol{\Gamma}_5^{-1}(\boldsymbol{\theta},\boldsymbol{\tau})\big), \tag{3.26}$$

where

$$\boldsymbol{\nu}_n(\boldsymbol{\theta},\boldsymbol{\tau}) = \left(\frac{1}{n}\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta},\boldsymbol{\tau})\right)^{-1} \bar{\boldsymbol{\mu}}_n(\boldsymbol{\tau})$$

and

$$\bar{\boldsymbol{\mu}}_n(\boldsymbol{\tau}) = \frac{1}{n}\sum_{i=1}^{k} n_i \gamma_{3i} w_2(\mathbf{z}_i,\boldsymbol{\tau})\mathbf{z}_i.$$

Proof. From Theorem 3.10 we have that:

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \\ \hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau} \end{pmatrix} = \left(\frac{1}{n}\boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta},\boldsymbol{\tau})\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i,j} \lambda(x_i, y_{ij}, \boldsymbol{\theta},\boldsymbol{\tau}) + o_p(1).$$

Then from Theorem 3.11 and the Slutsky Theorem we have the expression in (3.26).

□

Corollary 3.13. *Let the conditions* **[A1]**-**[A2]**, **[B1]**-**[B3]** *hold. Then*

$$\hat{\boldsymbol{\Gamma}}^{-\frac{1}{2}}\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \\ \hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau} - \boldsymbol{\nu}_n(\boldsymbol{\theta},\boldsymbol{\tau}) \end{pmatrix} \longrightarrow N_{p+q}(\mathbf{0}, \mathbf{I}_{p+q}), \tag{3.27}$$

where

$$\hat{\boldsymbol{\Gamma}} = \left(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{5n}(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\tau}}_n)\right)^{-1} \left(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{3n}(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\tau}}_n)\right) \left(\frac{1}{n}\hat{\boldsymbol{\Gamma}}_{5n}(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\tau}}_n)\right)^{-1}. \tag{3.28}$$

Proof. Using (3.26) and the Slustsky Theorem we have the expression in (3.27). □

Chapter 4

PTE PROCEDURES IN NONLINEAR REGRESSION MODELS

4.1 Introduction

In this chapter we introduce the PTE procedures for nonlinear regression models using WME and OME. We consider the problem of estimating the regression parameter in heteroscedastic nonlinear regression models when it is suspected that the error variances are possibly homoscedastic.

In Section 4.2 we introduce the heteroscedastic nonlinear model that we consider, and the M-estimation method. The PTE procedures using M-estimation methods are defined in Section 4.3. In Section 4.4, we derive some asymtptotic results to obtain the asymptotic covariance matrix of the PTE.

4.2 PTE Using M-estimation in Nonlinear models

4.2.1 Model and Estimation

Suppose that $(x_i,\, y_{ij}),\ i = 1, 2, \ldots, k,\ \ j = 1, \ldots, n_i$, are $n\ (= \sum n_i)$ observations from a fixed-effects nonlinear model with a known functional form f. Then,

$$y_{ij} = f(x_i; \boldsymbol{\theta}) + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_i^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \tag{4.1}$$

where ϵ_{ij} are assumed to be independent and $\boldsymbol{\theta}$ is known to belong to Θ, a subset of $\Re^p$. Here, we shall assume that all x values are the same in group i, $i = 1, 2, \ldots, k$ since it is common model setup in dose-response studies, so that k is meant to be dose groups. The OME of $\boldsymbol{\theta}$, denoted by $\tilde{\boldsymbol{\theta}}_n$, minimizes

$$S_o(\boldsymbol{\theta}) = \sum_{i,j} h^2\left(y_{ij} - f(x_i, \boldsymbol{\theta})\right), \tag{4.2}$$

assuming $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$. However, if the homoscedasticity assumption is not true, then we assume that $\log \sigma_i = \tau_0 + \tau_1 x_i$ for $i = 1, \ldots, k$. Under this assumption, the WME of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_n$, minimizes

$$S_w(\boldsymbol{\theta}) = \sum_{i,j} \frac{1}{w(\hat{\sigma}_i^2)} h^2\left(y_{ij} - f(x_i;\, \boldsymbol{\theta})\right), \tag{4.3}$$

where the functional form of w is known and $\hat{\sigma}_i = \exp(\hat{\tau}_{0n} + \hat{\tau}_{1n} x_i)$. Here, $\hat{\tau}_{0n}$ and $\hat{\tau}_{1n}$ are estimates of τ_0 and τ_1, respectively, and the detailed estimation procedure is given in the next section.

4.2.2 Preliminary Test Estimation

We now develop PTE-based methodology which naturally calibrates between homoscedasticity and heteroscedasticity. Since in toxicological studies the variability in the observed data often increases as the dose level increases and we parametrize the error variance using a simple linear form in the model (4.1), the error variance structure of the model is expressed as the following hypotheses in a preliminary test:

$$\begin{aligned} H_0 &: \quad \tau_1 = 0, \\ H_1 &: \quad \tau_1 > 0. \end{aligned}$$

If the sample variance within i^{th} group is given by

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \tag{4.4}$$

for $i = 1, \ldots, k$, then from the model (4.1), we have

$$(n_i - 1)s_i^2/\sigma_i^2 \sim \chi^2_{n_i - 1}. \tag{4.5}$$

Hence, by the definition of the chi-square distribution and the CLT,

$$\big\{(n_i - 1)s_i^2/\sigma_i^2 - (n_i - 1)\big\}/\sqrt{2(n_i - 1)} \longrightarrow N(0,1) \quad \text{as} \quad n_i \to \infty, \tag{4.6}$$

or

$$\sqrt{\frac{n_i - 1}{2}}(s_i^2 - \sigma_i^2) \longrightarrow N(0, \sigma_i^4). \tag{4.7}$$

By performing log-transformation, we have

$$\sqrt{2(n_i - 1)}(\log s_i - \log \sigma_i) \longrightarrow N(0,1) \quad \text{as} \quad n_i \to \infty. \tag{4.8}$$

From the above result for the asymptotic distribution of $\log s_i$, we can obtain the estimate of $\boldsymbol{\tau} = (\tau_0, \tau_1)^t$, $\hat{\boldsymbol{\tau}}_n = (\hat{\tau}_{0n}, \hat{\tau}_{1n})^t$, by minimizing

$$S(\boldsymbol{\tau}) = \sum_{i=1}^{k} (n_i - 1)(\log s_i - \tau_0 - \tau_1 x_i)^2 \tag{4.9}$$

Then,

$$\hat{\boldsymbol{\tau}}_n = (Z^t W Z)^{-1} Z^t W L_s, \quad \text{and} \quad \mathrm{Var}(\hat{\boldsymbol{\tau}}_n) = \frac{1}{2}(Z^t W Z)^{-1} \tag{4.10}$$

where

$$Z = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_k \end{pmatrix},$$

$W = diag.\{n_1 - 1, \ldots, n_k - 1\}$, and $L_s = (\log s_1, \ldots, \log s_k)^t$. Especially,

$$\hat{\tau}_{1n} = \frac{\sum_{i=1}^{k} (n_i - 1)(x_i - \bar{x}) \log s_i}{\sum_{i=1}^{k} (x_i - \bar{x})^2} \tag{4.11}$$

and

$$\mathrm{Var}(\hat{\tau}_{1n}) = \frac{1}{2} \left\{ \sum_{i=1}^{k} (x_i - \bar{x})^2 \right\}^{-1}, \tag{4.12}$$

where

$$\bar{x} = \frac{1}{n-k} \sum_{i=1}^{k} (n_i - 1) x_i.$$

Thus, we have the test statistic Z_n as follows:

$$Z_n = \hat{\tau}_{1n} / \sqrt{\mathrm{Var}(\hat{\tau}_{1n})}. \tag{4.13}$$

Under the null hypothesis, Z_n has the standard normal distribution asymptotically.

Then, the preliminary test estimator is defined as

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} = \tilde{\boldsymbol{\theta}}_n I(Z_n \le z_\alpha) + \hat{\boldsymbol{\theta}}_n I(Z_n > z_\alpha), \tag{4.14}$$

where z_α is the critical value of the standard normal distribution having probability $1-\alpha$, α is the significance level of the preliminary test, and $I(A)$ denotes an indicator function of the set A. Then, equivalently, we have

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} = \begin{cases} \tilde{\boldsymbol{\theta}}_n & \text{if } Z_n \le z_\alpha \\ \hat{\boldsymbol{\theta}}_n & \text{if } Z_n > z_\alpha. \end{cases}$$

4.2.3 Asymptotic Results

Since the PTE is defined as a weighted average of the OME and the WME, the asymptotic joint distribution of the OME, WME and the test statistic needs to be studied in order to obtain asymptotic properties of the PTE. In order to derive the asymptotic joint distribution, we need to have the uniform asymptotic linearity on the OME and the WME.

We have the estimating equation for the minimization in (4.3) given by

$$\sum_{i,j} \lambda_w(x_i, y_{ij}, \hat{\boldsymbol{\theta}}_n) = \mathbf{0}, \tag{4.15}$$

where

$$\lambda_w(x_i, y_{ij}, \boldsymbol{\theta}) = \frac{1}{w(\hat{\sigma}_i^2)} \psi(y_{ij} - f(x_i, \boldsymbol{\theta})) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}), \tag{4.16}$$

and ψ and $\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})$ are defined in Chapter 2.

First we make the following regularity conditions which are similar with ones in Chapter 2.

[A1]:

ψ is a nonconstant odd function which is absolutely continuous and differentiable with respect to $\boldsymbol{\theta}$.

[A2]: for $i = 1, \ldots, k$,

(i) $E\psi^2(\epsilon_{i1}) = \sigma^2_{\psi 1i} < \infty$, and $E\psi^2(\epsilon_{i1})\epsilon^2_{i1} = \sigma^2_{\psi 2i} < \infty$

(ii) $E|\psi'(\epsilon_{i1})|^{1+\delta} < \infty$ for some $0 < \delta \leq 1$, and $E\psi'(\epsilon_{i1}) = \gamma_{1i}$, and $E\psi'(\epsilon_{i1})\epsilon^2_{i1} = \gamma_{2i}$.

[A3]:

(i) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\| \leq \delta} \left|\psi(Y - f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})) - \psi(Y - f(\mathbf{x}, \boldsymbol{\theta}))\right|\right\} = 0$

(ii) $\lim_{\delta\to 0} E\left\{\sup_{\|\boldsymbol{\Delta}\| \leq \delta} \left|\psi'(Y - f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})) - \psi'(Y - f(\mathbf{x}, \boldsymbol{\theta}))\right|\right\} = 0$

[B1]:

$f(\mathbf{x}, \boldsymbol{\theta})$ is continuous and twice differentiable with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact subset of $\Re^p$.

[B2]:

(i) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_1(\boldsymbol{\theta})$, where

$$\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) = \sum_{i=1}^{k} n_i\gamma_{1i}\left\{1 - \frac{c_{ii}}{n_i}\left(\frac{\gamma_{2i}}{\sigma_i^2\gamma_{1i}} - 1\right)w_1(\sigma_i^2) - \sum_{i_1=1}^{k}\frac{c^2_{ii_1}}{n_{i_1}}w_1(\sigma_i^2) + 2\sum_{i_1=1}^{k}\frac{c^2_{ii_1}}{n_{i_1}}w_1^2(\sigma_i^2)\right\}\frac{\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})\mathbf{f}^t_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})}{w(\sigma_i^2)},$$

$c_{ii_1} = (n_{i_1} - 1)\mathbf{z}_i^t(Z^tWZ)^{-1}\mathbf{z}_{i_1}$, $\mathbf{z}_i = (1, x_i)^t$, and $w_1(z) = zw'(z)/w(z)$.

(ii) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_2(\boldsymbol{\theta})$, where

$$\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) = \sum_{i=1}^{k} n_i\sigma^2_{\psi 1i}\left\{1 - \frac{2c_{ii}}{n_i}\left(\frac{\sigma^2_{\psi 2i}}{\sigma_i^2\sigma^2_{\psi 1i}} - 1\right)w_1(\sigma_i^2) - 2\sum_{i_1=1}^{k}\frac{c^2_{ii_1}}{n_{i_1}}w_1(\sigma_i^2) + 6\sum_{i_1=1}^{k}\frac{c^2_{ii_1}}{n_{i_1}}w_1^2(\sigma_i^2)\right\}\frac{\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})\mathbf{f}^t_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})}{w^2(\sigma_i^2)},$$

and $\boldsymbol{\Gamma}_2(\boldsymbol{\theta})$ is a positive definite matrix.

(iii) $\lim_{n\to\infty} \frac{1}{n}\mathbf{\Gamma}_{3n}(\boldsymbol{\theta}) = \mathbf{\Gamma}_3(\boldsymbol{\theta})$, where

$$\mathbf{\Gamma}_{3n}(\boldsymbol{\theta}) = \sum_{i=1}^{k} n_i \gamma_{1i} \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(x_i, \boldsymbol{\theta}).$$

(iv) $\lim_{n\to\infty} \frac{1}{n}\mathbf{\Gamma}_{4n}(\boldsymbol{\theta}) = \mathbf{\Gamma}_4(\boldsymbol{\theta})$, where

$$\mathbf{\Gamma}_{4n}(\boldsymbol{\theta}) = \sum_{i=1}^{k} n_i \sigma_{\psi 1i}^2 \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^t(x_i, \boldsymbol{\theta}),$$

and $\mathbf{\Gamma}_4(\boldsymbol{\theta})$ is a positive definite matrix.

(v) $\max_i ch_1 \Big\{ n_i \mathbf{G}_1(x_i, \boldsymbol{\theta}) \big(\mathbf{G}_{2n}(\boldsymbol{\theta})\big)^{-1} \Big\} \longrightarrow 0$, as $n \to \infty$, where

$$\mathbf{G}_1(x_i, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{0} \\ \mathbf{G}_{12} & \mathbf{G}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & g_{33} \end{pmatrix},$$

$$\begin{aligned} \mathbf{G}_{11} = \frac{\sigma_{\psi 1i}^2}{w^2(\sigma_i^2)} \Bigg\{ 1 - \frac{2c_{ii}}{n_i} \left(\frac{\sigma_{\psi 2i}^2}{\sigma_i^2 \sigma_{\psi 1i}^2} - 1 \right) w_1(\sigma_i^2) - 2 \sum_{i_1=1}^{k} \frac{c_{ii_1}^2}{n_{i_1}} w_1(\sigma_i^2) \\ + 6 \sum_{i_1=1}^{k} \frac{c_{ii_1}^2}{n_{i_1}} w_1^2(\sigma_i^2) \Bigg\} \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})^t, \end{aligned}$$

$$\begin{aligned} \mathbf{G}_{12} = \frac{\sigma_{\psi 1i}^2}{w(\sigma_i^2)} \Bigg\{ 1 - \frac{c_{ii}}{n_i} \left(\frac{\sigma_{\psi 2i}^2}{\sigma_i^2 \sigma_{\psi 1i}^2} - 1 \right) w_1(\sigma_i^2) - \sum_{i_1=1}^{k} \frac{c_{ii_1}^2}{n_{i_1}} w_1(\sigma_i^2) \\ + 2 \sum_{i_1=1}^{k} \frac{c_{ii_1}^2}{n_{i_1}} w_1^2(\sigma_i^2) \Bigg\} \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})^t, \end{aligned}$$

$$\mathbf{G}_{22} = \sigma_{\psi 1i}^2 \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})^t, \quad \text{and} \quad g_{33} = \frac{2n^2 w_{i2}^2}{n_i^2}.$$

(vi) $\lim_{n\to\infty} \frac{1}{n}\mathbf{G_{2n}}(\boldsymbol{\theta}) = \mathbf{G_2}(\boldsymbol{\theta})$, where

$$\mathbf{G_{2n}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) & \boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta}) & \mathbf{0} \\ \boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta}) & \boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2n^2 \sum_{i=1}^{k} \frac{w_{i2}^2}{n_i}, \end{pmatrix}$$

$$\boldsymbol{\Gamma}_{5n}(\boldsymbol{\theta}) = \sum_{i=1}^{k} n_i \sigma_{\psi 1i}^2 \left\{ 1 - \frac{c_{ii}}{n_i} \left(\frac{\sigma_{\psi 2i}^2}{\sigma_i^2 \sigma_{\psi 1i}^2} - 1 \right) w_1(\sigma_i^2) - \sum_{i_1=1}^{k} \frac{c_{ii_1}^2}{n_{i_1}} w_1(\sigma_i^2) \right.$$
$$\left. + 2 \sum_{i_1=1}^{k} \frac{c_{ii_1}^2}{n_{i_1}} w_1^2(\sigma_i^2) \right\} \frac{\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})^t}{w(\sigma_i^2)},$$

and $\mathbf{G_2}(\boldsymbol{\theta})$ is a positive definite matrix.

[**B3**]:

(i) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big| (\partial/\partial\theta_j) f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})(\partial/\partial\theta_l) f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta})$

$- (\partial/\partial\theta_j) f(\mathbf{x}, \boldsymbol{\theta})(\partial/\partial\theta_l) f(\mathbf{x}, \boldsymbol{\theta}) \Big| = 0$ for $j, l = 1, \ldots, p$

(ii) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big| (\partial^2/\partial\theta_j\partial\theta_l) f(\mathbf{x}, \boldsymbol{\theta} + \boldsymbol{\Delta}) - (\partial^2/\partial\theta_j\partial\theta_l) f(\mathbf{x}, \boldsymbol{\theta}) \Big| = 0$ for $j, l = 1, \ldots, p$

Now we shall prove the uniform asymptotic linearity on the WME given in the following theorem.

Theorem 4.1. *Let the conditions* [**A1**]-[**A3**], *and* [**B1**]-[**B3**] *hold. Then*

$$\sup_{\|\mathbf{t}\|\le C} \left\| \frac{1}{\sqrt{n}} \sum_{i,j} \{\lambda_w(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}) - \lambda_w(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\} + \frac{1}{n} \boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\mathbf{t} \right\| = o_p(1) \quad (4.17)$$

as $n \to \infty$.

Proof. The sample variance within group, s_i^2 defined in (4.4) is a U-statistic with

$f(x_1, x_2) = (x_1 - x_2)^2/2$, which can be expressed as

$$s_i^2 = \frac{1}{2}\binom{n_i}{2}^{-1} \sum\sum_{1 \le j < j_1 \le n_i} (\epsilon_{ij} - \epsilon_{ij_1})^2 \tag{4.18}$$

Then, by the asymptotic property of U-statistic,

$$\begin{aligned} s_i^2 - \sigma_i^2 &= \frac{2}{n_i}\sum_{j=1}^{n_i}\left\{h_1(\epsilon_{ij}) - \sigma_i^2\right\} + O_p\Big(\frac{1}{n_i}\Big), \\ &= \frac{1}{n_i}\sum_{j=1}^{n_i}(\epsilon_{ij}^2 - \sigma_i^2) + O_p\Big(\frac{1}{n_i}\Big), \end{aligned} \tag{4.19}$$

where

$$h_1(\epsilon_{ij}) = E\left[\frac{1}{2}(\epsilon_{ij} - \epsilon_{ij_1})^2 \,\Big|\, \epsilon_{ij}\right] = \frac{1}{2}(\epsilon_{ij} + \sigma_i^2).$$

Thus, we have

$$\begin{aligned} \log s_i - \log \sigma_i &= \log\left\{1 + \frac{s_i^2 - \sigma_i^2}{\sigma_i^2}\right\} \\ &= \frac{s_i^2 - \sigma_i^2}{2\sigma_i^2} + O_p\Big(\frac{1}{n_i}\Big) \\ &= \frac{1}{2\sigma_i^2}\frac{1}{n_i}\sum_{j=1}^{n_i}(\epsilon_{ij}^2 - \sigma_i^2) + O_p\Big(\frac{1}{n_i}\Big), \end{aligned} \tag{4.20}$$

and hence,

$$\begin{aligned} \hat{\boldsymbol{\tau}}_n &= (Z^tWZ)^{-1}Z^tWL_s \\ &= \sum_{i=1}^{k} \mathbf{w}_i \log s_i \\ &= \sum_{i=1}^{k} \mathbf{w}_i\big\{\log \sigma_i + (\log s_i - \log \sigma_i)\big\} \\ &= \boldsymbol{\tau} + \sum_{i=1}^{k} \frac{\mathbf{w}_i}{2\sigma_i^2}\frac{1}{n_i}\sum_{j=1}^{n_i}(\epsilon_{ij}^2 - \sigma_i^2) + O_p\Big(\frac{1}{n}\Big), \end{aligned} \tag{4.21}$$

where $\mathbf{w}_i = (n_i - 1)(Z^tWZ)^{-1}\mathbf{z}_i$. Now, if we define

$$\mathbf{u}_n = \sqrt{n}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) = \frac{\sqrt{n}}{2}\sum_{i,j}\frac{\mathbf{w}_i}{n_i\sigma_i^2}(\epsilon_{ij}^2 - \sigma_i^2) + o_p(1), \tag{4.22}$$

then

$$\frac{\hat{\sigma}_i^2}{\sigma_i^2} = \frac{\exp(2\mathbf{z}_i^t\hat{\boldsymbol{\tau}}_n)}{\exp(2\mathbf{z}_i^t\boldsymbol{\tau})} = \exp\left\{2\mathbf{z}_i^t(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau})\right\} = 1 + \frac{2}{\sqrt{n}}\mathbf{z}_i^t\mathbf{u}_n + + \frac{2}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2 + o_p\Big(\frac{1}{n}\Big), \tag{4.23}$$

$$\begin{aligned} w(\hat{\sigma}_i^2) &= w\Big(\sigma_i^2\,\frac{\hat{\sigma}_i^2}{\sigma_i^2}\Big) = w\Big(\sigma_i^2 + \frac{2}{\sqrt{n}}\mathbf{z}_i^t\mathbf{u}_n\sigma_i^2 + \frac{2}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2\sigma_i^2 + o_p\Big(\frac{1}{n}\Big)\Big) \\ &= w(\sigma_i^2) + \left\{\frac{2}{\sqrt{n}}\mathbf{z}_i^t\mathbf{u}_n + \frac{2}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2\right\}\sigma_i^2 w'(\sigma_i^2) + o_p\Big(\frac{1}{n}\Big), \end{aligned} \tag{4.24}$$

and hence,

$$\frac{1}{w(\hat{\sigma}_i^2)} = \frac{1}{w(\sigma_i^2)}\left[1 - \left\{\frac{2}{\sqrt{n}}\mathbf{z}_i^t\mathbf{u}_n + \frac{2}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2\right\}w_1(\sigma_i^2) + \frac{4}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2 w_1^2(\sigma_i^2) + o_p\Big(\frac{1}{n}\Big)\right] \tag{4.25}$$

Therefore,

$$E\psi\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big)\frac{1}{w(\hat{\sigma}_i^2)} = 0, \tag{4.26}$$

and since

$$E\psi'(\epsilon_{ij})\frac{2}{\sqrt{n}}\mathbf{z}_i^t\mathbf{u}_n = \frac{c_{ii}}{n_i\sigma_i^2}(\gamma_{2i} - \sigma_i^2\gamma_{1i}) \tag{4.27}$$

and

$$\begin{aligned} &E\psi'(\epsilon_{ij})\frac{2}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2 \\ &= \frac{2}{n}\frac{n}{4}\left\{\gamma_{1i}\left(\sum_{i_1\neq i}\sum_{j_1=1}^{n_{i_1}}\frac{c_{ii_1}^2}{n_{i_1}^2\sigma_{i_1}^4}(2\sigma_{i_1}^4) + \sum_{i_1=1}^{k}\sum_{j_1\neq j}\frac{c_{ii_1}^2}{n_{i_1}^2\sigma_{i_1}^4}(2\sigma_{i_1}^4)\right.\right. \\ &\qquad\left.\left. - \sum_{i_1\neq i}\sum_{j_1\neq j}\frac{c_{ii_1}^2}{n_{i_1}^2\sigma_{i_1}^4}(2\sigma_{i_1}^4)\right) + o\Big(\frac{1}{n}\Big)\right\} \\ &= \gamma_{1i}\sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}} + o\Big(\frac{1}{n}\Big), \end{aligned} \tag{4.28}$$

$$
E\psi'\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big)\frac{1}{w(\hat{\sigma}_i^2)}
$$

$$
= \frac{\gamma_{1i}}{w(\sigma_i^2)}\left\{1 - \frac{c_{ii}}{n_i}\left(\frac{\gamma_{2i}}{\sigma_i^2\gamma_{1i}} - 1\right)w_1(\sigma_i^2) - \sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}}w_1(\sigma_i^2) + 2\sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}}w_1^2(\sigma_i^2)\right\} + o\Big(\frac{1}{n}\Big) \tag{4.29}
$$

We can prove this theorem with exactly the same method used in Theorem 2.3, and obtain (4.17) from (4.26) and (4.29). □

From (4.17), we obtain the following asymptotic representation of $\hat{\boldsymbol{\theta}}_n$ similarly as in Theorem 2.4:

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \left(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i,j}\psi\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big)\frac{\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta})}{w(\hat{\sigma}_i^2)} + o_p(1) \tag{4.30}
$$

We now consider the OME and have the estimating equation for the minimization in (4.2) given by

$$
\sum_{i,j}\lambda_o(x_i, y_{ij}, \tilde{\boldsymbol{\theta}}_n) = \mathbf{0}, \tag{4.31}
$$

where

$$
\lambda_o(x_i, y_{ij}, \boldsymbol{\theta}) = \psi(y_{ij} - f(x_i, \boldsymbol{\theta}))\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}). \tag{4.32}
$$

Then, similarly, we have the uniform asymptotic linearity on the OME:

$$
\sup_{\|\mathbf{t}\|\le C}\left\|\frac{1}{\sqrt{n}}\sum_{i,j}\{\lambda_o(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta} + n^{-\frac{1}{2}}\mathbf{t}) - \lambda_o(\mathbf{x}_i, y_{ij}, \boldsymbol{\theta})\} + \frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta})\mathbf{t}\right\| = o_p(1), \tag{4.33}
$$

and hence we obtain the follwing asymptotic representation of $\tilde{\boldsymbol{\theta}}_n$:

$$
\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \left(\frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta})\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i,j}\psi\big(y_{ij} - f(x_i, \boldsymbol{\theta})\big)\mathbf{f}_{\boldsymbol{\theta}}(x_i, \boldsymbol{\theta}) + o_p(1) \tag{4.34}
$$

Now we shall prove the asymptotic normality of $(\hat{\boldsymbol{\theta}}_n^t, \tilde{\boldsymbol{\theta}}_n^t, \hat{\tau}_{1n})^t$.

Theorem 4.2. *Let the conditions* **[A1]**-**[A3]**, *and* **[B1]**-**[B3]** *hold. Then*

$$\frac{1}{\sqrt{n}}\sum_{i,j}\lambda^*(x_i,y_{ij},\boldsymbol{\theta}) \longrightarrow N_{2p+1}(\mathbf{0},\mathbf{G}_2(\boldsymbol{\theta})) \quad as \quad n\to\infty, \tag{4.35}$$

where

$$\lambda^*(x_i,y_{ij},\boldsymbol{\theta}) = \Big(\lambda_w^t(x_i,y_{ij},\boldsymbol{\theta}),\lambda_o^t(x_i,y_{ij},\boldsymbol{\theta}),\frac{nw_{i2}}{n_i\sigma_i^2}(\epsilon_{ij}^2-\sigma_i^2)\Big)^t,$$

and w_{i2} *is the second element of* $\mathbf{w}_i$.

Proof. We consider an arbitrary linear compound:

$$Z_n^* = \boldsymbol{\eta}^t\frac{1}{\sqrt{n}}\sum_{i,j}\lambda^*(x_i,y_{ij},\boldsymbol{\theta}),$$

where $\boldsymbol{\eta}=(\boldsymbol{\eta}_1^t,\boldsymbol{\eta}_2^t,\eta_3)^t$ and $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2\in\Re^p$; $\eta\in\Re$. And we have that:

$$\begin{aligned} Z_n^* &= \sum_{i,j}\frac{1}{\sqrt{n}}\Big\{\boldsymbol{\eta}_1^t\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))\frac{1}{w(\hat{\sigma}_i^2)}+\boldsymbol{\eta}_2^t\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})\psi(y_{ij}-f(x_i,\boldsymbol{\theta})) \\ &\qquad +\frac{n\eta_3 w_{i2}}{n_i\sigma_i^2}(\epsilon_{ij}^2-\sigma_i^2)\Big\} \\ &= \sum_{i,j}\frac{1}{\sqrt{n}}\big(c_{i1}Z_{ij1}+c_{i2}Z_{ij2}+c_{i3}Z_{ij3}\big), \end{aligned}$$

where $c_{i1}=\boldsymbol{\eta}_1^t\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})$, $c_{i2}=\boldsymbol{\eta}_2^t\mathbf{f}_{\boldsymbol{\theta}}(x_i,\boldsymbol{\theta})$, $c_{i3}=\frac{n\eta_3 w_{i2}}{n_i\sigma_i^2}$, $Z_{ij1}=\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))\frac{1}{w(\hat{\sigma}_i^2)}$, $Z_{ij2}=\psi(y_{ij}-f(x_i,\boldsymbol{\theta}))$, and $Z_{ij3}=\epsilon_{ij}^2-\sigma_i^2$. Then,

$$EZ_{ij1}=EZ_{ij2}=EZ_{ij3}=0,$$

$$EZ_{ij2}^2=\sigma_{\psi 1i}^2;\quad EZ_{ij3}^2=2\sigma_i^4;$$

since

$$\left(\frac{1}{w(\hat{\sigma}_i^2)}\right)^2 = \frac{1}{w^2(\sigma_i^2)}\left[1 - 2\left\{\frac{2}{\sqrt{n}}\mathbf{z}_i^t\mathbf{u}_n + \frac{2}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2\right\}w_1(\sigma_i^2) + \frac{12}{n}(\mathbf{z}_i^t\mathbf{u}_n)^2 w_1^2(\sigma_i^2) + o_p\Big(\frac{1}{n}\Big)\right],$$

if we compute similarly as in (4.27) and (4.28),

$$EZ_{ij1}^2 = \frac{\sigma_{\psi 1i}^2}{w^2(\sigma_i^2)}\left\{1 - \frac{2c_{ii}}{n_i}\left(\frac{\sigma_{\psi 2i}^2}{\sigma_i^2\sigma_{\psi 1i}^2} - 1\right)w_1(\sigma_i^2) - 2\sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}}w_1(\sigma_i^2) + 6\sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}}w_1^2(\sigma_i^2)\right\} + o\Big(\frac{1}{n}\Big), \tag{4.36}$$

and $EZ_{ij1}Z_{ij3} = EZ_{ij2}Z_{ij3} = 0$;

$$EZ_{ij1}Z_{ij2} = \frac{\sigma_{\psi 1i}^2}{w(\sigma_i^2)}\left\{1 - \frac{c_{ii}}{n_i}\left(\frac{\sigma_{\psi 2i}^2}{\sigma_i^2\sigma_{\psi 1i}^2} - 1\right)w_1(\sigma_i^2) - \sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}}w_1(\sigma_i^2) + 2\sum_{i_1=1}^{k}\frac{c_{ii_1}^2}{n_{i_1}}w_1^2(\sigma_i^2)\right\} + o\Big(\frac{1}{n}\Big)$$

Therefore, if we let

$$Z_n^* = \sum_{i,j} c_{ni}Z_{nij}, \tag{4.37}$$

where

$$c_{ni} = \frac{1}{\sqrt{n}}\left\{\boldsymbol{\eta}^t\mathbf{G}_1(x_i,\boldsymbol{\theta})\boldsymbol{\eta}\right\}^{\frac{1}{2}}, \tag{4.38}$$

and

$$Z_{nij} = \left(c_{i1}Z_{ij1} + c_{i2}Z_{ij2} + c_{i3}Z_{ij3}\right)/\left\{\boldsymbol{\eta}^t\mathbf{G}_1(x_i,\boldsymbol{\theta})\boldsymbol{\eta}\right\}^{\frac{1}{2}}, \tag{4.39}$$

then

$$EZ_{nij} = 0 \quad \text{and} \quad \mathrm{Var}(Z_{nij}) = 1 + o\Big(\frac{1}{n}\Big).$$

Then by using the Hájek-Šidak Central Limit Theorem, we can show that Z_n^*

converges in law to a normal distribution as $n \to \infty$. In order to use this theorem we need to verify the regularity condition about c_{nij}, which is given by

$$\max_i n_i c_{ni}^2 / \sum_i n_i c_{ni}^2 \longrightarrow 0.$$

However, since

$$\sum_i n_i c_{ni}^2 = \frac{1}{n} \boldsymbol{\eta}^t \mathbf{G_{2n}}(\boldsymbol{\theta}) \boldsymbol{\eta},$$

this condition is reduced to the condition **[B2]** (v) similarly in Theorme 2.5. Thus, we conclude that

$$Z_n^* / \Big(\sum_{i=1}^k n_i c_{ni}^2 \Big)^{\frac{1}{2}} \longrightarrow N(0,1) \ \text{ as } \ n \to \infty$$

and by using the Cramer-Wold Theorem we prove the expression in (4.35) □

Corollary 4.3. *Let the conditions* **[A1]**-**[A3]**, *and* **[B1]**-**[B3]** *hold. Then,*

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right) \longrightarrow N_{2p+1}(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta})) \ \ as \ \ n \to \infty, \tag{4.40}$$

where

$$\boldsymbol{\beta} = \left(\boldsymbol{\theta}^t, \boldsymbol{\theta}^t, \tau_1 \right)^t, \qquad \hat{\boldsymbol{\beta}}_n = \left(\hat{\boldsymbol{\theta}}_n^t, \tilde{\boldsymbol{\theta}}_n^t, \hat{\tau}_{1n} \right)^t,$$

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{G}_3^{-1}(\boldsymbol{\theta}) \mathbf{G}_2(\boldsymbol{\theta}) \mathbf{G}_3^{-1}(\boldsymbol{\theta}),$$

and

$$\mathbf{G}_3(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Gamma}_1(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_3(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2 \end{pmatrix}.$$

Proof. From (4.22), (4.30), and (4.34), we have that:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = \mathbf{G}_{3n}^{-1}(\boldsymbol{\theta})\frac{1}{\sqrt{n}}\sum_{i,j}\lambda^*(x_i, y_{ij}, \boldsymbol{\theta}) + o_p(1),$$

where

$$\mathbf{G}_{3n}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2 \end{pmatrix}.$$

Then from Theorem 4.2 and the Slutsky Theorem we have the expression in (4.40).

□

From the result of the asymptotic joint distribution of $\hat{\boldsymbol{\theta}}_n$, $\tilde{\boldsymbol{\theta}}_n$, $\hat{\tau}_{1n}$, the asymptotic covariance matrix of the PTE can be obtained.

Corollary 4.4. *Let the conditions* [**A1**]-[**A3**], *and* [**B1**]-[**B3**] *hold. Then,*

$$\begin{aligned} &E\left[n\left(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}\right)^t\right] \\ &= \Phi\left(z_\alpha - \frac{\tau_1}{\sqrt{\mathrm{Var}(\hat{\tau}_{1n})}}\right)\left(\frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\boldsymbol{\Gamma}_{4n}(\boldsymbol{\theta})\right)\left(\frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta})\right)^{-1} \\ &\qquad + \left\{1 - \Phi\left(z_\alpha - \frac{\tau_1}{\sqrt{\mathrm{Var}(\hat{\tau}_{1n})}}\right)\right\}\left(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta})\right)\left(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}, \end{aligned} \tag{4.41}$$

where Φ *is the cdf of the standard normal random variable.*

Proof. From (4.14), we have, for arbitrary $\mathbf{x} \in \Re^p$,

$$\begin{aligned}
&P\left\{\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}\right) \le \mathbf{x}\right\} \\
&= P\left\{\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \le \mathbf{x},\, Z_n \le z_\alpha\right\} + P\left\{\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \le \mathbf{x},\, Z_n > z_\alpha\right\} \\
&= P\left\{\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \le \mathbf{x}\right\} P\left\{\sqrt{n}\left(\hat{\tau}_{1n} - \tau_1\right) \le \sqrt{n}\left(z_\alpha - \tau_1\right)\right\} \\
&\qquad + P\left\{\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \le \mathbf{x}\right\} P\left\{\sqrt{n}\left(\hat{\tau}_{1n} - \tau_1\right) > \sqrt{n}\left(z_\alpha - \tau_1\right)\right\} \qquad (4.42) \\
&= \Phi\left(z_\alpha - \frac{\tau_1}{\sqrt{\mathrm{Var}(\hat{\tau}_{1n})}}\right) P\left\{\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \le \mathbf{x}\right\} \\
&\qquad + \left\{1 - \Phi\left(z_\alpha - \frac{\tau_1}{\sqrt{\mathrm{Var}(\hat{\tau}_{1n})}}\right)\right\} P\left\{\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \le \mathbf{x}\right\}.
\end{aligned}$$

Then from (4.40),

$$\begin{aligned}
&E\left[n\left(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}} - \hat{\boldsymbol{\theta}}\right)^t\right] \\
&= \Phi\left(z_\alpha - \frac{\tau_1}{\sqrt{\mathrm{Var}(\hat{\tau}_{1n})}}\right) E\left[n\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\right] \qquad (4.43) \\
&\qquad + \left\{1 - \Phi\left(z_\alpha - \frac{\tau_1}{\sqrt{\mathrm{Var}(\hat{\tau}_{1n})}}\right)\right\} E\left[n\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\right],
\end{aligned}$$

and since from (4.40),

$$E\left[n\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)^t\right] = \left(\frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\boldsymbol{\Gamma}_{4n}(\boldsymbol{\theta})\right)\left(\frac{1}{n}\boldsymbol{\Gamma}_{3n}(\boldsymbol{\theta})\right)^{-1} \qquad (4.44)$$

and

$$E\left[n\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)^t\right] = \left(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}\left(\frac{1}{n}\boldsymbol{\Gamma}_{2n}(\boldsymbol{\theta})\right)\left(\frac{1}{n}\boldsymbol{\Gamma}_{1n}(\boldsymbol{\theta})\right)^{-1}, \qquad (4.45)$$

we have the expression in (4.41). □

Chapter 5

SIMULATION STUDY

5.1 Introduction

The theory and methodology developed in the previous chapters can be applied to many nonlinear models. We illustrate the PTE procedure proposed in Chapter 4 for homoscedastic and heteroscedastic data.

We divide this chapter in two parts. In Section 5.2, we illustrate OME, WME and PTE methodologies using simulated data sets. In that section we demonstrate how PTE chooses the appropriate method without costing too much in terms of the standard error of the estimates. We compare the performance of OME, WME and PTE in Section 5.3 using a simulation study.

5.2 Illustration of PTE using Simulated Data

We generated three data sets using the Hill model,

$$y_{ij} = f(x_i, \boldsymbol{\theta}) + \epsilon_{ij} = \theta_0 + \frac{\theta_1 x_i^{\theta_2}}{\theta_3^{\theta_2} + x_i^{\theta_2}} + \epsilon_{ij}, \quad i = 1, \ldots, 8, \quad j = 1, \ldots, 4. \tag{5.1}$$

We set x_i having the values of $0, 1, 3, 10, 30, 100, 400, 600$, and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1, 4, 1.5, 100)$. The error variances are as follows:

$$\begin{aligned} &\text{Data 1:} \quad \epsilon_{ij} \sim N(0,\, e^{-4}), \quad i = 1, \ldots, 8, \quad j = 1, \ldots, 4, \\ &\text{Data 2:} \quad \epsilon_{ij} \sim N(0,\, e^{-6+0.01x_i}), \quad i = 1, \ldots, 8, \quad j = 1, \ldots, 4, \\ &\text{Data 3:} \quad \epsilon_{ij} \sim N(0,\, 0.01 f^2(x_i; \boldsymbol{\theta})), \quad i = 1, \ldots, 8, \quad j = 1, \ldots, 4. \end{aligned} \tag{5.2}$$

Thus in the case of Data 1 and Data 2, the error variance model described in Chapter 4 are correct. However, in the case of Data 3 the error variance model of Chapter 4 is incorrect. Using Data 3, we evaluated the robustness of the proposed PTE methodology to mis-specified variance structure.

The results of OME, WME and PTE estimates (and their standard errors) for the three data sets are summarized in Table 5.1. As expected, for Data 1 the standard errors of OME were smaller than those of WME. Furthermore OME and WME were substantially different. PTE automatically selects OME and the standard errors of PTE were much less than those of WME. Similarly, as expected, the converse is true in the case of heteroscedastic data (Data 2 and Data 3).

Note that if the data are homoscedastic then the “correct” choice of estimator is OME, whereas for heteroscedastic data (Data 2), the “correct” choice is WME. However in a practical setting, for a given data set, one does not know *a priori* whether the data are homoscedastic or heteroscedastic. In all three data sets, PTE automatically chooses the “correct” estimation procedure (OME or WME) while

Table 5.1: Estimate and Standard Error for parameters of the models for Data 1, 2 and 3 using OME, WME and PTE methods.

		OME		*WME*		*PTE*	
		Estimate	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*
	θ_0	2.326	0.089	2.292	0.091	2.326	0.089
Data 1	θ_1	3.702	0.243	4.109	0.382	3.702	0.251
(homo.)	θ_2	1.585	0.324	1.247	0.257	1.585	0.321
	θ_3	98.997	9.888	122.361	23.263	98.997	10.952
	θ_0	2.301	0.040	2.284	0.013	2.284	0.013
Data 2	θ_1	4.175	0.864	4.496	0.240	4.496	0.250
(hetero.)	θ_2	1.541	0.442	1.377	0.062	1.377	0.073
	θ_3	106.152	33.676	120.161	10.154	120.161	10.515
	θ_0	2.255	0.108	2.280	0.085	2.280	0.090
Data 3	θ_1	4.487	1.167	4.137	0.400	4.137	0.630
(hetero.)	θ_2	1.082	0.324	1.225	0.247	1.225	0.264
	θ_3	121.974	64.695	102.845	19.331	102.845	33.560

keeping the standard error as small as that of the “correct” estimation procedure.

5.3 Comparison of OME, WME and PTE

Three data sets are generated from (5.1) and (5.2) with the total sample size of 40, and the same values of x_i and $\boldsymbol{\theta}$. Using 10,000 simulation runs, we compared OME, WME and PTE in terms of three standard criteria: (1) mean squared error (MSE) of individual parameters as well as all parameters simultaneously, (2) the coverage probabilities of 95% confidence intervals of individual parameters as well as the simultaneous confidence ellipsoid defined below, and (3) the size of the 95% confidence regions. The $100(1-\alpha)\%$ confidence region for the parameter vector $\boldsymbol{\theta}$ is defined as

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t \left[\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})\right]^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq pF_{p,n-p}(\alpha),$$

where $\hat{\boldsymbol{\theta}}$ is the point estimator and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})$ is the appropriate variance estimator.

We also consider the total MSE of the model based on the data sets generated from (5.1) and the following error structure:

$$\epsilon_{ij} \sim N(0,\, e^{-4+\tau_1 x_i}), \quad i = 1, \ldots, 8, \quad j = 1, \ldots, 5.$$

Total MSEs are computed based on 1,000 simulation runs using OME, WME and PTE with various values of τ_1.

5.3.1 Mean Squared Error (MSE)

Table 5.2 shows the results of the simulation. When data are generated from homoscedastic model, the estimated MSE of OME for θ_3 (ED_{50}) is smaller than that of WME and the estimated MSE of PTE is slightly larger than that of OME. On the other hand, when data are heteroscedastic, the estimated MSEs of WME are much smaller than those of OME and the estimated MSEs of the PTE are exactly same as those of WME (Data 2), or slightly larger than those of WME (Data 3). Therefore,

Table 5.2: The estimated MSE for parameters of the models for Data 1, 2 and 3 using OME, WME and PTE methods ($n = 40$; 10,000 simulations).

		OME	*WME*	*PTE*
	θ_0	0.001	0.001	0.001
Data 1	θ_1	0.017	0.019	0.017
(homo.)	θ_2	0.017	0.017	0.017
	θ_3	45.891	49.984	46.166
	Total	45.926	50.021	46.201
	θ_0	0.0008	0.0002	0.0002
Data 2	θ_1	0.535	0.069	0.069
(hetero.)	θ_2	0.063	0.006	0.006
	θ_3	1922	87.571	87.571
	Total	1922	87.645	87.645
	θ_0	0.0041	0.0038	0.0039
Data 3	θ_1	0.308	0.234	0.247
(hetero.)	θ_2	0.138	0.110	0.115
	θ_3	1510	1016	1073
	Total	1510	1017	1073

Table 5.3: The estimated total MSE of the models with various values of τ_1 using OME, WME and PTE methods ($n = 40$; 1,000 simulations).

$100\tau_1$	*OME*	*WME*	*PTE*
0	45.91	48.62	46.17
0.01	56.44	60.27	56.44
0.05	63.59	62.09	62.15
0.08	76.80	73.38	74.51
0.10	81.43	71.21	74.20
0.15	130.63	96.47	100.48
0.18	154.80	89.30	93.09
0.20	299.77	123.72	131.58
0.30	2060.17	203.94	204.36
0.50	3474.84	958.65	958.65

we see that the choice between OME and WME may affect the estimation result seriously and PTE improves the performance of parameter estimation.

From Table 5.3 we can see that as τ_1 increases, the total MSE increases in all cases using OME, WME and PTE methods. However, when τ_1 is 0 or very close to 0, i.e., for (almost) homoscedastic data, the total MSEs of OME are smaller than those of WME, and the total MSEs of PTE are slightly larger than those of OME. On the other hand, as τ_1 gets larger, the total MSEs of OME get much larger than those of WME, and the total MSEs of PTE are now slightly larger than those of WME. Also, it is seen that the difference between the total MSEs of WME and PTE gets decreased to 0 by increasing τ_1. Figure 5.1 also shows this result.

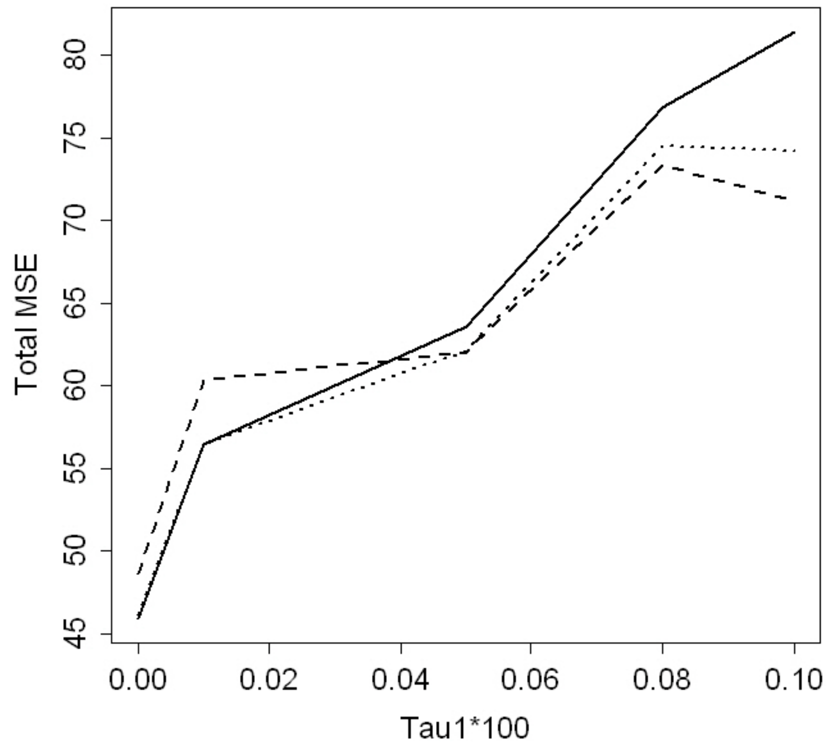


Figure 5.1: Plot of total MSE against τ_1 using OME (solid line), WME (dashed line), and PTE (dotted line).

5.3.2 Coverage Probability

The results in Table 5.4 show that all three methods are subject to under-coverage for homoscedastic as well as heteroscedastic data (except for OME of θ_0 for Data 2). However, as with the MSE, the coverage probability of PTE was closer to that of OME for homoscedastic data and closer to that of WME for heteroscedastic data.

5.3.3 Length of CI

From Table 5.5, we can see that for homoscedastic data, the length of 95% CIs for parameters are all quite similar using the OME, WME and PTE methods. However, for Data 2 (heteroscedastic data) WME has much shorter CIs than OME, and the

Table 5.4: Coverage probability for parameters of the models for Data 1, 2 and 3 using OME, WME and PTE methods. Nominal coverage probability = 0.95 ($n = 40$; 10,000 simulations).

		OME	*WME*	*PTE*
	θ_0	0.932	0.919	0.931
Data 1	θ_1	0.935	0.916	0.933
(homo.)	θ_2	0.924	0.904	0.921
	θ_3	0.919	0.908	0.918
	$\theta_0 + \theta_1$	0.934	0.914	0.933
	θ_0	0.958	0.939	0.946
Data 2	θ_1	0.915	0.886	0.897
(hetero.)	θ_2	0.922	0.927	0.944
	θ_3	0.907	0.896	0.911
	$\theta_0 + \theta_1$	0.917	0.884	0.895
	θ_0	0.938	0.928	0.930
Data 3	θ_1	0.930	0.919	0.924
(hetero.)	θ_2	0.934	0.912	0.916
	θ_3	0.902	0.884	0.888
	$\theta_0 + \theta_1$	0.929	0.916	0.922

lengths of CI for PTE are very similar with the corresponding ones for WME. For Data 3, which is heteroscedastic, the lengths of CI are shorter using WME than OME, and PTE are between the two.

Table 5.5: Length of 95% CI for parameters of the models for Data 1, 2 and 3 using OME, WME and PTE methods ($n = 40$; 10,000 simulations).

		OME	*WME*	*PTE*
	θ_0	0.067	0.064	0.067
Data 1	θ_1	0.246	0.235	0.245
(homo.)	θ_2	0.240	0.230	0.238
	θ_3	12.274	11.907	12.231
	$\theta_0 + \theta_1$	0.215	0.205	0.214
	θ_0	0.056	0.027	0.027
Data 2	θ_1	1.291	0.446	0.475
(hetero.)	θ_2	0.499	0.142	0.153
	θ_3	70.423	16.166	18.863
	$\theta_0 + \theta_1$	1.247	0.440	0.468
	θ_0	0.125	0.116	0.117
Data 3	θ_1	0.984	0.835	0.866
(hetero.)	θ_2	0.687	0.566	0.595
	θ_3	64.566	50.028	53.817
	$\theta_0 + \theta_1$	0.934	0.794	0.824

5.3.4 Simultaneous Confidence Region

We compute the coverage probability and the volume of the ellipsoid using the three estimation methods for various total sample sizes. The homoscedastic data (Data 1) are used and the total sample size are taken to be 24, 40, 104, 152, and 200. Simulation results are based on 1,000 simulation runs, except for $n = 40$ (10,000 runs for $n = 40$,).

Table 5.6: Coverage probability and the volume of the ellipsoid for Data 1 using OME, WME and PTE methods. Nominal coverage probability = 0.95 (1,000 simulations except for $n = 40$).

n		*OME*	*WME*	*PTE*
24	*Cov. Prob.*	0.755	0.684	0.747
	Volume	0.755	0.621	0.739
40	*Cov. Prob.*	0.845	0.813	0.843
	Volume	0.281	0.246	0.277
104	*Cov. Prob.*	0.903	0.895	0.902
	Volume	0.044	0.041	0.044
152	*Cov. Prob.*	0.938	0.930	0.936
	Volume	0.020	0.019	0.020
200	*Cov. Prob.*	0.934	0.934	0.933
	Volume	0.012	0.012	0.012

The results of simulation are summarized in Table 5.6. All methods suffer from severe under coverage for smaller sample sizes and the convergence is rather slow. Even at $n = 200$ the coverage probability is still smaller then 0.95 for all methods. From a practical poin of view, researchers are interested in “rectagular regions” rather than ellipses. So using Bonferroni adjusted simultaneous confidence intervals, we not only derive simultaneous confidence intervals that are easy to interpret by researchers, but we also attain close to nominal values.

Chapter 6

REAL DATA APPLICATION

6.1 Description of the Data

In this chapter we apply the PTE methodology proposed in Chapter 4 to several data sets from a toxicological study that was designed to examine the relationship between concentrations of Hexavalent Chromium (CrVI), as sodium dichromate dihydrate, in drinking water and accumulation of total chromium in tissue for three species (rats, mice, and guinea pigs) (NTP, 2007).

Groups of four Fischer 344 rats, four B6C3F1 mice, and four Hartley guinea pigs were randomly assigned to one of six concentrations of sodium dichromate dihydrate in their drinking water. All animals were between 6 and 10 weeks in age. Control groups were given water without added sodium dichromate dihydrate. The dose concentrations were 0, 2.87, 8.62, 28.7, 86.2, 287, and 862 mg sodium dichromate dihyrate/L (to yield 0, 1, 3, 10, 30, 100, and 300 mg chromium/L). When animals were sacrificed, total chromium concentrations in blood, kidneys, and femurs were measured.

6.2 Analysis and Results

The Hill model in Chapter 5 is used for fitting the data, where x is dose concentration, ranged from 0 to 300 and y is total chromium concentration. We illustrate the proposed methodology using the following three data sets where we model chromium concentration (y):

- Data 1: in blood for guinea pig,
- Data 2: in kidney for guinea pig,
- Data 3: in blood for rat.

We have 7 values of x and 4 observations at each x except for $x = 30$ for Data 1 and 2; $x = 0$ for Data 3 (3 observations at $x = 30$ for Data 1 and 2; 3 observations at $x = 0$ for Data 3), so that the total sample size is 27. For each of the data sets we estimate the parameters and their standard errors using the OME, WME, and PTE methods. the LSE methods are also used, which we can compare with the M-estimation methods. The results are summarized in Table 6.1 and 6.2, and the data and the fitted curve using the M-estimation and LSE methods for Data 1, 2 and 3 are plotted in Figure 6.1, 6.2 and 6.3, respectively.

First of all, from Table 6.1 and 6.2, we can see that the results from both the M-estimation and LSE methods are almost the same each other, which suggests that there may be no outliers or influential observations in Data 1, 2 and 3. Figure 6.1, 6.2 and 6.3 also show the same result, so that the fitted curves from both the M-estimation and LSE methods look no difference.

From Figure 6.1 and Table 6.1, we see that Data 1 appear to be heteroscedastic, and the point estimates dramatically differ between OME and WME. Also the estimated standard errors of WME are much smaller than those of OME, and the standard errors of PTE are very similar to those of WME.

Table 6.1: Estimate and Standard Error for parameters of the models for Chromium data using OME, WME and PTE methods.

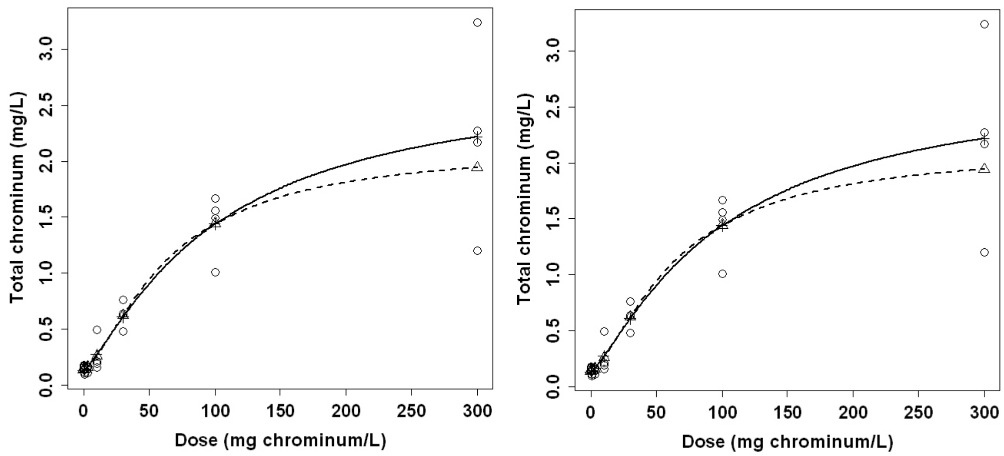
		OME		*WME*		*PTE*	
		Estimate	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*
Data 1	θ_0	0.134	0.018	0.141	0.011	0.141	0.011
	θ_1	2.620	1.060	1.995	0.500	1.995	0.501
	θ_2	1.244	0.343	1.468	0.198	1.468	0.198
	θ_3	100.766	70.396	65.254	26.304	65.254	26.372
Data 2	θ_0	0.119	0.033	0.120	0.012	0.120	0.012
	θ_1	3.272	0.911	3.166	0.648	3.166	0.649
	θ_2	1.506	0.632	1.542	0.267	1.542	0.269
	θ_3	71.118	41.699	67.698	24.188	67.698	24.263
Data 3	θ_0	0.108	0.009	0.111	0.008	0.111	0.008
	θ_1	0.751	0.113	0.643	0.062	0.643	0.086
	θ_2	0.988	0.174	1.128	0.159	1.128	0.165
	θ_3	95.859	33.666	68.453	14.715	68.453	24.202

On the other hand, although Data 2 also appear to be heteroscedastic (Figure 6.2), the point estimates are almost the same for OME and WME. However, the standard errors are very different between the two, and the standard errors of PTE are closer to those of WME.

For Data 3, Figure 6.3 shows that the data might be homoscedastic. However, since the point estimates and the standard errors are quite different between the OME and WME, it is important to decide correctly which of the two is optimal. PTE selects WME and the standard errors of the PTE lie between those of OME and WME.

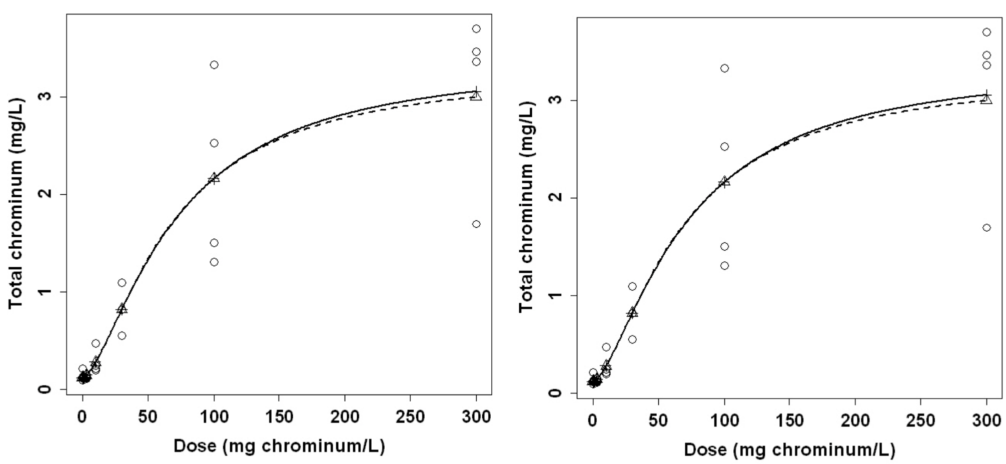
Table 6.2: Estimate and Standard Error for parameters of the models for Chromium data using OLSE, WLSE and PTE methods.

		OLSE		*WLSE*		*PTE*	
		Estimate	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*
Data 1	θ_0	0.134	0.018	0.141	0.011	0.141	0.011
	θ_1	2.620	1.060	1.995	0.500	1.995	0.500
	θ_2	1.244	0.343	1.469	0.198	1.469	0.198
	θ_3	100.736	70.374	65.238	26.293	65.238	26.362
Data 2	θ_0	0.119	0.033	0.120	0.012	0.120	0.012
	θ_1	3.272	0.911	3.166	0.648	3.166	0.649
	θ_2	1.506	0.632	1.542	0.268	1.542	0.270
	θ_3	71.117	41.705	67.704	24.200	67.704	24.275
Data 3	θ_0	0.108	0.009	0.111	0.008	0.111	0.008
	θ_1	0.751	0.113	0.643	0.062	0.643	0.086
	θ_2	0.988	0.174	1.128	0.159	1.128	0.165
	θ_3	95.863	33.665	68.444	14.712	68.444	24.201



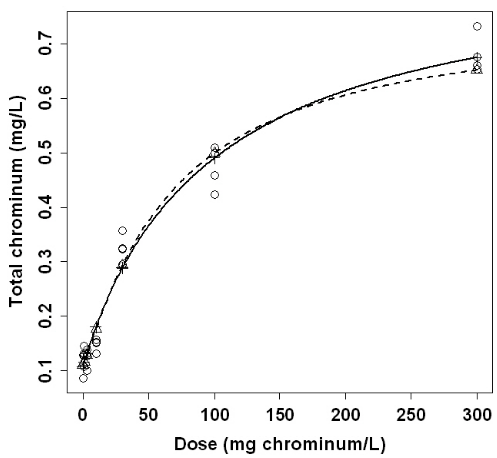
(a) M-estimation (b) LSE

Figure 6.1: Chromium concentration in blood for guinea pig (Data 1) using (a) M-estimation methods: plus sign and solid line (OME); triangles and dashed line (WME); circles (data); (b) LSE methods: plus sign and solid line (OLSE); triangles and dashed line (WLSE); circles (data).

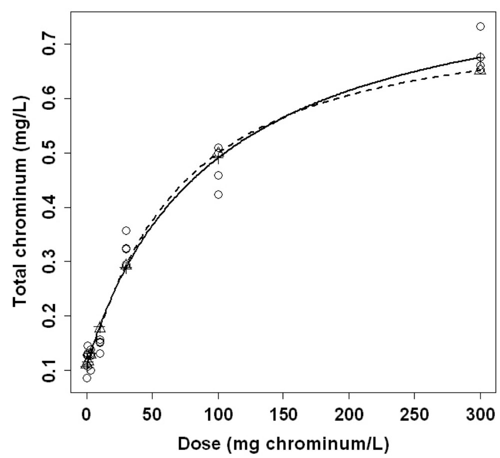


(a) M-estimation (b) LSE

Figure 6.2: Chromium concentration in kidney for guinea pig (Data 2) using (a) M-estimation methods: plus sign and solid line (OME); triangles and dashed line (WME); circles (data); (b) LSE methods: plus sign and solid line (OLSE); triangles and dashed line (WLSE); circles (data).



(a) M-estimation



(b) LSE

Figure 6.3: Chromium concentration in blood for rat (Data 3) using (a) M-estimation methods: plus sign and solid line (OME); triangles and dashed line (WME); circles (data); (b) LSE methods: plus sign and solid line (OLSE); triangles and dashed line (WLSE); circles (data).

Chapter 7

CONCLUDING REMARKS AND FUTURE RESEARCH PLAN

In this dissertation, M-estimation and PTE based methodology has been developed for analyzing nonlinear models that are possibly subject to heteroscedastic variance structure. The methodology proposed here would allow researchers to use estimation procedures that are robust to both potential influential or outlying observations and the error variance structure in nonlinear models. Although the methodology has been developed under the assumption of monotone increasing error variance, it can easily be modified to other patterns of the error variance.

The PTE-based methodology has been proposed by incorporating the M-estimation procedures. Since the M-estimation methods are well-known to be robust against potential outliers and influential observations, we have not evaluated in this dissertation the robustness of the proposed methodology in this aspect. Thus, further work is planned in the immediate future for evaluating the robustness of the methodology by simulation studies using various data sets generated from contaminated normal distribution.

In the weighted M-estimation method, the estimators for the variance parameters are based on the fact that the individual sample variance has the chi-square distribution when the errors are normally distributed. And these estimators are valid when the number of observation at each dose level is at least 2. Motivated by the above observation and some practical applications, I like to explore at least three different extensions to the research presented in my dissertation.

Firstly, in toxicological studies such as HTS assays, there might be data sets where the number of observations at each dose level may be small, even 1. For such data sets the theory and methodology developed in this dissertation may not be appropriate. When the number of observation at each dose level is 1, one natural method to estimate the variance at each dose level is to use the residuals obtained from OME for the regression parameters. However, they are not robust to possible heteroscedasticity. Therefore, further work needs to be done for developing alternative estimation methods for the variance parameters with small samples.

Secondly, since the sample variance is not robust to outliers, one may consider the Gini's mean difference as an alternative for estimating the variance parameters. The Gini's mean difference is defined as the sample mean absolute deviation. Since it is a U-statistic and hence the associated asymptotic theory can be exploited.

Thirdly, the theory and methodology developed in this dissertation rely on the assumption that the errors are normally distributed. Specifically, the normality assumption is required for estimating the variance parameters in the log-linear model. Such an assumption needs not be true in practice. Hence I like to develop suitable alternate methodology that relaxes this assumption.

Another important part of the proposed PTE methodology is to model heteroscedasticity log-linearly. Although we have evaluated from simulation studies the robustness of the proposed method when the model is not valid, it is still possible that the method is not very efficient for some data sets since the model is a first order approximation. Hence, further work is planned in the immediate future for investigating the robustness of the proposed method in various types of data sets. Also, one might want to study the relationship between the complexity of the model for the error variance and the efficiency of the proposed method.

Another extension of this research is to derive PTE-based methodology that accounts for not only variance structure for the error but also model mis-specification. In many practical situations, a researcher may believe that the intercept of the Hill model is zero. However, this may depend upon the bioassay and may not be true for all data sets. The presence or the absence of the intercept parameter may drastically impact the estimation of the remaining parameters of the model. So it would be useful from a practical point of view to develop a PTE-based methodology to select the suitable model and estimate the parameters.

The theory and methodology developed in my dissertation lends itself to various possible extensions. In the near future we would extend the methodology to nonlinear mixed effect models and to physiologically based pharmacokinetic (PBPK) models. PBPK models are mathematical characterizations which describe how a chemical is absorbed, distributed, metabolized and eliminated in various parts of the body, such as blood, liver, kidney, etc. This approach is often used for understanding mechanism of action of various chemicals. Thus they play an important role in drug development by pharmaceuticals and for risk assessment by regulatory agencies such as the EPA and FDA. A PBPK model consists of a system of differential equations. Solution to PBPK models requires estimation of several parameters, some of which are determined using nonlinear regression models and others are obtained from published literature. However, issues such as structure for the error variance in the model and the presence of potential outliers or influential observations, require the development of robust statistical methods for analyzing the PBPK models. Furthermore, since repeated measurements are obtained on some of the compartments of a PBPK model, such as, blood, urine and feces, robust statistical methods for mixed effects models would play an important role in analyzing these data. Also, since a PBPK model consists of several compartments, new methods will be needed for simultaneous estimation of parameters from all the compartments.

REFERENCES

[1] Ahmed, S. E. and Saleh, A. K. Md. E. (1988). Estimation strategy using a preliminary test in some normal models. *Soochow Journal of Mathematics* **14**, 135–165.

[2] Ahmed, S. E. and Krzanowski, W. J. (2004). Biased estimation in a simple multivariate regression model. *Computational Statistics & Data Analysis* **45**, 689–696.

[3] Ahsanullah, M. and Saleh, A. K. Md. E. (1972). Estimation of intercept in a linear regression model with one dependent variable after a preliminary test on the regression coefficient. *International Statistical Review* **40**, 139–145.

[4] Ali, A. M. and Saleh, A. K. Md. E. (1990). Estimation of the mean vector of a multivariate normal distribution under symmetry. *Journal of Statistical Computation and Simulation* **35**, 209–226.

[5] Al-Kandari, N. M., Buhamra, S. S., and Ahmed, S. E. (2005). Estimating and testing effect size from an arbitrary population. *Journal of Statistical Computation and Simulation* **75**, 987–1001.

[6] Arashi, M. and Tabatabaey, S. M. M. (2008). Estimation in multiple regression model with elliptically contoured errors under MLINEX loss. *Journal of Applied Probability and Statistics* **3**, 23–35.

[7] Avalos, M., Mak, C., Randall, P. K., Trzeciakowski, J. P., Abell, C., Kwan, S.-W., and Wilcox, R. E. (2001). Nonlinear analysis of partial dopamine agonist effects on cAMP in C6 glioma cells. *Journal of Pharmacological and Toxicological Methods* **45**, 17–37.

[8] Baklizi, A. (2008). Preliminary test estimation in the Pareto distribution using minmax regret significance levels. *International Mathematical Forum 3*, 473–478.

[9] Baklizi, A. and Ahmed, S. E. (2008). On the estimation of reliability function in a Weibull lifetime distribution. *Statistics* **42**, 351–362.

[10] BANCROFT, T. A. (1944). On biases in estimation due to the use of the preliminary test of significance. *The Annals of Mathematical Statistics* **15**, 190–204.

[11] BANCROFT, T. A. (1964). Analysis and inference for incompletely specified models involving the use preliminary test(s) of significance. *Biometrics* **20**, 427–444.

[12] BANCROFT, T. A. (1972). Som recent advances in inference procedures using preliminary tests of significance. In *Statistical Papers* in honour of G. W. Snedecor, pp. 19–30, Iowa State University Press.

[13] BARATA, C., BAIRD, D. J., NOGUEIRA, A. J. A., SOARES, A. M. V. M., AND RIVA, M. C. (2006). Toxicity of binary mixtures of metals and pyrethroid insecticides to Daphnia magna Straus. Implications for multi-substance risks assessment. *Aquatic Toxicology* **78**, 1–14.

[14] BHOJ, D. S. AND AHSANULLAH, M. (1993). Estimation of conditional mean in a linear regression model. *Biometrical Journal* **35**, 791–799.

[15] BHOJ, D. S. AND AHSANULLAH, M. (1994). Estimation of conditional mean in a regression model after a preliminary test on regression coefficient. *Biometrical Journal* **36**, 153–163.

[16] CARAPETO, M. AND HOLT, W. (2003). Testing for heteroscedasticity in regression models. *Journal of Applied Statistics* **30**, 13–20.

[17] CARROLL, R. J. (1982). Adapting for heteroscedasticity in linear models. *The Annals of Statistics* **10**, 1223–1233.

[18] CARROLL, R. J. AND RUPPERT, D. (1982). Robust estimation in heteroscedastic linear models. *The Annals of Statistics* **10**, 429–441.

[19] CARROLL, R. J. AND RUPPERT, D. (1988). *Transformation and Weighting in Regression.* New York: Chapman and Hall.

[20] CARROLL, R. J., WU, C. F. J., AND RUPPERT, D. (1988). The effect of estimating weights in weighted least squares. *Journal of the American Statistical Association* **83**, 1045–1054.

[21] CHAUBEY, Y. P. AND SEN, D. (2004). An investigation into properties of an estimator of mean of an inverse Gaussian population. *Technical Report No. 7/04.* Concordia University, Canada.

[22] CROFTON, K. M., PAUL, K. B., DEVITO, M. J., AND HEDGE, J. M. (2007). Short-term *in vivo* exposure to the water contaminant triclosan: evidence for disruption of thyroxine. *Environmental Toxicology and Pharmacology* **24**, 194–197.

[23] Davidian, M. and Carroll, R. J. (1987). Variance Function Estimation. *Journal of the American Statistical Association* **82**, 1079–1091.

[24] Davidian, M. and Giltinan, D. M. (1993). Some simple methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics* **49**, 59–73.

[25] Dette, H. and Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *Journal of the Royal Staistical Society. Series B* **60**, 693–708.

[26] Gallant, A. R. (1987). *Nonlinear Statistical Models.* New York: Wiley.

[27] Gaylor, D. W. and Aylward, L. L. (2004). An evaluation of benchmark dose methodology for non-cancer continuous-data health effects in animals due to exposures to dioxin (TCDD). *Regulatory Toxicology and Pharmacology* **40**, 9–17.

[28] Hampel, F. R., Rousseeuw, P. J., Ronchetti, E., and Stabel, W. (1986). *Robust Statistics – The Approach Based on Influence Functions.* New York: John Wiley & Sons.

[29] Han, C. P. and Bancroft, T. A. (1968). On pooling means when variance is unkown. *Journal of the American Stastical Association* **29**, 21–34.

[30] Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology* **40**(Suppl), iv–vii.

[31] Hoferkamp, C. (1999). *Anaysis of Fixed and Mixed Effects Linear Models under Heteroscedasticity.* Ph.D. dissertation, Division of Statistics, University of Virginia.

[32] Hoferkamp, C. and Peddada, S. D. (2001). Test of homogeneity of variances against ordered alternatives in fixed effects linear models. *Sankhya Series B* **48**, 354–371.

[33] Hoferkamp, C. and Peddada, S. D. (2002). Parameter estimation in linear models with heteroscedastic variances subject to order restrictions. *Journal of Multivariate Analysis* **82**, 65–87.

[34] Huber, P. J. (1981). *Robust Statistics.* New York: John Wiley & Sons.

[35] Jacquez, J. A., Mather, F. J., and Crawford, C. R. (1968). Linear regression with nonconstant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24**, 607–626.

[36] Judge, G. G. and Bock, M. E. (1978). *The Statistical Implications of Pretest and Stein-rule Estimators in Econometrics.* Amesterdam: North-Holland.

[37] Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures, Asymptotics and Interrelations.* New York: John Wiley & Sons.

[38] Kabir, Md. E. and Khan, S. (2008). Estimation of slope in a simple regression model using p-value. USQ Working Paper Series SC-MC-0709.

[39] Khan, S. (2000). Improved estimation of the mean vector for Student-t model. *Communications in Statistics – Theory and Methods* **29**, 507–527.

[40] Khan, S. (2005). Estimation of parameters of the simple multivariate linear model with student-t errors. *Journal of Statistical Research* **39**, 79–94.

[41] Khan, S. (2008). Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes. *Communications in Statistics – Theory and Methos* **37**, 247–260.

[42] Khan, S., Hoque, Z., and Saleh, A. K. Md. E. (2005). Estimation of the intercept parameter for linear regression model with uncertain non-sample prior information. *Statistical Papers* **46**, 379–395.

[43] Khan, S. and Saleh, A. K. Md. E. (1995). Preliminary test estimators of the mean based on p-samples from multivariate Student-t populations. *Bull. of the Int. Stat. Ins., 50th Session of the ISI, Beijing*, 599–600.

[44] Khan, S. and Saleh, A. K. Md. E. (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student's t-errors. *Biometrical Journal* **39**, 131–147.

[45] Khan, S. and Saleh, A. K. Md. E. (2008). Estimation of slope for linear regression model with uncertain prior information and student-t error. *Communications in Statics – Theory and Methods* **37**, 2564–2581.

[46] Kibria, B. M. G. (1996). On preliminary test ridge regression estimator for the restriced linear model with non-normal disturbances. *Communications in Statistics – Theory and Methods* **25**, 2349–2369.

[47] Kim, H. M. and Saleh, A. K. Md. E. (2005). Improved estimation of regression parameters in measurement error models. *Journal of Multivariate Analysis* **95**, 273–300.

[48] Kitagawa, T. (1963). Estimation after preliminary tests of significance. *University of California Publications of Statistics* **3**, 147–186.

[49] Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*, 5th edition. New York: McGraw-Hill/Irwin.

[50] Lin, J.-G. and Wei, B.-C. (2003). Testing heteroscedasticity in nonlinear regression models. *Communications in Statistics - Theory and Methods* **32**, 171–192.

[51] Mahdi, T. N., Ahmad, S. E., and Ahsanullah, M. (1998). Improved predictions: pooling two identical regression lines. *Journal of Applied Statistical Science* **7**, 63–86.

[52] Menéndez, M. L., Pardo, L., and Pardo, M. C. (2008). Preliminary test estimators and phi-divergence measures in generalized linear models with binary data. *Journal of Multivariate Analysis* **99**, 2265–2284.

[53] Menéndez, M. L., Pardo, L., and Pardo, M. C. (2008). Confidence sets and coverage probabilities based on preliminary estimators in logistic regression models. *Journal of Computational and Applied Mathematics* **224**, 193–203.

[54] Morris, J. B., Symanowicz, P., and Sarangapani, R. (2002). Regional distribution and kinetics of vinyl acetate hydrolysis in the oral cavity of the rat and mouse. *Toxicology Letters* **126**, 31–39.

[55] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**(4), 308–313.

[56] Nitcheva, D. K., Piegorsch, W. W., West, R. W., and Kodell, R. L. (2005). Multiplicity-adjusted inferences in risk assessment with non-quantal data. *Journal of Biopharmaceutical Statistics* **15**, 17–31.

[57] NTP. (2007). NTP Toxicity Studies of Sodium Dichromate Dihydrate (CAS No. 7789-12-0) Administered in Drinking Water to Male and Female F344/N Rats and B6C3F1 Mice and Male BALB/c and am3-C57BL/6 Mice. *Toxicity Report Series* **72**, 1–G4, National Toxicology Program, NC.

[58] Piegorsch, W. W. and West, R. W. (2005). Benchmark analysis: Shopping with proper confidence. *Risk Analysis* **25**, 913–920.

[59] Pounds, J. G., Haider, J., Chen, D. G., and Mumtaz, M. (2004). Interactive toxicity of simple chemical mixtures of cadmium, mercury, methylmercury and trimethyltin: model-dependent responses. *Environmental Toxicology and Pharmacology* **18**, 101–113. 2nd Ed., New York: Cambridge University Press.

[60] Saleh, A. K. Md. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications.* New York: Wiley.

[61] Saleh, A. K. Md. E. and, Sen, P. K. (1978). Nonparametric estimation of location parameter after a preliminary test on regression. *The Annals of Statistics* **6**, 154–168.

[62] Saleh, A. K. Md. E. and, Sen, P. K. (1985). On some shrinkage estimators of multivariate location. *The Annals of Statistics* **13**, 272–281.

[63] Saleh, A. K. Md. E. and Han, C. P. (1990). Shrinkage estimation in regression analysis. *Estadistica* **42**, 40–63.

[64] Sand, S., von Rosen, D., Eriksson, P, Fredriksson, A, Viberg, H., Victorin, K., and Filipsson, A. F. (2004). Dose-response modeling and benchmark calculations from spontaneous behavior data on mice neonatally exposed to 2,2′,4,4′,5-pentabromodiphenyl ether. *Toxicological Sciences* **81**, 491–501.

[65] Sanhueza, A. I. (2000). *Robust M-procedures in Nonlinear Regression Models.* PhD dissertation, University of North Carolina at Chapel Hill.

[66] Sanhueza, A. I. and Sen, P. K. (2001). M-methods in generalized nonlinear models. In *Probability and Statistical Models with Applications.* Charalambides, Ch. A., Koutras, M. V., and Balakrishnan, N., Eds., pp. 359–375. New York: Chapman & Hall/CRC.

[67] Sanhueza, A. I. and Sen, P. K. (2004). Robust M-procedures in univariate nonlinear regression models. *Brazilian Journal of Probability and Statistics* **18**, 183–200.

[68] Sanhueza, A. I., Sen, P. K., and Leiva, V. (2009). A robust procedure in nonlinear for repeated measurements. *Communications in Statistics – Theory and Methods* **38**, 138–155.

[69] Seber, G. A. F., and Wild, C. H. (1989). *Nonlinear Regression.* New York: Wiley.

[70] Sen, P. K. (1986). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhya: The Indian Journal of Statistics Series A* **48**, 354–371.

[71] Shao, J. (1992). Consistency of lest-squares estimator and its jackknife variance estimator in nonlinear model. *Canadian Journal of Statistics* **20**, 415–428.

[72] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* New York: Springer-Verlag.

[73] Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions.* New York: John Wiley and Sons.

[74] Swann, W. H. (1972). Direct search methods, In *Numerical Methods for Unconstrained Optimization.* W. Murray, Ed., pp. 13–28. London: Academic Press.

[75] Tabatabaey, S. M. (1995). *Preliminary Test Approach Estimation: Regression Model with Spherically Symmetric Errors.* PhD dissertation, Carleton University, Ottawa.

[76] Tabatabaey, S. M., Saleh, A. K. Md. E., and Kibria, B. M. G. (2004a). Estimation strategies for parameters of the linear regression model with spherically symmetric distributions. *J. Statist. Res.* **38**, 13–31.

[77] Tabatabaey, S. M., Saleh, A. K. Md. E., and Kibria, B. M. G. (2004b). Simultaneous estimation of regression parameters with spherically symmetric errors under possible stochastic constrains. *Int. J. Statist. Sci.* **3**, 1–20.

[78] Velarde, G., Ait-aissa, S., Gillet, C., Rogerieux, F., Lambre, C., Vindimian, E., and Porcher, J. M. (1999). Use of transepithelial electrical resistance in the study of pentachlorophenol toxicity. *Toxicology in Vitro* **13**, 723–727.

[79] Wan, A. T. K. and Zou, G. (2003). Optimal critical values of pre-tests when estimating the regression error variance: Analytical findings under a general loss structure. *Journal of Econometrics* **114**, 165–196.

[80] Wan, A. T. K., Zou, G., and Ohtani, K. (2006). Further results on optimal critical values of pre-test when estimating the regression error variance. *Econometrics Journal* **9**, 159–176.

[81] Wilcox, R. R. and Keselman, H. J. (2004). Robust regression methods: achieving small standard errors when there is heteroscedasticity. *Understanding Statistics* **3**(4), 349–364.

[82] Wu, Y., Piegorsch, W. W., West, W., Tang, D., Petkewich, M. O., and Pan, W. (2006). Multiplicity-adjusted inferences in risk assessment: benchmark analysis with continuous response data. *Environmental and Ecological Statistics* **13**, 125–141.

[83] You, J. and Chen, G. (2005). Testing heteroscedasticity in partially linear regression models. *Statistics & Probability Letters* **73**, 61–70.

[84] Yunus, R. M. and Khan, S. (2007). Increasing power of the test through pre-test – a robust method. *USQ Working Paper Series SC-MC-0713.*