ABSTRACT

Title of Dissertation: Nonparametric Quasi-likelihood in Longitudinal Data Analysis

Xiaoping Jiang, Doctor of Philosophy, 2004

Dissertation directed by: Professor Paul J. Smith
Statistics Program
Department of Mathematics

This dissertation proposes a nonparametric quasi-likelihood approach to estimate regression coefficients in the class of generalized linear regression models for longitudinal data analysis, where the covariance matrices of the longitudinal data are totally unknown but are smooth functions of means. This proposed nonparametric quasi-likelihood approach is to replace the unknown covariance matrix with a nonparametric estimator in the quasi-likelihood estimating equations, which are used to estimate the regression coefficients for longitudinal data analysis. Local polynomial regression techniques are used to get the nonparametric estimator of the unknown covariance matrices in the proposed nonparametric quasi-likelihood approach. Rates of convergence of the resulting estimators are established. It is shown that the nonparametric quasi-likelihood estimator is not only consistent but also has the same asymptotic distribution

as the quasi-likelihood estimator obtained with the true covariance matrix. The results from simulation studies show that the performance of the nonparametric quasi-likelihood estimator is comparable to other methods with given marginal variance functions and correctly specified correlation structures. Moreover, the results of the simulation studies show that nonparametric quasi-likelihood corrects some shortcomings of Liang and Zeger's GEE approach in longitudinal data analysis.

Nonparametric Quasi-likelihood in

Longitudinal Data Analysis

by

Xiaoping Jiang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor Paul J. Smith, Chairman/Advisor
Professor Jeffery Cooper
Professor Benjamin Kedem
Professor Grace Yang
Professor Mary Christman

DEDICATION

To my husband, my daughter and my son.

ACKNOWLEDGEMENTS

I would like to thank Dr. Paul Smith, my advisor, for his great effort and valuable guidance to help me finish my dissertation. I would like to thank Drs. Christman, Cooper, Kedem and Yang for their willingness to serve in my dissertation committee. In addition, I would like to thank my husband Zhenshao for his great support.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

Nelder and Wedderburn introduced Generalized Linear Models (GLM) in 1972. Let $\mathbf{Y} = (y_1, y_2, \ldots, y_n)$ be a vector of observations, assumed to be a realization of a random variable $\mathbf{Y}$ and independently distributed with mean vector $\boldsymbol{\mu}$. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ be the $n \times 1$ vectors of covariates. GLM consists of three components:

(a) The random component: Each component of $\mathbf{Y}$ independently has a exponential family distribution.

(b) The systematic component: Covariates $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$ produce a linear predictor $\boldsymbol{\eta}$ given by

$$\boldsymbol{\eta}_k = \sum_1^p \mathbf{x}_{kj}\beta_j$$

where $\beta_1, \beta_2, \ldots, \beta_p$ are unknown parameters.

(c) The link between the random and systematic components: There is a function g called the link function which relates the linear predictor vector $\boldsymbol{\eta}$ and the expected value $\boldsymbol{\mu}$ of $\mathbf{Y}$, such that

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}).$$

GLM extends linear models from the Gaussian case to a broad class of outcomes. Because the distribution of observations can be specified as an exponential family, one can construct a likelihood function, and therefore maximum likelihood estimation is the principal method of estimation used for all generalized linear models. The Gauss-Newton method is a well-known algorithm for calculating maximum likelihood estimates for GLM. This method produces maximum likelihood estimates by iterative weighted least squares.

Maximum likelihood estimation (MLE) has many good analytical properties. For example, the estimators are consistent, asymptotically normal and asymptotically efficient under mild regularity conditions (McCullagh and Nelder, 1983). However, the full distributions of observations have to be specified in order to define a likelihood function.

Unfortunately, it is unclear how to specify the full distribution in many practical situations. Wedderburn (1974) proposed an important extension of likelihood function, the quasi-likelihood function, for the situations where there is insufficient information to construct a likelihood function.

Suppose that we have independent observations z_i $(i = 1, 2, \ldots, n)$ with expectations μ_i and variances $\mathrm{Var}\,(z_i) \propto V(\mu_i)$, where μ_i is some known function of a set of parameters $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_p\}$ and $V(\cdot)$ is some known function. The quasi-likelihood function (in fact log quasi-likelihood function), is a function $\mathcal{K} = \sum_{i=1}^{n} \mathcal{K}_i$ such that

$$\frac{\partial \mathcal{K}_i\,(z_i, \mu_i)}{\partial \mu_i} = \frac{z_i - \mu_i}{V(\mu_i)}. \tag{1.1}$$

Then $\mathcal{K}$ has statistical properties similar to those of a log-likelihood function. For example, the expectation of the derivative of $\mathcal{K}$ with respect to μ equals 0; the expectation of the derivative of $\mathcal{K}$ with respect to β_i equals 0 and the expectation of the square of the derivative of $\mathcal{K}$ with respect to the mean μ equals the negative expectation of second derivative of $\mathcal{K}$ with respect to mean μ, which is the reciprocal of the variance function.

Wedderburn's introduction of quasi-likelihood greatly widened the scope of generalized linear models by allowing the full distributional assumption about the random component in the model to be replaced by a much weaker assumption in which only the mean and a relation between the mean and the variance (variance function) of observations need to be specified. A quasi-likelihood function then can be used for estimation in the same way as a likelihood function for generalized linear models. When certain mean-variance relationships are specified, the quasi-likelihood function sometimes turns out to be a recognizable likelihood function. For example, according to Wedderburn (1974), for a constant coefficient of variation the quasi-likelihood function is the same as the likelihood function obtained by treating the observations as if they have a gamma distribution. Wedderburn showed that the log likelihood function is identical to the log quasi-likelihood if and only if this family of distributions is a one-parameter exponential family.

Wedderburn's original quasi-likelihood model required knowing the variance function up to a multiplicative constant. Since the variance function is an essential determinant of the quasi-likelihood, its specification is an important problem in the quasi-likelihood approach. In many applications, it is a priori unclear how the variance function should be specified. There are parametric and nonparametric quasi-likelihood functions based on the methods of specification of the unknown variance function.

Nelder and Pregibon (1987) proposed an extended parametric quasi-likelihood function which replaces the unknown variance function by a family of functions indexed by an unknown parameter. They embedded the variance function into a family of functions indexed by an unknown parameter θ, so that

$$\mathrm{Var}\,(z_i) = \phi V_\theta(\mu_i)\ .$$

A useful family is obtained by considering powers of μ:

$$V_\theta(\mu) = \mu^\theta\ . \tag{1.2}$$

Most common values of θ in (1.2) are the values $0, 1, 2, 3$ which correspond to variance functions associated with normal, Poisson, Gamma, and Inverse Gaussian distributions respectively. It has been shown that an exponential family with variance function

$$V_\theta(\mu) = \mu^\theta$$

exists for $\theta = 0$ and $\theta \geq 1$. Another parametric approach is the pseudo-likelihood method introduced by Carroll and Ruppert(1982).

A nonparametric quasi-likelihood approach was proposed by Chiou and Müller (1999), who extended the quasi-likelihood approach to situations where the variance functions are unknown but can be assumed to be smooth. Their nonparametric quasi-likelihood function is obtained by substituting a nonparametrically estimated variance function in the place of the unknown true variance function in the usual definition of the quasi-likelihood function (1.1). The nonparametric variance function estimate which is used in the nonparametric quasi-likelihood is obtained by smoothing squared residuals obtained from a preliminary model fit. This approach consists of a two-stage iterative estimating procedure. The regression parameters are first estimated by assuming $V(\mu) = 1$ to obtain GLM parameter estimates $\widehat{\boldsymbol{\beta}}_0$. Then a variance function is estimated nonparametrically, treating the regression parameters as known to be $\widehat{\boldsymbol{\beta}}_0$. This procedure is iterated by using the updated model parameters in order to obtain new residuals and estimated means and thus an updated nonparametric variance function estimate, which then in turn can be used to obtain improved parameter estimates. They showed that the asymptotic distribution of the nonparametric quasi-likelihood estimator is the same as that of quasi-likelihood estimator under known variance function,assuming that the unknown variance function is replaced by a consistent nonparametric variance function estimates.

In Chiou and Müller's nonparametric quasi-likelihood approach, they chose local polynomial fitting regression by locally weighted least squares as their smoothing method. Local polynomial regression was systematically studied by Stone (1977, 1980, 1982) and Cleveland (1979). Cleveland (1979) introduced local weighted polynomial regression using LOcally WEighted Scatterplot Smoothing (Lowess) and Cleveland (1988) extended Lowess to multivariate settings. Lowess is one of several nonparametric regression methods that can be used to estimate the mean response profile as a function of some covariates. Fan (1992, 1993), Fan and Gijbels (1992), and Ruppert and Wand (1994) published papers detailing the advantages of local polynomial fitting. The book of Fan and Gijbels (1996) gave a thorough study of local polynomial regression. There is extensive literature on nonparametric variance function estimation. Carroll (1982) developed kernel estimators in the context of linear regression. Müller and Stadtmüller (1987) and Hall and Carroll (1989) proposed and analyzed kernel-type variance function estimators by assuming a nonparametric mean function. Fan and Gijbels (1995) proposed a type of local polynomial variance function estimator as part of their bandwidth selection procedure. Ruppert and Wand (1997) had some results about local polynomial smoothers by using linear smoothing of squared residuals in estimation of variance functions under the assumption that both mean and variance functions are smooth, but neither is assumed to be in a parametric family. Wedderburn's Quasi-likelihood, Nelder and Pregibon's extended parametric quasi-likelihood, Carroll and Ruppert's pseudo-likelihood method and Chiou and Müller's nonparametric quasi-likelihood function are useful for independent observations.

McCullagh (1983) extended Wedderburn's quasi-likelihood to multivariate settings. Given the vector of random variables $\mathbf{Y}$ with length N mean vector $\boldsymbol{\mu}$ and covariance matrix $\sigma^2\mathbf{V}(\boldsymbol{\mu})$, the log quasi-likelihood $\mathcal{L}$, a function of $\boldsymbol{\mu}$, will be given by the system of partial differential equations

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}} = \mathbf{V}^{-1}(\boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu}).$$

According to McCullagh (1989), the statistical properties of quasi-likelihood functions, in terms of score function, estimator of regression parameters $\boldsymbol{\beta}$ and the distribution of the quasi-likelihood-ratio statistic, are very similar to those of ordinary likelihood functions except that the nuisance parameter, σ^2, when it is unknown, is treated separately from $\boldsymbol{\beta}$ and is not estimated by weighted least squares. The quasi-likelihood score function

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \mathbf{D}^T\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$$

has zero mean and covariance matrix

$$\sigma^2 i_{\boldsymbol{\beta}} = \sigma^2 \mathbf{D}^T\mathbf{V}^{-1}\mathbf{D}$$

where $-i_{\boldsymbol{\beta}}$ is the expected derivative matrix of the log quasi-likelihood function $\mathcal{L}(\mathbf{Y}; \boldsymbol{\mu})$. Under some weak conditions on the third derivative of the link function and assuming that $N^{-1}i_{\boldsymbol{\beta}}$ has a positive definite limit and that the third moments of $\mathbf{Y}$ are finite, the $\sqrt{n}$-consistent quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$ asymptotically follows a normal distribution with mean $\boldsymbol{\beta}$ and covariance matrix $\left(N\sigma^2 i_{\boldsymbol{\beta}}\right)^{-1}$. The quasi-likelihood approach is very useful in many situations by only using first and second moment assumptions to avoid the complete specification of underlying distribution of the observations. For various analyses of independent observations, generalized linear models (McCullagh and Nelder, 1983) and Quasi-likelihood (Wedderburn, 1974) have recently unified regression methods for a variety of discrete and continuous variables.

There are many situations where the dependence relationships among the data are so significant that we can not ignore them. Longitudinal data are one example of dependent data. Longitudinal data consist of repeated measurements through time for each subject, and these repeated measurements are correlated or exhibit variability that changes. They can be collected either prospectively (such as clinical trial data), following subjects forward in time, or retrospectively, by extracting multiple measurements on each subject from historical records. The main interest in a longitudinal study is to determine the dependence of the outcome variable on covariates, such as the dependence of the clinical outcome on the treatment and other factors in clinical study. Since longitudinal data are characterized by the fact that repeated measurements made on the same subject are usually intercorrelated, the statistical analysis of longitudinal data requires special methods to take the correlation structure into account to increase the efficiency of estimators.

Liang and Zeger (1986) applied the quasi-likelihood approach to longitudinal data analysis and proposed the generalized estimated equations (GEE) approach, which is very useful for longitudinal data analysis. Suppose that there is a longitudinal data set $\{y_{kj}, \mathbf{x}_{kj}\}$ with mean $\mu_{kj} = E(y_{kj})$, and $g\left(\mu_{kj}\right) = x_{kj}^T \boldsymbol{\beta}$ for k-th subject at time point t_{kj}, $j = 1, 2, \ldots, T_k$ and subjects $k = 1, 2, \ldots, n$. Here y_{kj} is the response variable and $\mathbf{x_{kj}}$ is a $p \times 1$ vector of covariates at time point t_{kj}. Let $\mathbf{Y_k}$ be the $T_k \times 1$ vector $(y_{k1}, \ldots, y_{kT_k})$, with mean vector $\boldsymbol{\mu} = (\mu_{k1}, \ldots, \mu_{kT_k})$ and and covariance matrix $\boldsymbol{\Sigma}_k = \mathrm{Cov}(\mathbf{Y}_k)$, and let $\mathbf{X}_k$ be the $T_k \times p$ matrix $(\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kT_k})^T$ for the k-th subject. Assuming that the form of the first two marginal moments $E(y_{kj})$ and $\mathbf{A}_k = \mathrm{diag}\left[\mathrm{Var}\left(y_{k1}\right), \mathrm{Var}\left(y_{k2}\right), \ldots, \mathrm{Var}\left(y_{kT_k}\right)\right]$ are known, Liang and Zeger's (1986) GEE approach used a working correlation matrix $R(\alpha)$, which is assumed to be a matrix dependent on a parameter α, to replace the covariance matrix $\mathbf{V}_k^{-1} = A_k^{1/2} R(\alpha) A_k^{1/2}$ in the following general estimating equations,

$$\sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}_k^{-1} \mathbf{S_k} = 0 \tag{1.3}$$

where the matrix $\mathbf{D}_k = \partial \mu_k / \partial \boldsymbol{\beta}$ and the vector $\mathbf{S}_k = \mathbf{Y}_k - \boldsymbol{\mu}_k$. Given α, one solves the equation (1.3) to obtain consistent estimators of regression parameters $\boldsymbol{\beta}$ in the class of generalized linear models for repeated measures data. Liang and Zeger also prove that the same results hold if α is replaced by $\widehat{\alpha}$, a quantity estimated from the data. In fact, the estimating equation (1.3) is the quasi-score equation derived from McCullagh (1983), as Liang and Zeger pointed out (1986). In terms of the method of specification of correlation structure, Liang and Zeger's GEE is an very important parametric approach to longitudinal data analysis, provided one knows the marginal mean and variance functions. However, the GEE estimators will be less efficient than the quasi-likelihood estimator when the working correlation matrix is misspecified, even though they are still consistent (Liang and Zeger (1986)). According to Crowder (1995), there may not even exist any solution for $\widehat{\alpha}$ for various possible reasons, so that the uncertainty of definition of the working correlation matrix can lead to a complete breakdown of the estimation of regression parameters in some cases. Sutradhar and Das (1999) show that even though the Liang-Zeger approach in many dependant data situations yields consistent estimators for the regression parameters, in some cases, these estimators are inefficient as compared to the regression estimators obtained by using the independence estimating equation approach. Wang and Carey (2001) provided two approaches to supplement and enhance GEE by constructing unbiased estimating equations from general correlation models for irregularly timed repeat measures.

Besides GEE, semiparametric regression modeling is also useful for longitudinal data analysis. Fan and Li (2004) proposed two new approaches for estimating the regression coefficients in the following semiparametric model for longitudinal data analysis:

$$y(t) = \alpha(t) + \boldsymbol{\beta}^T \mathbf{x}(t) + \epsilon(t),$$

where $y(t)$ is the response variable and $\mathbf{x}$ is a covariate vector at time t, $\alpha(t)$ is an unspecified baseline function of t, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, and $\epsilon(t)$ is a mean-0 stochastic process. Fan and Li used local polynomial regression to estimate the baseline function $\alpha(t)$, given a so called difference-based estimator (DBE) of $\boldsymbol{\beta}$.

Similar to Wedderburn's quasi-likelihood approach, Liang and Zeger's GEE requires knowing marginal variance functions. It is unclear how to specify both the marginal variance function and the correlation structure in some longitudinal studies. Nonparametric procedures let the data speak for themselves, instead of picking one matrix arbitrarily as a working correlation matrix when we have no idea about the data correlation structure. Modern computer technology makes nonparametric techniques much more feasible than they used to be since it is much easier to perform extensive computation on datasets.

A nonparametric quasi-likelihood approach will be proposed in this dissertation to estimate parameters in the class of generalized linear regression models for longitudinal data analysis where the covariance structures (i.e. marginal variance functions and correlation structure) are unknown. This proposed extended nonparametric quasi-likelihood approach is to estimate regression model parameters $\boldsymbol{\beta}$ in the class of generalized linear models for longitudinal data analysis where the covariance matrix is totally unknown but its elements are smooth functions of the means. Since this proposed nonparametric quasi-likelihood approach can be used for longitudinal data (dependent data) analysis and does not need to specify the marginal variance functions, it is a multivariate extension of Chiou and Müller's nonparametric quasi-likelihood approach and also is a generalization of Liang and Zeger's GEE. The proposed nonparametric quasi-likelihood approach for longitudinal data consists of following two major procedures.

First Procedure: Initially set a value as the initial estimate $\widehat{\boldsymbol{\beta}}_0$ for the regression parameters. Obtain the nonparametric estimator of the covariance matrix by smoothing squares of residuals and cross terms of residuals generated from the previous model fit (obtained by substituting $\widehat{\boldsymbol{\beta}}_0$ into the quasi-likelihood model).

Second Procedure: Obtain the nonparametric quasi-likelihood function by replacing the unknown true covariance matrix in quasi-likelihood function score equation with the nonparametric estimator of the covariance matrix obtained from the first procedure, and solve this nonparametric quasi-likelihood score equation to obtain the updated estimator of model parameters.

The proposed nonparametric quasi-likelihood approach is achieved by iterating those two procedures until a convergence criterion is satisfied. The updated nonparametric estimator of the covariance matrix is obtained by smoothing residuals generated from the previous model fit. The quasi-likelihood estimator for the parameter in the generalized linear model is in turn updated by solving the nonparametric quasi-likelihood score equation with the updated nonparametric estimator of the covariance matrix. In this proposed extended nonparametric quasi-likelihood approach, local polynomial smoothing in multivariate settings by locally weighted least squares is chosen as the smoothing method. Some definitions and properties of local polynomial smoothers will be discussed in Chapter 2. The quasi-likelihood functions with true covariance matrix and unknown covariance matrix for longitudinal data will be introduced in Chapter 3.

In Chapter 4, the model assumptions of nonparametric quasi-likelihood for longitudinal data will be introduced and the nonparametric estimator for unknown covariance matrix will be defined.

The consistency and the rate of convergence of the nonparametric estimator of covariance matrix will be established in Chapter 5. The asymptotic properties of the nonparametric quasi-likelihood estimator of $\boldsymbol{\beta}$ will be established under certain regularity conditions in Chapter 5. It will be shown that when the unknown covariance matrix is replaced with the consistent nonparametric covariance matrix estimate, the $\sqrt{n}$-consistency and the asymptotic normality properties of the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$ of the regression parameter $\boldsymbol{\beta}$ are the same as those for the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ obtained from the quasi-likelihood score equation with the true covariance matrix.

Finite sample behaviors are examined by simulation in Chaper 6. All proofs and auxiliary results will be compiled in Chapter 7. Some conclusions and future research will be discussed in Chapter 8.

Chapter 2

Local Polynomial Regression

2.1 Local Polynomial Regression with a Univariate Explanatory Variable

Consider the bivariate data $(X_1, Y_1), \ldots, (X_n, Y_n)$, which form an i.i.d sample from a population (X, Y). We would like to estimate the regression function $m(x_0) = E(Y|X = x_0)$ and its derivates $m'(x_0), m''(x_0), \cdots, m^{(p)}(x_0)$.

Suppose that the data satisfy the following model:

$$Y = m(X) + \sigma(X)\epsilon$$

where $E(\epsilon) = 0, \mathrm{Var}(\epsilon) = 1$, and X and ϵ are independent. Assume that the $(p+1)^{th}$ derivative of $m(x)$ at the point x_0 exists. We approximate the unknown regression function $m(x)$ locally by a polynomial of order p. A Taylor expansion gives, for x in a neighborhood of x_0,

$$m(x) \approx m(x_0) + m'(x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p. \quad (2.1)$$

This polynomial is fitted locally at x_0 by weighted least squares regression: minimize

$$\sum_{k=1}^{n}\left\{Y_k-\sum_{j=0}^{p}\beta_j(X_k-x_0)^j\right\}^2 K_h(X_k-x_0) \qquad (2.2)$$

where h is a bandwidth, K is a kernel function, and $K_h(\cdot)=K(\cdot/h)/h$ assigns weights to each data point.

Denote by $\widehat{\beta}_j, j=0,1,\cdots,p$, the solution to the least squares problem (2.2). From the Taylor expansion in (2.1) one sees that $\widehat{m}_\nu(x_0)=\nu!\widehat{\beta}_\nu$ is an estimator for $m^{(\nu)}(x_0),\ \ \nu=0,1,\cdots,p.$

Let $\mathbf{X}$ be the design matrix of problem (2.2):

$$\begin{pmatrix} 1 & (X_1-x_0) & \dots & (X_1-x_0)^p \\ 1 & (X_2-x_0) & \dots & (X_2-x_0)^p \\ \vdots & \vdots & \dots & \vdots \\ 1 & (X_n-x_0) & \dots & (X_n-x_0)^p \end{pmatrix},$$

and let

$$\mathbf{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad \boldsymbol{\beta}=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}.$$

Also, let $\mathbf{W}=\operatorname{diag}\{K_h(X_k-x_0)\}_{k=1}^{n}$. Then the weighted least squares problem (2.2) can be written as:

$$\widehat{\boldsymbol{\beta}}=\arg\min_{\beta}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).$$

with $\boldsymbol{\beta}=(\beta_0,\dots,\beta_p)^T$. The solution vector is provided by the weighted least squares method and is given by

$$\widehat{\boldsymbol{\beta}}=(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}.$$

Take $p = 1$. Then we have the locally weighted linear regression estimator of $m(x)$, $\widehat{m}(x_0)$.

2.2 Local Linear Regression in the Multivariate Setting

Given d-dimensional covariates $\mathbf{X}$ and a response variable Y, we want to estimate the mean regression function

$$m(\mathbf{x}) = E(Y|\mathbf{X} = x).$$

Let K be a d-variate nonnegative kernel function. For simplicity, we assume that K is a multivariate probability density function, such that (a) $\int K(\mathbf{u})d\mathbf{u} = 1$ and $\int \mathbf{u}K(\mathbf{u})d\mathbf{u} = 0$; (b) K has compact support and (c)

$$\int u_k u_j K(\mathbf{u})d\mathbf{u} = \delta_{kj} m_2(K),$$

with $m_2(K) = \int u_k^2 K(\mathbf{u})d\mathbf{u} \geq 0$. and δ_{kj} is the Kronecker delta.

Define $K_B(\mathbf{u}) = |\mathbf{B}|^{-1} K(\mathbf{B}^{-1}\mathbf{u})$, where $\mathbf{B}$ is a nonsingular $d \times d$ matrix, the bandwidth matrix, and $|\mathbf{B}|$ denotes its determinant.

Suppose that there are observations $\left\{ \left(\mathbf{X}_k^T, Y_k\right) : \quad k = 1, 2, \cdots, n \right\}$, with vector $\mathbf{X}_k = (X_{k1}, \cdots, X_{kd})^T$. Let $\mathbf{x} = (x_1, x_2, \cdots, x_d)$ be a point in $\mathcal{R}^d$. Using a local linear approximation (take $p = 1$ in problem (2.2)), we have the multivariate version of the weighted least squares problem (2.2): minimize

$$\sum_{k=1}^{n} \left\{ Y_k - \beta_0 - \sum_{j=1}^{d} \beta_j (X_{kj} - x_j) \right\}^2 K_B(\mathbf{X}_k - \mathbf{x}), \tag{2.3}$$

with respect to $\beta = (\beta_0, \beta_1, \cdots, \beta_d)^T$, where $\beta_0 = m(\mathbf{x})$, $\beta_j = (\partial m / \partial x_j)(\mathbf{x})$, and $j = 1, 2, \cdots, d$.

Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_2, \cdots, \widehat{\beta}_d)^T$ denote the estimator of $\boldsymbol{\beta} = (\beta_0, \beta_2, \cdots, \beta_d)^T$ resulting from problem (2.3).

Let

$$\mathbf{X}_D = \begin{pmatrix} 1 & (X_{11} - x_1) & \dots & (X_{1d} - x_d) \\ 1 & (X_{21} - x_1) & \dots & (X_{2d} - x_d) \\ \vdots & \vdots & \dots & \vdots \\ 1 & (X_{n1} - x_1) & \dots & (X_{nd} - x_d) \end{pmatrix}, \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

and $\mathbf{W} = \text{diag}\left\{K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x})\right\}_{k=1}^{n}$.

Then the weighted least squares problem (2.3) can be written with matrix notation as :

$$\widehat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

The solution to this weighted least regression problem is

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} \mathbf{X}_D^T \mathbf{W} \mathbf{Y}.$$

The estimates of $m(\mathbf{x})$ and its partial derivatives are given by

$$\widehat{m}(\mathbf{x}) = \widehat{\beta}_0, \qquad (\widehat{\partial m}/\partial x_j)(\mathbf{x}) = \widehat{\beta}_j; \qquad j = 1, 2, \cdots, d$$

We consider a special case with $d = 2$. We have

$$\mathbf{X}_D^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ (X_{11} - x_1) & (X_{21} - x_1) & \dots & (X_{n1} - x_1) \\ (X_{12} - x_2) & (X_{22} - x_2) & \dots & (X_{n2} - x_2) \end{pmatrix}$$

and $\mathbf{W} = \text{diag}\left(K_{\mathbf{B}}(\mathbf{X}_1 - \mathbf{x}), K_{\mathbf{B}}(\mathbf{X}_2 - \mathbf{x}), \dots, K_{\mathbf{B}}(\mathbf{X}_n - \mathbf{x})\right)$. Then

$$(\mathbf{X}_D^T \mathbf{W})\mathbf{Y} = \begin{pmatrix} \sum_{k=1}^n K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x}) \mathbf{Y}_k \\ \sum_{k=1}^n K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x})(X_{k1} - x_1) Y_k \\ \sum_{k=1}^n K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x})(X_{k2} - x_2) Y_k \end{pmatrix} = \begin{pmatrix} G_{n0} \\ G_{n1} \\ G_{n2} \end{pmatrix}$$

where $G_{n0} = \sum_{k=1}^{n} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x})Y_k$ and

$$G_{nj} = \sum_{k=1}^{n} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x})(\mathbf{X}_{kj} - x_k)Y_k, \qquad j = 1, 2.$$

Let

$$N_{npq} = \sum_{k=1}^{n} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x})(X_{k1} - x_1)^p (X_{k2} - x_2)^q$$

where

$$0 \leq p, q \leq 2, \ \ 0 \leq p + q \leq 2.$$

Then

$$(\mathbf{X}_D^T \mathbf{W})\mathbf{X}_D = \begin{pmatrix} N_{n00} & N_{n10} & N_{n01} \\ N_{n10} & N_{n20} & N_{n11} \\ N_{n01} & N_{n11} & N_{n02} \end{pmatrix}.$$

Therefore,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} (\mathbf{X}_D^T \mathbf{W} \mathbf{Y}) = \begin{pmatrix} N_{n00} & N_{n10} & N_{n01} \\ N_{n10} & N_{n20} & N_{n11} \\ N_{n01} & N_{n11} & N_{n02} \end{pmatrix}^{-1} \begin{pmatrix} G_{n0} \\ G_{n1} \\ G_{n2} \end{pmatrix};$$

that is,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} N_{n00} & N_{n10} & N_{n01} \\ N_{n10} & N_{n20} & N_{n11} \\ N_{n01} & N_{n11} & N_{n02} \end{pmatrix}^{-1} \begin{pmatrix} G_{n0} \\ G_{n1} \\ G_{n2} \end{pmatrix}.$$

Hence,

$$
\begin{aligned}
\widehat{\beta}_0 &= \frac{1}{\det \mathbf{N}} \left[(N_{n20}N_{n02} - N_{n11}^2)G_{n0} - (N_{n10}N_{n02} - N_{n01}N_{n11})G_{n01} \right. \\
&\qquad \left. + (N_{n10}N_{n11} - N_{n01}N_{n20})G_{n2} \right] \\
&= \frac{1}{\det \mathbf{N}} \sum_{k=1}^{n} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x}) \left[(N_{n20}N_{n02} - N_{n11}^2) \right. \\
&\quad - (X_{k1} - x_1) \times (N_{n10}N_{n02} - N_{n01}N_{n11}) \\
&\quad \left. + (X_{k2} - x_2) \times (N_{n10}N_{n11} - N_{n01}N_{n20}) \right] Y_k \\
&= \sum_{k=1}^{n} W_{nk} Y_k
\end{aligned}
$$

where $\mathbf{N}$ is 3×3 matrix and W_{nk} is a weight function defined as follows:

$$
\begin{aligned}
W_{nk} &= \frac{1}{\det \mathbf{N}} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x}) \\
&\quad \times \left[\left(N_{n20}N_{n02} - N_{n11}^2 \right) - (N_{n10}N_{n02} - N_{n01}N_{n11}) \right. \\
&\quad \left. + (N_{n10}N_{n11} - N_{n01}N_{n20}) \right].
\end{aligned}
$$

Define $F_{npq} = (1/n^2)\, N_{npq}$. Then

$$
\begin{aligned}
W_{nk} &= \frac{n^4}{\det \mathbf{N}} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x}) \left[(F_{n20}F_{n02} - F_{n11}^2) \right. \\
&\quad -(X_{k1} - x_1)(F_{n10}F_{n02} - F_{n01}F_{n11}) \qquad (2.4) \\
&\quad \left. + (X_{k2} - x_2)(F_{n10}F_{n11} - F_{n01}F_{n20}) \right] \\
&= \frac{1}{n^2} K_{\mathbf{B}}(\mathbf{X}_k - \mathbf{x}) \frac{F_N}{F_D} \qquad (2.5)
\end{aligned}
$$

with

$$
\begin{aligned}
F_N &= (F_{n20}F_{n02} - F_{n11}^2) - (X_{k1} - x_1)(F_{n10}F_{n02} - F_{n01}F_{n11}) \\
&\quad +(X_{k2} - x_2)(F_{n10}F_{n11} - F_{n01}F_{n20}), \\
F_D &= F_{n00}F_{n20}F_{n02} + 2F_{n10}F_{n01}F_{n11} \\
&\quad -F_{n01}^2 F_{n20} - F_{n10}^2 F_{n02} - F_{n11}^2 F_{n00}.
\end{aligned}
$$

The following lemma states one of the properties of the weight function.

Lemma 2.1 *Let W_{nk} be defined as in (2.5). Then*

$$\sum_{k=1}^{n}\left(\mathbf{X}_{kq}-\mathbf{x}_q\right)W_{nk}=\begin{cases}1 & \textit{if q=0;}\\ 0 & \textit{if q=1,2.}\end{cases}$$

In particular,

$$\sum_{k=1}^{n}W_{nk}=1.$$

We define $\mathbf{X}_{kq}-\mathbf{x}_q=1$ if $q=0$.

Proof: Let $e_{v+1}^T=(0,\cdots,0,1,0,\cdots,0)$ be a vector with 1 as its $(v+1)th$ component and 0 otherwise. Also, let $\mathbf{S}_n=\mathbf{X}_D^T\mathbf{W}\mathbf{X}_D$. Then

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_0 &= e_1^T\widehat{\boldsymbol{\beta}}\\ &= e_1^T\left(\mathbf{S}_n^{-1}\mathbf{X}_D^T\mathbf{W}\mathbf{Y}\right)\\ &= \sum_{k=1}^{n}W_{nk}Y_k\end{aligned}$$

where

$$W_{nk}=e_1^T\mathbf{S}_n^{-1}\left\{1,X_{k1}-x_1,X_{k2}-x_2\right\}^T K_{\mathbf{B}}\left(\mathbf{X}_k-\mathbf{x}\right).$$

Observe that

$$\begin{aligned}&\sum_{k=1}^{n}\left(X_{kq}-x_q\right)W_{nk}\\ &= \sum_{k=1}^{n}\left(X_{kq}-x_q\right)e_1^T\mathbf{S}_n^{-1}\left\{1,X_{k1}-x_1,X_{k2}-x_2\right\}^T K_{\mathbf{B}}\left(\mathbf{X}_k-\mathbf{x}\right)\\ &= \sum_{k=1}^{n}\left(X_{kq}-x_q\right)e_1^T\mathbf{S}_n^{-1}\begin{pmatrix}1\\ X_{k1}-x_1\\ X_{k2}-x_2\end{pmatrix}K_{\mathbf{B}}\left(\mathbf{X}_k-\mathbf{x}\right)\\ &= e_1^T\mathbf{S}_n^{-1}\left(\mathbf{S}_n^{-1}e_{q+1}\right)=\begin{cases}1 & \text{if } q=0;\\ 0 & \text{if } q=1,2.\end{cases}\end{aligned}$$

In particular,

$$\sum_{k=1}^{n} W_{nk} = \sum_{k=1}^{n} (\mathbf{X}_0 - x_0) W_{nk} = 1. \qquad \square$$

Chapter 3

Quasi-likelihood Function for Longitudinal Data

3.1 Quasi-likelihood Function with Known Covariance Matrix

Suppose that there are longitudinal observations $\{y_{kj}, \mathbf{x}_{kj}\}$ taken at times t_{kj}, $j = 1, 2, \cdots, T_k$ and subjects $k = 1, 2, \cdots, n$. Here y_{kj} is the response variable and $\mathbf{x_{kj}}$ is a $p \times 1$ vector of covariates at time point t_{kj}.

Let $\mathbf{Y_k}$ be the $T_k \times 1$ vector $(y_{k1}, \cdots, y_{kT_k})$ with mean vector $\boldsymbol{\mu}_k = E(\mathbf{Y}_k)$, covariance matrix $\boldsymbol{\Sigma}_k = \text{Cov}(\mathbf{Y}_k)$ and $\mathbf{X}_k$ be the $T_k \times p$ matrix $(\mathbf{x}_{k1}, \cdots, \mathbf{x}_{kT_k})^T$ for the kth subject. The main interest of longitudinal data analysis is to investigate the dependence of the outcome variable on the covariate variables. A generalized linear model will be established for this purpose. The framework for generalized linear models and maximum quasi-likelihood estimation, derived from the multivariate settings in McCullagh (1983) and McCullagh and Nelder (1983), can be set out as two main components:

i) Model specifications for the mean vector $\boldsymbol{\mu}_k = (\mu_{k1}, \cdots, \mu_{kT_k})$ and covariance matrix $\mathbf{\Sigma}_k$, with

$$\mu_{kj} = g(\eta_{kj}), \qquad \eta_{kj} = x_{kj}^T \boldsymbol{\beta}, \qquad \mathbf{\Sigma}_k = \phi \mathbf{V}(\boldsymbol{\mu}_k)$$

where $j = 1, 2, \cdots, T_k$; $k = 1, 2, \cdots, n$ and g is a known link function having bounded third derivatives. Notice that $g(\cdot)$ is often referred as the inverse link function in the literature on generalized linear models. The $p \times 1$ vector $\boldsymbol{\beta}$ consists of regression parameters, $\mathbf{V}(\cdot)$ is a symmetric positive definite matrix of known variance and covariance functions, and $\phi > 0$ is a scale factor, either a known constant or an unknown parameter.

ii) The log quasi-likelihood function is given by the following system of partial differential equations

$$\frac{\partial \mathcal{L}(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\mu}} = \sum_{k=1}^{n} \mathbf{V}^{-1}(\boldsymbol{\mu}_k) \left(\mathbf{Y}_k - \boldsymbol{\mu}_k \right) / \phi. \tag{3.1}$$

Estimation of the regression model parameter $\boldsymbol{\beta}$ is based on the quasi-score function:

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\beta}} \tag{3.2}$$

$$= \sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}^{-1}(\boldsymbol{\mu}_k)(\mathbf{Y}_k - \boldsymbol{\mu}_k)/\phi \tag{3.3}$$

where $\mathbf{D}_k$ is a $p \times T_k$ matrix with (j, l)th element $(\partial/\partial\beta_l)\mu_{kj}$ and $\mathbf{D}_k$ has rank p for all $\boldsymbol{\beta}$. (This would imply that distinct $\boldsymbol{\beta}$'s imply distinct $\boldsymbol{\mu}$'s). The quasi-score function has the following properties according to the results of McCullagh (1983) and McCullagh and Nelder (1983):

(i)

$$E\left(\mathbf{U}(\boldsymbol{\beta})\right) = \mathbf{0},$$

(ii)

$$\mathrm{Cov}\left(\mathbf{U}(\boldsymbol{\beta})\right) = \sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}(\boldsymbol{\mu}_k)^{-1} \mathbf{D}_k / \phi = \mathbf{i}_\beta,$$

(iii)

$$-E\left(\partial\left(\mathbf{U}(\boldsymbol{\beta})\right)/\partial\boldsymbol{\beta}\right) = \sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}\left(\boldsymbol{\mu}_k\right)^{-1} \mathbf{D}_k/\phi = \mathrm{Cov}\left(U(\boldsymbol{\beta})\right) = \mathbf{i}_\beta,$$

where $-\mathbf{i}_\beta$ is the expected second derivative matrix of the log quasi-likelihood function $L(\boldsymbol{\mu}, \mathbf{Y})$.

By the results in McCullagh (1983), we also have following facts:

1. $U(\boldsymbol{\beta}) = O_p(n)$;
2. $\mathbf{I}_\beta = -\left(\partial^2/\partial\beta_r\beta_s\right)\left(U(\boldsymbol{\beta})\right) = O_p(n)$;
3. $\mathbf{I}_\beta - \mathbf{i}_\beta = O_p(n^{1/2})$;
4. There exists $\widehat{\boldsymbol{\beta}}$, a solution of $U(\boldsymbol{\beta}) = 0$, such that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(n^{1/2})$.

Furthermore, the maximum quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$, a solution of the quasi-likelihood equation $U(\boldsymbol{\beta}) = 0$, satisfies:

(i) Consistency:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p\left(n^{-1}\right) \qquad \text{as } n \to \infty$$

(ii) Asymptotic Normality:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to \mathcal{N}\left(\mathbf{0}, n\mathbf{V}_{\boldsymbol{\beta}}\right) \mathrm{as} n \to \infty$$

where

$$\mathbf{V}_{\boldsymbol{\beta}} = \phi\left[\sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}(\mu_k)^{-1} \mathbf{D}_k\right]^{-1} = \mathbf{i}_\beta^{-1},$$

and the covariance matrix of estimator $\widehat{\boldsymbol{\beta}}$ is $\mathbf{V}_{\boldsymbol{\beta}}$, provided that the eigenvalues λ of $\mathbf{i}_\beta$ satisfy $0 < c_1 < \lambda < c_2 < \infty$ for sufficiently large n (Weddernburn 1974).

3.2 Nonparametric Quasi-likelihood Function with Unknown Covariance Matrix

When the covariance matrix $\mathbf{V}(\boldsymbol{\mu})$ is unknown, an extended nonparametric approach is proposed in this dissertation by replacing $\mathbf{V}(\boldsymbol{\mu})$ in (3.1) with a consistent nonparametric estimator $\widehat{\mathbf{V}}_n(\widehat{\boldsymbol{\mu}})$ to construct a nonparametric quasi-likelihood function $\mathcal{L}^*(\boldsymbol{\mu}, \mathbf{Y})$. The nonparametric quasi-likelihood function $\mathcal{L}^*(\boldsymbol{\mu}, \mathbf{Y})$ will be given as follows:

$$\frac{\partial \mathcal{L}^*(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\mu}} = \sum_{k=1}^{n} \widehat{\mathbf{V}}_n^{-1}(\boldsymbol{\mu}_k)\left(\mathbf{Y}_k - \boldsymbol{\mu}_k\right)/\phi. \tag{3.4}$$

Then we have the nonparametric quasi-likelihood score function:

$$\begin{aligned} \mathbf{U}^*(\boldsymbol{\beta}) &= \frac{\partial \mathcal{L}^*(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\beta}} \\ &= \sum_{k=1}^{n} \mathbf{D}_k^T \widehat{\mathbf{V}}_n^{-1}(\boldsymbol{\mu}_k)(\mathbf{Y}_k - \boldsymbol{\mu}_k)/\phi. \end{aligned}$$

In Chapter 5, we will show that the maximum nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$, a solution of the nonparametric quasi-likelihood score equation:

$$\mathbf{U}^*(\boldsymbol{\beta}) = 0, \tag{3.5}$$

is still consistent and has asymptotically normal distribution $N(\boldsymbol{\beta}, \mathbf{V}_{\boldsymbol{\beta}})$, with

$$\mathbf{V}_{\boldsymbol{\beta}} = \phi \left[\sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}(\mu_k)^{-1} \mathbf{D}_k\right]^{-1},$$

the same asymptotic distribution as the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$ obtained with known covariance matrix. We will discuss this approach in detail in Chapter 4 and Chapter 5.

Chapter 4

Nonparametric Quasi-likelihood Model for Longitudinal Data

Suppose that there are longitudinal observations $\{y_{kj}, \mathbf{x}_{kj}\}$ for times t_{kj}, $j = 1, 2, \ldots, T_k$, and subjects $k = 1, 2, \ldots, n$. Here y_{kj} is the response variable and $\mathbf{x_{kj}}$ is a $q \times 1$ vector of covariates at time point t_{kj}.

Let $\mathbf{Y}_k$ be the $T_k \times 1$ vector $(y_{k1}, \ldots, y_{kT_k})^T$ with mean vector $\boldsymbol{\mu}_k = E(\mathbf{Y}_k)$ and unknown covariance matrix $\boldsymbol{\Sigma}_k = \mathrm{Cov}(\mathbf{Y}_k)$, and let $\mathbf{X}_k$ be the $T_k \times p$ matrix $(\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kT_k})^T$ for the kth subject. The vectors $\mathbf{Y}_k, k = 1, 2, \ldots, n$ are independent. For simplicity, we assume that $T_k = T$ for each k and $\phi = 1$.

The following assumptions about the proposed nonparametric quasi-likelihood model for longitudinal data will be used throughout the remainder of the thesis.

(N1) Model specifications for response variable y_{kj} and mean μ_{kj} :

$$y_{kj} = g(\eta_{kj}) + \epsilon_{kj};$$

$$\mu_{kj} = g(\mathbf{x}_{kj}^T \boldsymbol{\beta});$$

$$\eta_{kj} = \mathbf{x}_{kj}^T \boldsymbol{\beta};$$

$$k = 1, \ldots, n; \qquad j = 1, \ldots, T;$$

where $g(\cdot)$ is a known function, and called the link function. Notice that $g(\cdot)$

is often called as the inverse link function in the literature on generalized linear models. Suppose that $\mathbf{x}_{kj}$ is the nonrandom p-dimensional predictor variable vector corresponding to the observation y_{kj} for the subject at time point j. The error vector $\boldsymbol{\varepsilon}_k = (\epsilon_{k1}, \ldots, \epsilon_{kT})$ satisfies

$$E(\boldsymbol{\varepsilon}_k) = 0, \qquad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_l) < \infty$$

for any $1 \le l, k \le n$, and $1 \le j \le T$.

(N2) There exists a positive definite matrix of covariance functions (depending on the means) $\mathbf{V}(\boldsymbol{\mu}) = (\sigma_{st}(\mu_{ks}, \mu_{kt}))_{T \times T}$, $1 \le s, t \le T$, where $\|\mathbf{V}^{-1}\|_\infty \ge r$ for some $r > 0$, such that

$$\begin{aligned} E(\epsilon_{ks}\epsilon_{kt}) &= \text{Cov}(\epsilon_{ks}, \epsilon_{kt}) \\ &= \text{Cov}(y_{ks}, y_{kt}) \\ &= \sigma_{st}(\mu_{ks}, \mu_{kt}). \end{aligned}$$

Here $\{\sigma_{st}(u_s, u_t)\}_{1 \le s,t, \le T}$ is an array of covariance functions (depending on the unknown means) and $\|\cdot\|_\infty$ is the matrix L_∞ -norm. That is, if $A = (a_{ij})$ is a $n \times n$ matrix, then $\|A\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|$. Assume this matrix $\mathbf{V}(\boldsymbol{\mu})$ of variance functions is the unknown covariance matrix $\text{Cov}(\mathbf{Y}_k)$ and is going to be estimated by the proposed nonparametric approach. The dispersion parameter ϕ is a known constant. For simplicity, assume that $\phi = 1$.

(N3) There exists a constant $M > 0$ such that $\max_{1 \le k \le n} \|\mathbf{x}_{kj}\|_\infty \le M < \infty$, for all $1 \le k \le n$ and $1 \le j \le T$. Let $\mathbf{X}_k = (\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kT})$ be a matrix for any k with $1 \le k \le n$.

Condition (N3) implies that the predictor variable vectors are bounded, since the covariate vector $\mathbf{x}_{kj}$ at time j is fixed.

(N4) Given the link function g and the generalized linear model parameter vector $\boldsymbol{\beta}$, we assume that $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ form a sequence of matrices such that the mean vectors $\boldsymbol{\mu}_k$ or $\mu_{kj} = g(\mathbf{x}_{kj}^T \boldsymbol{\beta})$ are generated by a design density $f_{\boldsymbol{\mu}}(\mathbf{u})$ which is assumed to satisfy the following conditions:

Let $\mathcal{C} \subseteq \mathcal{R}^{\mathcal{T}}$ be a subspace with a design measure such that all 2-dimensional marginals are absolutely continuous with respect to Lebesgue measure. This design measure has a T-dimensional density, positive everywhere in $\mathcal{C}$. Let $f(\mathbf{x})$ be a T dimensional density, so that $\int_{\mathcal{C}} f(\mathbf{x})\, d\mathbf{x} = 1$. The support of $f(\mathbf{x})$ is a compact set D in $\mathcal{C}$ and D must contain a t-dimensional rectangle ($1 \leq t \leq T$). The function $f(\mathbf{x})$ is twice conditionally differentiable, exchangeably differentiable and has other regular analytical properties. Also $f(\mathbf{x})$ satisfies $0 \leq \inf f(\mathbf{x}) \leq \sup f(\mathbf{x}) < \infty$. The design matrices $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ are chosen in such way that the mean values $\mu_{kt} = g\left(\mathbf{x}_{kt}^T \boldsymbol{\beta}\right)$ satisfy

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\mu_{kt}} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) d\mathbf{x} = \frac{k-1}{n-1}$$

for any $1 \leq k \leq n$ and $1 \leq t \leq T$.

Let f_{st} be the two-dimensional marginal density of $f(\mathbf{x})$ for any pair (s, t), $1 \leq s, t \leq T$, and let D_{st} be the support of f_{st} such that

$$D_{st} = \{(u_s, u_t) | (u_1, \ldots, u_s, \ldots, u_t, \ldots, u_T) \in D\}$$

For simplicity, let $f_{st}(x, y)$ be the marginal density of $(\mathbf{X}_s, \mathbf{X}_t)$ for a pair (s, t), $1 \leq s, t \leq T$.

(N5) For any $1 \leq s, t \leq T$, there is a constant c such that

$$E(\epsilon_{ks}^2 \epsilon_{kt}^2) \leq c.$$

(N6) The link function $g(\cdot)$ is three times and the variance functions $\{\sigma_{st}(\cdot)\}$ in the covariance matrix $\mathbf{V}(\cdot)$ are twice continuously differentiable with bounded derivatives, for any pair (s, t).

(N7) There is a positive definite matrix $\mathbf{\Sigma}$, such that

$$\frac{1}{n} \sum_{k=1}^{n} \mathbf{D}_k^T V^{-1}(\boldsymbol{\mu}_k) \mathbf{D}_k \to \mathbf{\Sigma} \qquad \text{as } n \to \infty$$

where $\mathbf{D}_k$ is a $T \times p$ matrix with (j, l)th element $(\partial / \partial \beta_l) \mu_{kj}$.

Given the known covariance matrix $\mathbf{V}(\cdot)$, the log quasi-likelihood function will be given as the same as in (3.1):

$$\frac{\partial \mathcal{L}(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\mu}} = \sum_{k=1}^{n} \mathrm{V}^{-1}(\boldsymbol{\mu}_k) (\mathbf{Y}_k - \boldsymbol{\mu}_k) \tag{4.1}$$

and the estimators of the regression parameters will be given by solving the quasi-likelihood score equation

$$\mathbf{U}(\boldsymbol{\beta}) = 0$$

where

$$\begin{aligned} \mathbf{U}(\boldsymbol{\beta}) &= \frac{\partial \mathcal{L}(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\beta}} \qquad (4.2) \\ &= \sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}^{-1} (\boldsymbol{\mu}_k) (\mathbf{Y}_k - \boldsymbol{\mu}_k). \end{aligned}$$

If the covariance matrix $\mathbf{V}(\cdot)$ is unknown, then we will obtain the nonparametric log quasi-likelihood function $\mathcal{L}^*(\boldsymbol{\mu}_k, \mathbf{Y})$ by substituting the nonparametric estimator $\mathbf{V}_n(\boldsymbol{\mu}) = (\sigma_{nst}(\mu_{ks}, \mu_{kt}))_{T \times T}$ for the covariance matrix $\mathbf{V}(\cdot)$ in the log quasi-likelihood function $\mathcal{L}(\boldsymbol{\mu}, \mathbf{Y})$ in (4.1) as follows:

$$\frac{\partial \mathcal{L}^*(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\mu}} = \sum_{k=1}^{n} \mathbf{V}_n^{-1}(\boldsymbol{\mu}_k)(\mathbf{Y}_k - \boldsymbol{\mu}_k), \tag{4.3}$$

and the nonparametric quasi-score function $\mathbf{U}^*(\beta)$ for this nonparametric quasi-likelihood is

$$\mathbf{U}^*(\boldsymbol{\beta}) = \sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}_n^{-1}(\boldsymbol{\mu}_k)(\mathbf{Y}_k - \boldsymbol{\mu}_k) \tag{4.4}$$

where $\mathbf{D}_k$ is a $T \times p$ matrix with (j, l)th element $(\partial/\partial\beta_l)\mu_{kj}$; $\mathbf{V}_n(\cdot) = (\sigma_{nst}(\cdot))$ is a $T \times T$ matrix of variance functions. The nonparametric quasi-likelihood estimator (NQLE) $\widehat{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$ is a solution of the nonparametric quasi-likelihood scoring equation

$$\mathbf{U}^*(\boldsymbol{\beta}) = 0. \tag{4.5}$$

Assume the bandwidth matrix B is a nonsingular symmetric 2×2 matrix and K is a 2-variate probability density function which satisfies the following conditions:

(K1) $\int K(\mathbf{u})d(\mathbf{u}) = 1, \qquad \int \mathbf{u}K(\mathbf{u})d\mathbf{u} = \mathbf{0}.$

(K2) K has support $[-1, 1] \times [-1, 1]$ and that

$$\int u_k u_j K(\mathbf{u})d(\mathbf{u}) = \delta_{kj} m_2(K)$$

where $m_2(K) = \int u_k^2 K(\mathbf{u})d\mathbf{u} \leq 0$ and δ_{kj} is kronecker delta. Moreover K is continuously differentiable on $[-1, 1] \times [-1, 1]$ and

$$K(-u, -v) = K(u, v), \; K(u, v) \geq 0.$$

In other words, the mean of the density function $K(\cdot)$ is zero and the covariance matrix of K is $m_2(K)I_2$, with I_2 the 2×2 identity matrix.

Further we assume that K is Lipschitz. This means that for any small number $\epsilon > 0$, if $\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \epsilon$ then there is a constant k, such that

$$\|K(\mathbf{x}_1) - K(\mathbf{x}_2)\| \leq k\|\mathbf{x}_1 - \mathbf{x}_2\|.$$

Define

$$K_{\mathbf{B}}(\mathbf{u}) = \frac{1}{|\mathbf{B}|} K\left(\mathbf{B}^{-1}\mathbf{u}\right)$$

where $|\mathbf{B}|$ denotes the determinant of $\mathbf{B}$. From $(K2)$, we have

$$\iint u^p v^q K(u, v)\, dudv = 0 \qquad \text{if } p + q \text{ is odd;}$$

$$\int\int u^p v^q K(u, v)\, dudv < \infty \qquad \text{if } p + q \text{ is even.}$$

Let

$$\alpha_{pq} = \int\int u^p v^q K(u, v)\, dudv < \infty \qquad \text{if } p + q \text{ is even.}$$

It is clear that

$$\alpha_{00} = 1.$$

(K3) The sequence of bandwidth matrices $\mathbf{B} = \mathbf{B}(n) = \text{diag}\{h_s(n), h_t(n)\}$ satisfies:

1. $h_s = h_s(n) > 0$, and $h_t = h_t(n) > 0$;
2. $h_s \to 0$ and $h_t \to 0$ as $n \to \infty$;
3. $h_s/h_t = O(1)$ as $n \to \infty$;
4. $nh_s^2 \to \infty$ and $nh_t^2 \to \infty$ as $n \to \infty$.
5. $(\log n/nh_s)^{1/2} = o(1)$
6. $(\log n/nh_t)^{1/2} = o(1)$

7. $((\log n)/(n^2 h_s h_t))^{1/2} = o(1)$

From (N1) and (N2), we have

$$\epsilon_{kj} = y_{kj} - \mu_{kj},$$

$$\epsilon_{ks}\epsilon_{kt} = (y_{ks} - \mu_{ks})(y_{kt} - \mu_{kt}).$$

Then

$$\begin{aligned} E(\epsilon_{ks}\epsilon_{kt}) &= E((y_{ks} - \mu_{ks})(y_{kt} - \mu_{kt})) \\ &= \text{Cov}(y_{ks}, y_{kt}) \quad (4.6) \\ &= \sigma_{st}(\mu_{ks}, \mu_{kt}) \quad (4.7) \end{aligned}$$

Therefore, we have the model:

$$\epsilon_{ks}\epsilon_{kt} = \sigma_{st}(\mu_{ks}, \mu_{kt}) + \delta_{kst},$$

where δ_{kst} is an error term with $E\delta_{kst} = 0$ and $k = 1, 2, \ldots, n$.

Let the positive definite matrix $\mathbf{V}_n = (\sigma_{nst}(u_s, u_t))$ be a nonparametric estimator of the covariance matrix. Then we have

$$\sigma_{nst}(u_s, u_t) = \sum_{k=1}^{n} W_{nk}(u_s, u_t; \mu_{ks}, \mu_{kt})\epsilon_{ks}\epsilon_{kt}. \quad (4.8)$$

Here $W_{nk}(u_s, u_t; \mu_{ks}, \mu_{kt})$ is a local linear weight function, defined as the same as (2.5) in Chapter 2. That is,

$$W_{nk}(\mu_{ks}, \mu_{kt}; u_s, u_t) = \frac{1}{n^2} K_{\mathbf{B}}\left((\mu_{ks}, \mu_{kt}) - (u_s, u_t)\right) F(\mu_{ks}, \mu_{kt}; u_s, u_t) \quad (4.9)$$

where

$$F(\mu_{ks}, \mu_{kt}; u_s, u_t) = \frac{F_N}{F_D}$$

and

$$
\begin{aligned}
F_N &= (F_{n20}F_{n02} - F_{n11}^2) - (\mu_{ks} - u_s)(F_{n10}F_{n02} - F_{n01}F_{n11}) && (4.10)\\
&\quad -(\mu_{kt} - u_t)(F_{n10}F_{n11} - F_{n01}F_{n20}) \\
F_D &= F_{n00}F_{n20}F_{n02} + 2F_{n10}F_{n01}F_{n11} - F_{n20}F_{n01}^2 && (4.11)\\
&\quad -F_{n10}^2F_{n02} - F_{n11}^2F_{n00}
\end{aligned}
$$

where

$$
F_{npq} = \frac{1}{n^2}\sum_{k=1}^{n} K_{\mathbf{B}}\left((\mu_{ks},\mu_{kt}) - (u_s,u_t)\right)(\mu_{ks} - u_s)^p(\mu_{kt} - u_t)^q \quad (4.12)
$$

for $0 \le p, q \le 2$; $0 \le p + q \le 2$.

Since $\{\mu_{ks}, \mu_{kt}; \epsilon_{ks}, \epsilon_{kt}\}$ are unknown in (4.8), we are unable to calculate $W_{nk}(u_s, u_t; \mu_{ks}, \mu_{kt})$, nor can we calculate the element $\sigma_{nst}(u_s, u_t)$ in the proposed nonparametric estimator of covariance matrix $(\sigma_{nst}(u_s, u_t))_{T\times T}$. In other words, $\mathbf{V}_n(\cdot)$ is not a statistic.

Suppose that we are given an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. Then we have estimated means and observed residuals $\{\widehat{\mu}_{ks}, \widehat{\mu}_{kt}; \widehat{\epsilon}_{ks}, \widehat{\epsilon}_{kt}\}$:

$$
\begin{aligned}
\widehat{\mu}_{ks} &= g_s(\mathbf{x}_{ks}^T\widehat{\boldsymbol{\beta}}); && (4.13)\\
\widehat{\mu}_{kt} &= g_t(\mathbf{x}_{kt}^T\widehat{\boldsymbol{\beta}}); && (4.14)\\
\widehat{\epsilon}_{ks} &= y_{ks} - \widehat{\mu}_{ks}; && (4.15)\\
\widehat{\epsilon}_{kt} &= y_{kt} - \widehat{\mu}_{kt}. && (4.16)
\end{aligned}
$$

so, we can compute

$$
\widehat{\sigma}_{nst}(u_s, u_t) = \sum_{k=1}^{n} \widehat{W}_{nk}(u_s, u_t; \mu_{ks}, \mu_{kt})\widehat{\epsilon}_{ks}\widehat{\epsilon}_{kt} \quad (4.17)
$$

where

$$
\begin{aligned}
&\widehat{W}_{nk}(u_s, u_t; \mu_{ks}, \mu_{kt}) \\
&\quad = \quad W_{nk}(u_s, u_t; \widehat{\mu}_{ks}, \widehat{\mu}_{kt}) \\
&\quad = \quad \frac{1}{n^2} K_{\mathbf{B}}((\widehat{\mu}_{ks}, \widehat{\mu}_{kt}) - (u_s, u_t)) \widehat{F}(u_s, u_t; \widehat{\mu}_{ks}, \widehat{\mu}_{kt}),
\end{aligned} \tag{4.18}
$$

$$
\widehat{F}(u_s, u_t; \widehat{\mu}_{ks}, \widehat{\mu}_{kt}) = \widehat{F}_N \big/ \widehat{F}_D,
$$

$$
\begin{aligned}
\widehat{F}_N \quad = \quad & (\widehat{F}_{n20}\widehat{F}_{n02} - \widehat{F}^2_{n11}) - (\widehat{\mu}_{ks} - u_s) I_{\{|\widehat{\mu}_{ks} - u_s| \leq h_s\}} (\widehat{F}_{n10}\widehat{F}_{n02} - \widehat{F}_{n01}\widehat{F}_{n11}) \\
& -(\widehat{\mu}_{kt} - u_t) I_{\{|\widehat{\mu}_{kt} - u_t| \leq h_t\}} (\widehat{F}_{n10} F_{n11} - \widehat{F}_{n01}\widehat{F}_{n20}),
\end{aligned} \tag{4.19}
$$

$$
\begin{aligned}
\widehat{F}_D \quad = \quad & \widehat{F}_{n00}\widehat{F}_{n20}\widehat{F}_{n02} + 2\widehat{F}_{n10}\widehat{F}_{n01}\widehat{F}_{n11} - \widehat{F}_{n20}\widehat{F}^2_{n01} \\
& -\widehat{F}^2_{n10}\widehat{F}_{n02} - \widehat{F}^2_{n11}\widehat{F}_{n00},
\end{aligned} \tag{4.20}
$$

and

$$
\widehat{F}_{npq} \quad = \quad \frac{1}{n^2} \sum_{k=1}^{n} K_{\mathbf{B}} \left((\widehat{\mu}_{ks}, \widehat{\mu}_{kt}) - (u_s, u_t) \right) (\widehat{\mu}_{ks} - u_s)^p (\widehat{\mu}_{kt} - u_t)^q \tag{4.21}
$$

where $0 \leq p, q \leq 2$; $0 \leq p + q \leq 2$. Unlike $\mathbf{V}_n, \widehat{\mathbf{V}}_n$ is a statistic which can be computed from the observations.

In the following chapter, we will investigate the asymptotic properties of the random matrices $\mathbf{V}_n = (\sigma_{nst}(u_s, u_t))$ and $\widehat{\mathbf{V}}_n = (\widehat{\sigma}_{nst}(u_s, u_t))$, defined in (4.8) and (4.17).

Since (4.5) is a nonlinear equation, the Newton-Raphson method will be applied to solve this equation. In order to obtain $\widehat{\boldsymbol{\beta}}^*$, the solution of equation (4.5), the following iterative formula resulting from applying the Newton-Raphson method to the nonparametric quasi-likelihood scoring equation $\mathbf{U}^*(\boldsymbol{\beta}) = 0$ and updating the nonparametric smoothing technique, will be applied iteratively until a convergence criterion is satisfied:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^*_{(j+1)} &= \widehat{\boldsymbol{\beta}}^*_{(j)} + \left\{ \sum_{k=1}^{n} \mathbf{D}_k^T \left(\widehat{\boldsymbol{\beta}}^*_{(j)} \right) \widehat{\mathbf{V}}_n^{-1} \left(\widehat{\boldsymbol{\mu}}_k \right)_{(j)} \mathbf{D}_k \left(\widehat{\boldsymbol{\beta}}^*_{(j)} \right) \right\}^{-1} \\ &\quad \times \left\{ \sum_{k=1}^{n} \mathbf{D}_k^T \left(\widehat{\boldsymbol{\beta}}^*_{(j)} \right) \widehat{\mathbf{V}}_n^{-1} \left(\widehat{\boldsymbol{\mu}}_k \right)_{(j)} S_k \left(\widehat{\boldsymbol{\beta}}^*_{(j)} \right) \right\} \end{aligned} \tag{4.22}$$

where

$$\widehat{\mathbf{V}}_n^{-1} \left(\widehat{\boldsymbol{\mu}}_k \right)_{(j)} = \widehat{\mathbf{V}}_n^{-1} \left(g \left(\mathbf{X}_k^T \widehat{\boldsymbol{\beta}}^*_{(j)} \right) \right)$$

$$\mathbf{D}_k{}^T (\widehat{\boldsymbol{\beta}}^*_{(j)}) = \mathbf{D}_k{}^T (\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^*_j}$$

and

$$S_k \left(\widehat{\boldsymbol{\beta}}^*_{(j)} \right) = \mathbf{Y}_k g \left(\mathbf{X}_k^T \widehat{\boldsymbol{\beta}}^*_{(j)} \right).$$

The proposed method of obtaining the nonparametric covariance matrix estimator $\widehat{\mathbf{V}}_n(\cdot) = (\widehat{\sigma}_{nst}(\cdot))$ is as follows:

1. Assign a guess value as the initial estimator $\widehat{\boldsymbol{\beta}}^*_{(0)}$ of $\boldsymbol{\beta}$ into regression model.

2. For any $1 \leq s, t \leq T$, obtain the products of residuals $\{\widehat{\epsilon}_{ks} \widehat{\epsilon}_{kt}\}$ from a previously fitted model based on the estimated value $\widehat{\boldsymbol{\beta}}^*_{(0)}$ of $\boldsymbol{\beta}$ from the last step and smooth them by applying the local polynomial smoothing method with $\{\widehat{\mu}_{ks}, \widehat{\mu}_{kt}\}$ as the two predictors. The predicted value from smoothing will be the element $(\widehat{\sigma}_{nst}(\cdot))$ of the nonparametric estimator $\widehat{\mathbf{V}}_n(\cdot)$ of covariance matrix.

3. Repeat Step 1 and Step 2. In other words, substitute the updated nonparametric estimator of covariance matrix $\widehat{\mathbf{V}}_n(\cdot)$ from step 2 into the nonparametric quasi-likelihood score equation to get the updated estimator $\widehat{\boldsymbol{\beta}}^*_{(n)}$ of regression parameters $\boldsymbol{\beta}$ until convergence occurs.

The estimation of the variance function by using the local polynomial fitting of square residuals obtained from a nonparametric regression fit was studied in detail by Ruppert, Wand, Holst and Hössjer (1997). Chou and Müller (1999) used the nonparametric estimator of variance function obtained by smoothing square residuals from a previous regression model for their nonparametric quasi-likelihood. The main idea of the proposed nonparametric quasi-likelihood for longitudinal data is to combine a nonparametric smoothing technique such as local polynomial smoothing and quasi-likelihood estimation method to get the estimate of regression parameter estimators. In other words, use local polynomial smoothing to get the nonparametric covariance matrix estimator, and then replace the unknown covariance matrix with this nonparametric covariance matrix estimator in quasi-likelihood function in order to get the estimate the regression parameters $\boldsymbol{\beta}$. These two procedures will be used iteratively by updating regression parameters and obtaining new residuals and estimated means and thus an updated nonparametric covariance matrix. The updated nonparametric covariance matrix then can be in turn used to update the estimator of regression parameters.

Chapter 5

Asymptotic Properties of Nonparametric Quasi-likelihood Estimator.

In the previous chapter, an extended nonparametric quasi-likelihood approach for longitudinal data was proposed. When the covariance matrix is unknown in a longitudinal data analysis, it can be replaced by an asymptotically consistent nonparametric estimator in quasi-likelihood function to get the nonparametric quasi-likelihood estimator for regression parameters. In this chapter, we will discuss the convergence rates of the nonparametric covariance matrix estimators based on the theoretical residuals $\{\boldsymbol{\varepsilon}_k\}, k = 1, 2, \ldots, n$ and sample residuals $\{\widehat{\boldsymbol{\varepsilon}}_k\}, k = 1, 2, \ldots, n$ and of the asymptotic properties of the nonparametric quasi-likelihood of regression coefficients obtained by replacing the unknown covariance matrix with the nonparametric estimator. The following theorems will show that the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$ of regression parameter $\boldsymbol{\beta}$ obtained from the nonparametric estimator of covariance matrix is not only consistent but also has the same asymptotic distribution as the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$, the solution of the quasi-score equation with known covariance matrix.

The first theorem shows that the element in the nonparametric estimator of unknown covariance matrix defined in (4.8) is consistent when the true means $\boldsymbol{\mu}_k$ and the theoretical residuals $\epsilon_{ks}(k = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, T)$ are known. This theorem also gives the sizes of mean square error and maximum of error of this nonparametric estimator in terms of sample size and the bandwidth matrix. As defined in (4.8) in Chapter 4, an element of the nonparametric estimator of the (s, t) entry covariance matrix will be

$$\sigma_{nst}(u_s, u_t) = \sum_{k=1}^{n} W_{nk}(u_s, u_t; \mu_{ks}, \mu_{kt})\epsilon_{ks}\epsilon_{kt}.$$

Theorem 5.1 *Under* $(N1) - (N7)$ *and* $(K1) - (K3)$ *of Chapter* 4,

(i)

$$\sup_{(u_s,u_t)\in D_{st}} |E\sigma_{nst}(u_s, u_t) - \sigma_{st}(u_s, u_t)| = O\left(h_s^2\right) + O\left(h_t^2\right);$$

(ii)

$$|\sigma_{nst}(u_s, u_t) - E\sigma_{nst}(u_s, u_t)| = O_p\left(\left[\frac{\log n}{n^2 h_s h_t}\right]^{1/2}\right);$$

(iii)

$$\sup_{(u_s,u_t)\in D_{st}} E\left[(\sigma_{nst}(u_s, u_t) - \sigma_{st}(u_s, u_t))^2\right] = O\left(\frac{1}{n^2 h_s h_t} + h_s^2 h_t^2 + h_s^4 + h_t^4\right);$$

(iv)

$$\sup_{(u_s,u_t)\in D_{st}} |\sigma_{nst}(u_s, u_t) - \sigma_{st}(u_s, u_t)| = O_p\left(\left[\frac{\log n}{n^2 h_s h_t}\right]^{1/2} + h_s^2 + h_t^2\right).$$

Since the definition of nonparametric estimator of unknown covariance matrix in (4.8) depends on the values of $\{\boldsymbol{\mu}_k, \boldsymbol{\varepsilon}_k\}$, we can not obtain σ_{nst} because we can not observe $\{\boldsymbol{\mu}_k, \boldsymbol{\varepsilon}_k\}$. Instead of using the nonparametric covariance matrix estimator $\mathbf{V}_n = (\sigma_{nst})$, $\widehat{\mathbf{V}}_n = (\widehat{\sigma}_{nst})_{T\times T}$, which is a statistic, will be used as a nonparametric covariance matrix estimator. Does the estimator $\widehat{\mathbf{V}} = (\widehat{\sigma}_{nst})_{T\times T}$ behave like the estimator $\mathbf{V}_n = (\sigma_{nst})_{T\times T}$? The next theorem shows the nonparametric covariance matrix estimator $\widehat{\mathbf{V}}_n = (\widehat{\sigma}_{nst})_{T\times T}$, obtained by smoothing observable $\{\widehat{\boldsymbol{\mu}}_k, \widehat{\boldsymbol{\varepsilon}}_k\}$, converges uniformly to the true covariance matrix $\mathbf{V}(\cdot) = (\sigma_{st}(\cdot))_{T\times T}$, provided that the regression parameter estimator $\widehat{\boldsymbol{\beta}}$ is consistent estimator of $\boldsymbol{\beta}$.

Theorem 5.2 *Under* $(N1)-(N7)$ *and* $(K1)-(K3)$, *if* $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\| = O_p\left(1/\sqrt{n}\right)$, *then for* $(u_s, u_t) \in D$,

$$\sup_{(u_s,u_t)\in D} |\widehat{\sigma}_{nst} - \sigma_{st}| = O_p\left(\left[\frac{\log n}{n^2 h_s h_t}\right]^{1/2} + h_s^2 + h_t^2 + \frac{1}{\sqrt{n}h_s} + \frac{1}{\sqrt{n}h_t}\right)$$

Therefore,

$$\widehat{\mathbf{V}}_n \xrightarrow{p} \mathbf{V} \qquad as \quad n \to \infty$$

for any $(u_s, u_t) \in D$.

The proof of this theorem will be presented in Chapter 7. By substituting the consistent nonparametric covariance matrix $\widehat{\mathbf{V}}_n$ for the unknown covariance matrix in the quasi-likelihood score function, the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$, the solution of nonparametric quasi-likelihood score equation, is asymptotically normally distributed. The following theorem shows that this nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$ will have the same efficiency as the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$, the solution of quasi-likelihood score equation with known covariance matrix.

Theorem 5.3 *Assume that (N1)-(N7) and (K1)-(K3) are satisfied in the nonparametric quasi-likelihood model. Assume that the covariance matrix* $\mathbf{V}(\cdot)$ *is estimated by a positive definite matrix* $\widehat{\mathbf{V}}_n(\cdot) = (\widehat{\sigma}_{nst}(\cdot))_{T\times T}$ *which satisfies*

*(***K4***) There is a constant c such that cond(*$\widehat{\mathbf{V}}_n$*), the condition number of* $\widehat{\mathbf{V}}_n$ *is bounded above by c, for all n.*

In addition assume that for each $\widehat{\sigma}_{nst}$*, there is a sequence* $\lambda_n > 0$ *such that* $|\widehat{\sigma}_{nst}| > \lambda_n$ *and as* $n \to \infty$,

(i) $\lambda_n \to 0$;

(ii) $h_s/\lambda_n \to 0, \qquad h_t/\lambda_n \to 0$;

(iii) $nh_s^2{\lambda_n}^2 \to \infty, \qquad nh_t^2{\lambda_n}^2 \to \infty$;

(iv) $\left(n^2 h_s h_t {\lambda_n}^2\right)/\log\ n \to \infty$.

Then the NQLE $\widehat{\boldsymbol{\beta}}^*$ *in* (3.5) *has an asymptotically normal distribution such that, as* $n \to \infty$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}^{-1}\right). \tag{5.1}$$

From (5.1) and the results in Chapter 3, we know that the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$ has the same asymptotic distribution as the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$, the solution of the quasi-likelihood score equation with known covariance matrix. Also, an asymptotic test statistic can be derived from following Corollary, which follows from Theorem 5.3.

Corollary 5.1 *Let* $\widehat{\boldsymbol{\beta}}^*$ *be the nonparametric quasi-likelihood estimator, the solution of* (3.5)*, and let*

$$\widehat{\boldsymbol{\Sigma}}^{-1} = \frac{1}{n}\left(\sum_{k=1}^{n} \widehat{\mathbf{D}}_k^T \widehat{\mathbf{V}}_n \widehat{\mathbf{D}}_k\right)^{-1},$$

where $\widehat{\mathbf{D}}_k$ *is a* $T \times p$ *matrix with* (j,l)*th element* $(\partial/\partial\beta_l)\mu_{kj}$ *and*

$$\widehat{\mathbf{V}}_n = (\widehat{\sigma}_{nst}(\widehat{\mu}_{ks}, \widehat{\mu}_{kt}))$$

with $\widehat{\mu}_{ks} = g\left(\mathbf{x}_{ks}^T\widehat{\boldsymbol{\beta}}^*\right)$ *and* $\widehat{\mu}_{kt} = g\left(\mathbf{x}_{kt}^T\widehat{\boldsymbol{\beta}}^*\right)$. *Then we have*

$$\widehat{\boldsymbol{\Sigma}}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}^{-1}, \tag{5.2}$$

and

$$\left[\sum_{k=1}^{n} \widehat{\mathbf{D}}_k^T \widehat{\mathbf{V}}_n \widehat{\mathbf{D}}_k\right]^{-1/2} \left(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right) \tag{5.3}$$

Since we know the asymptotic distribution of the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$ and can estimate its covariance matrix, we can develop an asymptotic test statistic for a class of hypotheses:

$$H_0 : A\boldsymbol{\beta} = \mathbf{c}_0 \qquad \text{versus} \qquad H_{1n} : A\boldsymbol{\beta} = \mathbf{c}_{1n} \tag{5.4}$$

where A is an $m \times n$ matrix with rank m and $\mathbf{c}_0$ and $\mathbf{c}_{1n}$ are vectors.

By this result (McCullagh (1983)), we know the test statistic

$$T_n = n\left(A\boldsymbol{\beta} - \mathbf{c}_0\right)^T \left(A\widehat{\boldsymbol{\Sigma}}^{-1}A^T\right)^{-1} \left(A\boldsymbol{\beta} - \mathbf{c}_0\right) \tag{5.5}$$

has an asymptotic χ^2_m distribution under the null hypothesis H_0. Here χ^2_m denotes a central χ^2 distribution with m degrees of freedom. Under the alternative hypothesis H_{1n}, T_n has an asymptotic $\chi^2_m(\nu^2)$ distribution, where $\chi^2_m(\nu^2)$ is a noncentral χ^2 distribution with m degrees of freedom and the noncentrality parameter ν^2 is a fixed real constant such that

$$n\left(\mathbf{x}_{1n} - \mathbf{x}_0\right)^T \left(A\widehat{\Sigma}^{-1}A^T\right)^{-1} \left(\mathbf{x}_{1n} - \mathbf{x}_0\right) \longrightarrow \nu^2.$$

If $T_n > \chi^2_{m;\alpha}$, then the null hypothesis H_0 in (5.4) will be rejected at level α. Here $\chi^2_{m;\alpha}$ is the $100(1-\alpha)\%$ quantile of the central to χ^2 distribution with m degrees of freedom. Furthermore, a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$ is given by

$$\left\{\boldsymbol{\beta} : n\left(\widehat{\boldsymbol{\beta}}^{*}-\boldsymbol{\beta}\right)^{T}\widehat{\boldsymbol{\Sigma}}^{-1}\left(\widehat{\boldsymbol{\beta}}^{*}-\boldsymbol{\beta}\right) \leq \chi^{2}_{\nu;1-\alpha}\right\}.$$

Chapter 6

Simulations

In order to examine the efficiency of the nonparametric estimator of covariance matrix in quasi-likelihood model estimation, four simulation studies were run with a univariate predictor variable.

Let $\widehat{\mathbf{V}}_n(\cdot) = (\widehat{\sigma}_{nst}(\cdot))_{T\times T}$ be the nonparametric estimator of unknown covariance matrix and $(\widehat{\mu}_{ks}, \widehat{\mu}_{kt})_{B_{st}}$ be the estimated value of (μ_{ks}, μ_{kt}), where the nonparametric estimator of the covariance matrix was obtained by using bandwidth matrix B_{st}. Define

$$\begin{aligned} & G\left(B_{st}, (\widehat{\mu}_{ks}, \widehat{\mu}_{kt})_{B_{st}}, \widehat{\sigma}_{nst}\,(\widehat{\mu}_{ks}, \widehat{\mu}_{kt})_{B_{st}}\right) \\ = & \left|\sum_{k=1}^{n} \frac{(y_{ks} - \widehat{\mu}_{ks})(y_{kt} - \widehat{\mu}_{kt})}{\widehat{\sigma}_{nst}\,(\widehat{\mu}_{ks}, \widehat{\mu}_{kt})} - (n-p)\right|. \end{aligned}$$

The optimal bandwidth is B^*_{st}, the minimizer of

$$G\left(B_{st}, (\widehat{\mu}_{ks}, \widehat{\mu}_{kt})_{B_{st}}, \widehat{\sigma}_{nst}\,(\widehat{\mu}_{ks}, \widehat{\mu}_{kt})_{B_{st}}\right). \tag{6.1}$$

This bandwidth selection generalizes Chiou and Müller's (1999) bandwidth selection and was developed for this problem. The selections of bandwidth in nonparametric quasi-likelihood in following four simulation studies were automatically based on the bandwidth selector (6.1).

In the first three simulations, we considered the examples which had the same marginal expectations as examples in Liang and Zeger (1986) and Sutradhar and Das (1999), satisfying $\mu_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$ for $k = 1, 2, \cdots, n$, with $\beta_0 = 1$, $\beta_1 = 1$, $n = 200$ and a total of T time points. The fourth simulation study uses longitudinal over-dispersed Poisson data. One thousand Monte Carlo simulations were run for each study to compare methods from following methods in regression estimation in terms of the bias, sample standard error (S.E.), relative efficiency (Rel. Efficiency, ratio of true sample variance of QLE to the compared method), mean square error (MSE) and relative MSE. The following abbreviations appear in the tables.

QLE The quasi-likelihood method with true covariance matrix.

NQLE The nonparametric quasi-likelihood method with unknown but smooth covariance matrix replaced by nonparametric covariance matrix estimator.

GEEar(1) The Liang-Zeger GEE method with known marginal variance functions and working correlation matrix $A(\alpha) = (a_{kj})_{T \times T}$ specified as AR(1) structure, $a_{kk} = 1$, $a_{kj} = \alpha^{|k-j|}$ for $k, j = 1, 2, \cdots, T$.

GEEma(1) The Liang-Zeger GEE method with known marginal variance functions and working correlation matrix $A(\alpha) = (a_{kj})_{T \times T}$ specified as MA(1) structure, $a_{kk} = 1$, $a_{kj} = \alpha$ if $|k - j| = 1$ for $k, j = 1, 2, \cdots, T$; otherwise $a_{kj} = 0$.

GEEexch The Liang-Zeger GEE method with known marginal variance functions and working correlation matrix $A(\alpha) = (a_{kj})_{T \times T}$ specified as exchangeable structure, $a_{kj} = \alpha$ if $k \neq j$ for $k, j = 1, 2, \cdots, T$.

GEEunst The Liang-Zeger GEE method with known marginal variance functions and working correlation structure specified as unstructured when one has no idea about correlation structure of data.

GEEfix The Liang-Zeger GEE method with known marginal variance functions and working correlation structure specified as fixed structure $R.wrong$, where $R.wrong$ is a matrix defined as follows:

$$\begin{pmatrix} 1 & 0 & -0.98 \\ 0 & 1 & 0 \\ -0.98 & 0 & 1 \end{pmatrix}.$$

Indp The GLM method or GEE with known marginal variance functions and working correlation matrix specified as independent structure.

In the first simulation study, 1000 Monte Carlo runs were created where the data had the same marginal expectations as Zeger and Liang's example (1986), satisfying $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$ for $k = 1, 2, \cdots, n$, with $\mathrm{Var}\,(Y_{kj}) = \mu_{kj}^2 + \mu_{kj} + 1$, true correlation matrix $A(\alpha) = (a_{kj})_{T\times T}$ having AR(1) structure with $\alpha = -0.7$ and total time points $T = 5$. According to Sutradhar and Das (1999), the efficiencies of the GEE estimators $\widehat{\beta}_{0G}$, and $\widehat{\beta}_{1G}$ specifying the incorrect working correlation structure, such as exchangeable correlation structure, were the same as the efficiencies of the estimators $\widehat{\beta}_{0I}$, and $\widehat{\beta}_{1I}$ specifying the independent structure as working correlation structure. They were 71% and 73% of the efficiencies of the estimators with the correctly specified correlation structure.

Table 6.1 displays the results of comparison of four methods of covariance/correlation matrix estimation/specifications in terms of sample standard error, bias and mean square error etc.

Table 6.1: Simulation Results of the Estimated Regression Parameters for $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$, $(\beta_0, \beta_1) = (1, 1)$, $T = 5$, $n = 200$ and AR(1) as True Correlation Structure.

Method	Bias	S.E.	Rel. Efficiency	MSE	Relative MSE
	Estimation of Intercept ($\beta_0 = 1$)				
QLE	-0.00795	0.08277	1.00000	0.00691	1.00000
NQLE	-0.00779	0.08509	0.94611	0.00729	1.05606
GEEar(1)	-0.00782	0.08498	0.94853	0.00728	1.05346
Indp	-0.01118	0.09840	0.70755	0.00980	1.41846
	Estimation of Slope ($\beta_1 = 1$)				
QLE	0.01489	0.13872	1.00000	0.01945	1.00000
NQLE	0.01519	0.1413	0.96378	0.02018	1.03761
GEEar(1)	0.01468	0.14211	0.95279	0.02039	1.04867
Indp	0.01815	0.17193	0.65096	0.02986	1.53562

The results in Table 6.1 show that compared to QLE, NQL performed as well as GEEar(1) in the regression parameter estimation, even though GEEar(1) has the advantage of estimating with known marginal distribution and specifying the right working correlation structure. Also, NQL did better than method Indp, which had the same efficiency as GEEexch with specifying exchangable as a wrong working correlation structure (Sutradhar and Das (1999)). The relative efficiency of NQLE are about 95% and 96% for β_0 and β_1, respectively, compared to 70% and 65% for Indp, which confirms the results in Sutradhar and Das (1999).

Table 6.2 compares confidence intervals obtained from these estimation methods. The intervals were based on estimated asymptotic standard errors. The empirical coverage frequencies and average lengths for 90% and 95% intervals for each regression parameter derived from (5.1) are listed in Table 6.2. From Table 6.2, one can see that three methods QLE, NQLE and GEEar(1) performed well and their performances are pretty similar, while Indp method performed rather poorly, especially on estimating β_1.

In the second simulation study, we again considered the same example as the first simulation, but chose exchangeable as the true correlation structure, with $\alpha = 0.49$ and time points $t = 1, \cdots, 10$. The results of methods NQLE, GEEma(1) and GEEexch are in Table 6.2. In this simulation study, 30 out of 1000 Monte Carlo runs were not convergent when MA(1) was specified as the working correlation structure, while NQL and GEEexch did converge to reasonable values in all runs. Excluding the 30 runs, we have the results of GEEma(1) in Table 6.3. Table 6.3 shows that the efficiency of specifying the wrong working correlation structure, such as MA(1) correlation structure, is worse than specifying the correlation structure as independent. The efficiencies of regression estimators were only 75% and 51% for intercept and slope, respectively. Table 6.3 presents the efficiencies of regression estimation of the three following methods of covariance/correlation matrix specifications in terms of sample standard error, bias, mean square error etc.

Table 6.3 shows that the performance of NQLE is almost as good as GEEexch, given that GEEexch uses the correct marginal variance functions and specifies right correlation structure as working correlation structure. GEEma(1) did poorly since it specified the wrong correlation structure MA(1), while exchangeable is true correlation structure.

From Table 6.4, we also can see that the two methods NQLE and GEEexch

Table 6.2: Coverage of Confidence Intervals for $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$, True Value $(\beta_0, \beta_1) = (1, 1)$, $T = 5$, $n = 200$ and AR(1) as True Correlation Structure

	90 % Confidence Interval			95 % Confidence Interval		
	% Miss			% Miss		
Method	Left	Right	Length	Left	Right	Length
	Estimation of Intercept ($\beta_0 = 1$)					
QLE	4.85	5.35	0.27148	1.58	2.08	0.32445
NQLE	5.25	4.75	0.27910	1.58	2.28	0.33356
GEEar(1)	5.35	4.85	0.27875	1.68	2.48	0.33314
GEEindp	5.54	5.74	0.32274	2.08	2.08	0.38572
	Estimation of Slope ($\beta_1 = 1$)					
QLE	4.85	5.35	0.45499	2.67	2.28	0.54377
NQLE	4.95	5.54	0.46346	2.08	2.67	0.55389
GEEar(1)	3.96	5.35	0.46613	2.18	2.28	0.55708
GEEindp	3.66	7.82	0.56393	1.39	1.78	0.67397

Table 6.3: Simulation Results of the Estimated Regression Parameters for $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$, True Value $(\beta_0, \beta_1) = (1, 1),\ T = 10,\ n = 200$ and Exchangeable as True Correlation Structure.

Method	Bias	S.E.	Rel. Efficiency	MSE	Relative MSE
	Estimation of Intercept ($\beta_0 = 1$)				
NQLE	-0.00137	0.11936	0.89073	0.01423	1.12252
GEEma(1)	-0.00469	0.15755	0.51124	0.02482	1.95738
GEEexch	-0.00189	0.11265	1.00000	0.01268	1.00000
	Estimation of Slope ($\beta_1 = 1$)				
NQLE	0.00896	0.13837	0.92000	0.01921	1.08900
GEEma(1)	0.01192	0.23265	0.32544	0.05421	3.07313
GEEexch	0.00669	0.13272	1.00000	0.01764	1.00000

Table 6.4: Coverage of Confidence Intervals for $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$, True Value $(\beta_0, \beta_1) = (1, 1)$, $T = 5$, $n = 200$ and Exchangable as True Correlation Structure

	90 % Confidence Interval			95 % Confidence Interval		
	% Miss			% Miss		
Method	Left	Right	Length	Left	Right	Length
	Estimation of Intercept ($\beta_0 = 1$)					
NQLE	5.1	4.8	0.39149	2.4	2.6	0.46788
GEEma(1)	4.8	4.4	0.51678	2.4	3.2	0.61761
GEEexch	4.8	4.4	0.36948	2.3	2.3	0.44157
	Estimation of Slope ($\beta_1 = 1$)					
NQLE	6.1	5.2	0.45385	2.5	2.8	0.54241
GEEma(1)	4.3	4.1	0.76309	2.8	2.5	0.91198
GEEexch	6.2	4.6	0.43533	2.9	2.0	0.52027

had good performance in inference.

The following tables display the results of the third simulation study. In this simulation study, we increased the sample size from $n = 200$ in last simulation study to $n = 800$.

Table 6.5: Simulation Results of the Estimated Regression Parameters for $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$, True Value $(\beta_0, \beta_1) = (1, 1)$, $T = 10$, $n = 800$ and Exchangeable as True Correlation Structure.

Method	Bias	S.E.	Rel. Efficiency	MSE	Relative MSE
	Estimation of Intercept ($\beta_0 = 1$)				
QLE	0.00196	0.05303	1.00000	0.00281	1.00000
NQLE	0.00205	0.05382	0.97092	0.00290	1.03004
GEEunst	0.00272	0.05505	0.92791	0.00304	1.07884
GEEma(1)	0.00247	0.07834	0.45824	0.00614	2.18142
GEEexch	0.00247	0.05343	0.98509	0.00286	1.01591
	Estimation of Slope ($\beta_1 = 1$)				
QLE	-0.00492	0.06712	1.00000	0.00452	1.00000
NQLE	-0.00554	0.06799	0.97465	0.00465	1.02731
GEEunst	-0.00691	0.07239	0.85972	0.00528	1.16749
GEEma(1)	-0.00600	0.12911	0.27027	0.01669	3.68812
GEEexch	-0.00591	0.06785	0.97872	0.00463	1.02400

Table 6.5 shows that the performance of NQLE is better as the sample size increases compared to the results from last simulation study. In particular, as the sample size increases by a factor of four, the efficiency relative to GEEexch for estimating intercept increased from 89% to 98.6%, and from 92% to 99.6% for estimating slope, even though GEEexch uses the true marginal variance functions and specifies the right correlation structure as working correlation structure.

Table 6.6: Coverage of Confidence Intervals for $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $x_{kj} = j/T$, True Value $(\beta_0, \beta_1) = (1, 1)$, $T = 10$, $n = 800$ and Exchangable as True Correlation Structure

	90 % Confidence Interval			95 % Confidence Interval		
	% Miss			% Miss		
Method	Left	Right	Length	Left	Right	Length
	Estimation of Intercept ($\beta_0 = 1$)					
QLE	4.1	5.6	0.17395	2.5	3.3	0.20789
NQLE	4.3	5.7	0.17653	2.3	3.5	0.21098
GEEunst	3.8	6.0	0.18058	2.2	3.6	0.21581
GEEma(1)	3.1	3.9	0.25696	1.7	2.4	0.30710
GEEexch	3.5	5.7	0.17526	2.1	3.6	0.20945
	Estimation of Slope ($\beta_1 = 1$)					
QLE	5.0	4.4	0.22015	2.5	2.7	0.26311
NQLE	4.7	4.2	0.22300	2.3	2.5	0.26651
GEEunst	4.9	4.6	0.23743	2.8	2.2	0.28376
GEEma(1)	3.3	3.1	0.42347	2.2	2.0	0.50610
GEEexch	5.2	4.4	0.22253	2.6	2.4	0.26595

In the fourth simulation study, the underlying longitudinal data marginal distributions were over-dispersed Poisson. In this simulation study, longitudinal over-dispersed Poisson data $\{y_{kj}, x_{kj}\}$ were generated via a Gamma-Poisson mixture with sample size $n = 300$ and $T = 3$ time points as follows. Suppose that U_k is Gamma distributed with expectation 1 and variance $\tau = 0.8$ and that U_k is independent of x_{kj}, where $k = 1, \cdots, 300$ and $j = 1, 2, 3$. The link function is $\eta_{kj} = \log \mu_{kj}$, and $\eta_{kj} = \beta_0 + \beta_1 x_{kj}$ with $\beta_0 = 1$ and $\beta_1 = 0.5$. Given U_k, $y_{kj} \sim$ Poisson$(U_k \mu_{kj})$. Therefore Var$\,(y_{kj}) = \mu_{kj}(1 + \tau\mu_{kj})$ and Cov$(y_{ks}, y_{kt}) = \tau\mu_{ks}\mu_{kt}$. The design points x_{kj} were drawn from $D_x = \{Tp \times n$ random numbers generated from $unif(0, 0.5)\}$ in the first run and then fixed for the remainder of runs, $k = 1, \cdots, n = 300; j = 1, 2, 3$.

The results of the fourth simulation study are displayed in following Table 6.7 and Table 6.8.

From Table 6.7, the NQLE performed better than GEEfix, and GEEfix performed worse than NQLE and Indp. NQLE and Indp had almost same performance in prediction of intercept and slope, given that Indp method was run using the correct marginal variance function. But NQLE is best in terms of efficiency (smallest S.E and MSE) compared to GEEfix and Indp. Because the working correlation was misspecified, GEEfix is less efficient than NQLE and much worse than Indp, which confirms the conclusions of Sutradhar and Das (1999).

Table $6-8$ shows how well the asymptotic approximations made for the inference obtained from different three estimate methods. The empirical coverage frequencies and average lengths for 90% and 95% intervals for each regression parameter are derived from (5.1).

From the table 6.8, one can see that NQLE did best in the inference while GEEfix

Table 6.7: Simulation Results of the Estimated Regression Parameters for Longitudinal Over-dispersed Poisson Data with True Value $(\beta_0, \beta_1) = (1, 0.5)$, $T = 3$ and $n = 300$

Method	Bias	S.E.	Rel. Effeciency	MSE	Relative MSE
	Estimation of Intercept ($\beta_0 = 1$)				
NQLE	-0.00701	0.09096	1.00000	0.00830	1.00000
GEEfix	0.00977	0.14436	0.39702	0.02088	2.51538
Indp	0.00094	0.09391	0.93813	0.00880	1.05975
	Estimation of Slope ($\beta_1 = 0.5$)				
NQLE	-0.02785	0.20678	1.00000	0.04342	1.00000
GEEfix	-0.05832	0.45428	0.20719	0.20924	4.81853
Indp	-0.02301	0.27164	0.57948	0.07413	1.70708

Table 6.8: Coverage of Confidence Intervals for Longitudinal Over-dispersed Poisson Data with True Value $(\beta_0, \beta_1) = (1, 0.5)$, $T = 3$ and $n = 300$

	90 % Confidence Interval			95 % Confidence Interval		
	% Miss			% Miss		
Method	Left	Right	Length	Left	Right	Length
	Estimation of Intercept ($\beta_0 = 1$)					
NQLE	2.28	6.09	0.29836	1.52	2.79	0.35657
GEEfix	4.82	4.06	0.47352	2.54	2.28	0.56591
Indp	6.09	4.57	0.30804	2.79	1.02	0.36815
	Estimation of Slope ($\beta_1 = 0.5$)					
NQLE	4.57	4.06	0.67824	2.54	1.02	0.81058
GEEfix	4.82	5.58	1.49003	2.03	2.03	1.78076
Indp	3.55	4.82	0.89097	1.52	2.28	1.06482

did the worst in terms of the lengths of confidence intervals.

Chapter 7

Proofs of Theorems.

This chapter contains detailed proofs of the asymptotic theorems which are described in Chapter 5. Throughout, we refer the condition $(N1)-(N7)$ and $(K1)-(K3)$ of Chapter 4 and $(K4)$ in the statement of Theorem 5.3.

Let

$$F_t(t) = \int_{-\infty}^{t}\left(\int_{-\infty}^{\infty} f(x,y)\,dx\right)\,dy$$
$$F_s(s) = \int_{-\infty}^{s}\left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)\,dx$$

where f is the same as in $(N4)$ in Chapter 4. Then F_t^{-1}, F_s^{-1} exist, and the μ_{kt} are chosen to satisfy

$$F_t\left(\mu_{kt}\right) = \int_{-\infty}^{\mu_{kt}}(\int_{-\infty}^{\infty} f(x,y)dx)\,dy = \frac{k-1}{n-1},$$
$$F_s\left(\mu_{ks}\right) = \int_{-\infty}^{\mu_{ks}}(\int_{-\infty}^{\infty} f(x,y)dy)\,dx = \frac{k-1}{n-1}.$$

So then, we have

$$\frac{dF_t}{dt} = f_t(t) = \int_{-\infty}^{\infty} f(x,t)dx, \tag{7.1}$$
$$\frac{dF_s}{ds} = f_s(s) = \int_{-\infty}^{\infty} f(s,y)dy; \tag{7.2}$$

and

$$F_t^{-1}\left(\frac{k-1}{n-1}\right) = \mu_{kt},$$
$$F_s^{-1}\left(\frac{k-1}{n-1}\right) = \mu_{ks}.$$

Throughout the rest of the Chapter, assume that we have the following conditions.

1. The bandwidth matrix $\mathbf{B} = \operatorname{diag}\{h_s, h_t\}$ satisfies $(K3)$ in Chapter 4;
2. There exist two constants $C1$, $C2$, such that if

$$C_n(u_s, u_t) = \sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}},$$

   then

$$C1 \le \frac{C_n(u_s, u_t)}{n^2 h_s h_t} \le C2 \tag{7.3}$$

   for $(u_s, u_t) \in D_{st}$.

By two dimensional Riemann sum approximation and $(K1)-(K3)$, we have

$$\begin{aligned}
&F_{npq}(u_s, u_t) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} K_{\mathbf{B}}\left((\mu_{ks}, \mu_{kt})^T - (u_s, u_t)^T\right)(\mu_{ks}-u_s)^p(\mu_{kt}-u_t)^q \\
&= \frac{1}{n^2|B|}\sum_{k=1}^{n} K\left(\mathbf{B}^{-1}((\mu_{ks}, \mu_{kt})^T - (u_s, u_t)^T)\right)(\mu_{ks}-u_s)^p(\mu_{kt}-u_t)^q \\
&= \frac{1}{n^2 h_s h_t}\sum_{k=1}^{n} K\left(\frac{\mu_{ks}-u_s}{h_s}, \frac{\mu_{kt}-u_t}{h_t}\right)(\mu_{ks}-u_s)^p(\mu_{kt}-u_t)^q \\
&= \frac{h_s^p h_t^q}{h_s h_t}\sum_{k=1}^{n}\left[\frac{1}{n^2}K\left(\frac{F_s^{-1}\left(\frac{k-1}{n-1}\right)-u_s}{h_s}, \frac{F_t^{-1}\left(\frac{k-1}{n-1}\right)-u_t}{h_t}\right)\right. \\
&\qquad \left.\times\left(\frac{F_s^{-1}\left(\frac{k-1}{n-1}\right)-u_s}{h_s}\right)^p\left(\frac{F_t^{-1}\left(\frac{k-1}{n-1}\right)-u_t}{h_t}\right)^q\right] \\
&= h_s^p h_t^q F_{int}
\end{aligned}$$

where

$$\begin{aligned}
F_{int} &= \int_0^1 \int_0^1 \frac{1}{h_s h_t} \left[K\left(\frac{F_s^{-1}(x) - u_s}{h_s}, \frac{F_t^{-1}(y) - u_t}{h_t} \right) \right. \\
&\qquad \times \left(\frac{F_s^{-1}(x) - u_s}{h_s} \right)^p \left(\frac{F_t^{-1}(y) - u_t}{h_t} \right)^q dxdy \\
&\qquad \left. + O(1/n^2) \right].
\end{aligned}$$

Let

$$\begin{aligned}
F_s^{-1}(x) &= u, \\
F_t^{-1}(y) &= v;
\end{aligned}$$

and

$$\begin{aligned}
\left(F_s^{-1}(x) - u_s \right) / h_s &= \xi, \\
\left(F_t^{-1}(x) - u_t \right) / h_t &= \eta.
\end{aligned}$$

Then

$$dx = F_s'(u)du = f_s(u_s + \xi h_s) h_s d\xi,$$

$$dy = F_t'(v)dv = f_t(u_t + \eta h_t) h_t d\eta.$$

Therefore,

$$\begin{aligned}
F_{int} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{h_s h_t} \left[K(\xi, \eta) \xi^p \eta^q f_s(u_s + \xi h_s) f_t(u_t + \eta h_t) h_s h_t d\xi d\eta \right. \\
&\qquad \left. + O(1/n^2) \right] \\
&= \int \int K(\xi, \eta) \xi^p \eta^q \left[f_s(u_s) + \xi h_s f_s'(u_s) + \xi^2 O(h_s^2) \right] \\
&\qquad \times \left[f_t(u_t) + \eta h_t f_t'(u_t) + \eta^2 O(h_t^2) \right] d\xi d\eta + O\left(\frac{h_s^{-1} h_t^{-1}}{n^2} \right) \\
&= \int \int K(\xi, \eta) \xi^p \eta^q \left[(A1) + (A2) + (A3) \right] d\xi d\eta + O\left(\frac{h_s^{-1} h_t^{-1}}{n^2} \right)
\end{aligned}$$

where

$$
\begin{aligned}
(A1) &= f_s(u_s)f_t(u_t) + \eta h_t f_t'(u_t) f_s(u_s) + f_s(u_s)\eta^2 O(h_t^2) \\
(A2) &= \xi h_s f_s'(u_s) f_t(u_t) + \xi\eta h_s h_t f_s'(u_s) f_t'(u_t) + \xi\eta^2 h_s f_s'(u_s) O(h_t^2) \\
(A3) &= \xi^2 O(h_s^2) f_t(u_t) + \xi^2 O(h_s^2)\eta h_t f_t'(u_t) + \xi^2\eta^2 O(h_s^2 h_t^2)
\end{aligned}
$$

By (K2), we have the following results:

1. If $p+q$ is odd, then

$$
\begin{aligned}
F_{int} &= \alpha_{p,q+1} f_s(u_s) f_t'(u_t) h_t + \alpha_{p+1,q} f_s'(u_s) f_t(u_t) h_s \\
&\quad + \alpha_{p+1,q+2} f_s'(u_s) O(h_s h_t^2) + \alpha_{p+2,q+1} f_t'(u_t) O(h_s^2 h_t) \\
&\quad + O\left(\frac{h_s^{-1} h_t^{-1}}{n^2}\right)
\end{aligned}
$$

2. If $p+q$ is even, then

$$
\begin{aligned}
F_{int} &= \alpha_{p,q} f_s(u_s) f_t(u_t) + \alpha_{p+1,q+1} f_s'(u_s) f_t'(u_t) h_s h_t \\
&\quad + \alpha_{p,q+2} f_s(u_s) O(h_t^2) + \alpha_{p+2,q} f_t(u_t) O(h_s^2) \\
&\quad + \alpha_{p+2,q+2} O(h_s^2 h_t^2) + O\left(\frac{h_s^{-1} h_t^{-1}}{n^2}\right)
\end{aligned}
$$

Hence, we have proved the following lemma.

Lemma 7.1 *Under (K2),*

(i) *if $p+q$ is odd,*

$$
\begin{aligned}
F_{npq} &= \alpha_{p,q+1} f_s(u_s) f_t'(u_t) h_s^p h_t^{q+1} + \alpha_{p+1,q} f_s'(u_s) f_t(u_t) h_s^{p+1} h_t^q \\
&\quad + O\left(h_s^{p+1} h_t^{q+2}\right) + O\left(h_s^{p+2} h_t^{q+1}\right) \\
&\quad + O\left(\frac{h_s^{p-1} h_t^{q-1}}{n^2}\right);
\end{aligned}
$$

(ii) *if $p+q$ is even,*

$$
\begin{aligned}
F_{npq} &= \alpha_{p,q} f_s(u_s) f_t(u_t) h_s^p h_t^q + \alpha_{p+1,q+1} f_s'(u_s) f_t'(u_t) h_s^{p+1} h_t^{q+1} \\
&\quad + O\left(h_s^p h_t^{q+2}\right) + O\left(h_s^{p+2} h_t^q\right) \\
&\quad + O\left(\frac{h_s^{p-1} h_t^{q-1}}{n^2}\right),
\end{aligned}
$$

where f_s and f_t are defined in (7.2) *and* 7.1.

Lemma 7.2 *For $(u_s, u_t) \in D_{st}$, F_N, F_D and W_{nk} have the following explicit expressions:*

$$
\begin{aligned}
F_N &= (\alpha_{02}\alpha_{20} - \alpha_{11}^2) f_s^2 f_t^2 h_s^2 h_t^2 + O(h_s^2 h_t^3) + O(h_s^3 h_t^2); \\
F_D &= (\alpha_{02}\alpha_{20} - \alpha_{11}^2) f_s^3 f_t^3 h_s^2 h_t^2 + O(h_s^3 h_t^3) + O(h_s^2 h_t^4) + O(h_s^4 h_t^2); \\
W_{nk} &= \frac{1}{n^2 h_s h_t} K\left(\frac{\mu_{ks} - u_s}{h_s}, \frac{\mu_{kt} - u_t}{h_t}\right) \frac{1}{f_s f_t} \\
&\quad + O\left(\frac{1}{n^2 h_s}\right) + O\left(\frac{1}{n^2 h_t}\right).
\end{aligned}
$$

Proof: By Lemma 7.1, we have

$$
\begin{aligned}
& F_{n20} F_{n02} - F_{n11}^2 \\
&= (\alpha_{02}\alpha_{20} - \alpha_{11}^2) f_s^2 f_t^2 h_s^2 + O(h_s^3 h_t^3) + O(h_s^2 h_t^4) + O(h_s^4 h_t^2)
\end{aligned}
$$

$$
F_{n10} F_{n02} - F_{n01} F_{n11} = O(h_s^2 h_t^3) + O(h_s^3 h_t^2)
$$

$$
F_{n10} F_{n11} - F_{n01} F_{n20} = O(h_s^2 h_t^3) + O(h_s^3 h_t^2).
$$

Then

$$
F_N = (\alpha_{02}\alpha_{20} - \alpha_{11}^2) f_s^2 f_t^2 h_s^2 h_t^2 + O(h_s^2 h_t^3) + O(h_s^3 h_t^2).
$$

Similarly,

$$
F_D = (\alpha_{02}\alpha_{20} - \alpha_{11}^2) f_s^3 f_t^3 h_s^2 h_t^2 + O(h_s^3 h_t^3) + O(h_s^2 h_t^4) + O(h_s^4 h_t^2).
$$

Therefore

$$\begin{aligned} W_{nk} &= \frac{1}{n^2 h_s h_t} K\left(\frac{\mu_{ks}-u_s}{h_s}, \frac{\mu_{kt}-u_t}{h_t}\right) \frac{F_N}{F_D} \\ &= \frac{1}{n^2 h_s h_t} K\left(\frac{\mu_{ks}-u_s}{h_s}, \frac{\mu_{kt}-u_t}{h_t}\right) / f_s f_t \\ &\quad + O\left(\frac{1}{n^2 h_s}\right) + O\left(\frac{1}{n^2 h_t}\right). \qquad \square \end{aligned}$$

As we mentioned in Chapter 4, the nonparametric covariance matrix estimator $\widehat{\mathbf{V}}_n = (\widehat{\sigma}(\cdot))_{T\times T}$ rather than $\mathbf{V}_n = (\sigma(\cdot))_{T\times T}$ must be used in practice, since we can only observe $\{\widehat{\boldsymbol{\mu}}_k; \widehat{\boldsymbol{\varepsilon}}_k\}$ and not $\{\boldsymbol{\mu}_k; \boldsymbol{\varepsilon}_k\}$. The following lemma will show the relationship between those two nonparametric covariance matrix estimators.

Lemma 7.3 *For* $(u_s, u_t) \in D_{st}$, *let* F_{npq} *and* $\widehat{F}_{npq}$ *be as defined in (4.12) and (4.21). Under (K1)-(K3), if* $\max_{1\le k\le n} |\widehat{\mu}_{ks} - \mu_{ks}| = O_p(1/\sqrt{n})$ *and* $\max_{1\le k\le n} |\widehat{\mu}_{kt} - \mu_{kt}| = O_p(1/\sqrt{n})$, *then*

$$\widehat{F}_{npq}(u_s, u_t) = F_{npq}(u_s, u_t) + O_p\left(\frac{h_s^{p-1} h_t^q}{\sqrt{n}}\right) + O_p\left(\frac{h_s^p h_t^{q-1}}{\sqrt{n}}\right)$$

Proof : Let

$$h = \widehat{\mu}_{ks} - \mu_{ks}, \qquad k = \widehat{\mu}_{kt} - \mu_{kt},$$

and

$$G(x,y) = K\left(\frac{x-u_s}{h_s}, \frac{y-u_t}{h_t}\right)(x-u_s)^p (y-u_t)^q.$$

Then by two dimensional Taylor expansions and the assumptions, we have

$$
\begin{aligned}
& K\left(\frac{\widehat{\mu}_{ks}-u_s}{h_s},\frac{\widehat{\mu}_{kt}-u_t}{h_t}\right)(\widehat{\mu}_{ks}-u_s)^p(\widehat{\mu}_{kt}-u_t)^q \\
& \quad -K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)(\mu_{ks}-u_s)^p(\mu_{kt}-u_t)^q \\
= & \left(h\frac{\partial G}{\partial x}+k\frac{\partial G}{\partial y}\right)(\bar{x},\bar{y}) \\
= & h(\bar{y}-u_t)^q\left\{K'_{\frac{\bar{x}-u_s}{h_s}}\left(\frac{\bar{x}-u_s}{h_s},\frac{\bar{y}-u_t}{h_t}\right)\frac{1}{h_s}(\bar{x}-u_s)^p\right. \\
& \qquad \left.+\,p(\bar{x}-u_s)^{p-1}K\left(\frac{\bar{x}-u_s}{h_s},\frac{\bar{y}-u_t}{h_t}\right)\right\} \\
& +k(\bar{x}-u_s)^p\left\{K'_{\frac{\bar{y}-u_t}{h_t}}\left(\frac{\bar{x}-u_s}{h_s},\frac{\bar{y}-u_t}{h_t}\right)\frac{1}{h_t}(\bar{y}-u_t)^q\right. \\
& \qquad \left.+\,q(\bar{y}-u_t)^{q-1}K\left(\frac{\bar{x}-u_s}{h_s},\frac{\bar{y}-u_t}{h_t}\right)\right\} \\
= & \left\{O_p\left(\frac{1}{\sqrt{n}}\right)h_t^q\left[h_s^{p-1}+ph_s^{p-1}O(1)\right]+O_p\left(\frac{1}{\sqrt{n}}\right)h_s^p\left[h_t^{q-1}+qh_t^{q-1}O(1)\right]\right\} \\
& \times I_{\{|\widehat{\mu}_{ks}-u_s|\le h_s,|\widehat{\mu}_{kt}-u_t|\le h_t\}}I_{\{|\mu_{ks}-u_s|\le h_s,|\mu_{kt}-u_t|\le h_t\}} \\
= & O_p\left(\frac{1}{\sqrt{n}}\right)\left[O(h_s^{p-1}h_t^q)+O(h_s^ph_t^{q-1})\right] \\
& \times I_{\{|\widehat{\mu}_{ks}-u_s|\le h_s,|\widehat{\mu}_{kt}-u_t|\le h_t\}}I_{\{|\mu_{ks}-u_s|\le h_s,|\mu_{kt}-u_t|\le h_t\}}
\end{aligned}
$$

where $\bar{x}\in[\mu_{ks},\widehat{\mu}_{ks}]$, and $\bar{y}\in[\mu_{kt},\widehat{\mu}_{kt}]$.

Therefore, by (7.3) we have

$$
\begin{aligned}
& \widehat{F}_{npq}(u_s,u_t)-F_{npq}(u_s,u_t) \\
& \quad = \frac{1}{n^2h_sh_t}\sum_{k=1}^n\left[K\left(\frac{\widehat{\mu}_{ks}-u_s}{h_s},\frac{\widehat{\mu}_{kt}-u_t}{h_t}\right)(\widehat{\mu}_{ks}-u_s)^p(\widehat{\mu}_{kt}-u_t)^q\right. \\
& \qquad\qquad \left.-\,K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)(\mu_{ks}-u_s)^p(\mu_{kt}-u_t)^q\right] \\
& \quad = O_p\left(\frac{1}{\sqrt{n}}\right)\left[O(h_s^{p-1}h_t^q)+O(h_s^ph_t^{q-1})\right] \\
& \qquad \times\frac{1}{n^2h_sh_t}\sum_{k=1}^n I_{\{|\widehat{\mu}_{ks}-u_s|\le h_s,|\widehat{\mu}_{kt}-u_t|\le h_t\}}I_{\{|\mu_{ks}-u_s|\le h_s,|\mu_{kt}-u_t|\le h_t\}} \\
& \quad = O_p\left(\frac{h_s^{p-1}h_t^q}{\sqrt{n}}\right)+O_p\left(\frac{h_s^ph_t^{q-1}}{\sqrt{n}}\right). \qquad \square
\end{aligned}
$$

By the result of Lemma 7.3, we have following results about the relationships among F_N, F_D and $\widehat{F}_N$, $\widehat{F}_D$.

Lemma 7.4 *For* $(u_s, u_t) \in D_{st}$*, Let* F_N *and* F_D*;* $\widehat{F}_N$ *and* $\widehat{F}_D$ *are as defined in (4.11)and (4.12); (4.19)and (4.20). Under (K1)–(K3), if* $\max_{1\le k\le n} |\widehat{\mu}_{ks} - \mu_{ks}| = O_p(1/\sqrt{n})$ *and* $\max_{1\le k\le n} |\widehat{\mu}_{kt} - \mu_{kt}| = O_p(1/\sqrt{n})$*, then*

$$\widehat{F}_D = F_D + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right),$$

$$\widehat{F}_N = F_N + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right).$$

Proof: By Lemma 7.3, we have

$$\begin{aligned}
\widehat{F}_{n20} &= F_{n20} + O_p\left(\frac{h_s}{\sqrt{n}}\right) + O_p\left(\frac{h_s^2 h_t^{-1}}{\sqrt{n}}\right), \\
\widehat{F}_{n11} &= F_{n11} + O_p\left(\frac{h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s}{\sqrt{n}}\right), \\
\widehat{F}_{n10} &= F_{n10} + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^{-1}}{\sqrt{n}}\right), \\
\widehat{F}_{n02} &= F_{n02} + O_p\left(\frac{h_s^{-1} h_t^2}{\sqrt{n}}\right) + O_p\left(\frac{h_t}{\sqrt{n}}\right), \\
\widehat{F}_{n01} &= F_{n01} + O_p\left(\frac{h_s^{-1} h_t}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right), \\
\widehat{F}_{n00} &= F_{n00} + O_p\left(\frac{h_s^{-1}}{\sqrt{n}}\right) + O_p\left(\frac{h_t^{-1}}{\sqrt{n}}\right).
\end{aligned}$$

After multiplications we have

$$
\begin{aligned}
\widehat{F}_{n00}\widehat{F}_{n20}\widehat{F}_{n02} &= F_{n00}F_{n20}F_{n02} + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right) + \left(\frac{h_s^2 h_t}{\sqrt{n}}\right), \\
\widehat{F}_{n10}\widehat{F}_{n01}\widehat{F}_{n11} &= F_{n10}F_{n01}F_{n11} + O_p\left(\frac{h_s h_t^3}{\sqrt{n}}\right) + O_p\left(\frac{h_s^2 h_t^2}{\sqrt{n}}\right) \\
&\quad + O_p\left(\frac{h_s^3 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_t^2}{n}\right) + O_p\left(\frac{h_s h_t}{n}\right), \\
\widehat{F}_{n01}^2\widehat{F}_{n20} &= F_{n01}^2 F_{n20} + O_p\left(\frac{h_s h_t^3}{\sqrt{n}}\right) + O_p\left(\frac{h_s^2 h_t^2}{\sqrt{n}}\right) + O_p\left(\frac{h_s^3 h_t}{\sqrt{n}}\right) \\
&\quad + O_p\left(\frac{h_s h_t}{n}\right) + O_p\left(\frac{h_s^2}{n}\right) + O_p\left(\frac{h_t^2}{n}\right), \\
\widehat{F}_{n10}^2\widehat{F}_{n02} &= F_{n10}^2 F_{n02} + O_p\left(\frac{h_s h_t^3}{\sqrt{n}}\right) + O_p\left(\frac{h_s^2 h_t^2}{\sqrt{n}}\right) + O_p\left(\frac{h_s^3 h_t}{\sqrt{n}}\right) \\
&\quad + O_p\left(\frac{h_s^2}{n}\right) + O_p\left(\frac{h_t^2}{n}\right) + O_p\left(\frac{h_s^3 h_t^{-1}}{n}\right), \\
\widehat{F}_{n11}^2\widehat{F}_{n00} &= F_{n11}^2 F_{n00} + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right) + O_p\left(\frac{h_s^2 h_t^2}{\sqrt{n}}\right), \\
\widehat{F}_{n20}\widehat{F}_{n02} &= F_{n20}F_{n02} + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right),
\end{aligned}
$$

$$
\begin{aligned}
\widehat{F}_{n11}^2 &= F_{n11}^2 + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right), \\
\widehat{F}_{n10}\widehat{F}_{n02} &= F_{n10}F_{n02} + O_p\left(\frac{h_s h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_t^2}{\sqrt{n}}\right), \\
\widehat{F}_{n01}\widehat{F}_{n11} &= F_{n01}F_{n11} + O_p\left(\frac{h_t^2}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t}{\sqrt{n}}\right), \\
\widehat{F}_{n10}\widehat{F}_{n11} &= F_{n10}F_{n11} + O_p\left(\frac{h_s^2}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t}{\sqrt{n}}\right), \\
\widehat{F}_{n01}\widehat{F}_{n20} &= F_{n01}F_{n20} + O_p\left(\frac{h_s^2}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t}{\sqrt{n}}\right).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&(\widehat{\mu}_{ks} - u_s) I_{\{|\widehat{\mu}_{ks} - u_s| \le h_s\}} (\widehat{F}_{n10}\widehat{F}_{n02} - \widehat{F}_{n01}\widehat{F}_{n11}) \\
&= (\mu_{ks} - u_s) I_{\{|\mu_{ks} - u_s| \le h_s\}} (F_{n10}F_{n02} - F_{n01}F_{n11}) + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&(\widehat{\mu}_{kt}-u_t)I_{\{|\widehat{\mu}_{kt}-u_t|\le h_t\}}(\widehat{F}_{n10}\widehat{F}_{n11}-\widehat{F}_{n01}\widehat{F}_{n20})\\
&= (\mu_{kt}-u_t)I_{\{|\mu_{kt}-u_t|\le h_t\}}(F_{n10}F_{n11}-F_{n01}F_{n20})+O_p\left(\frac{h_s^2h_t}{\sqrt{n}}\right)+O_p\left(\frac{h_sh_t^2}{\sqrt{n}}\right)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\widehat{F}_D &= \widehat{F}_{n00}\widehat{F}_{n20}\widehat{F}_{n02}+2\widehat{F}_{n10}\widehat{F}_{n01}\widehat{F}_{n11}-\widehat{F}_{n20}\widehat{F}_{n01}^2\\
&\quad -\widehat{F}_{n10}^2\widehat{F}_{n02}-\widehat{F}_{n11}^2\widehat{F}_{n00}\\
&= F_D+O_p\left(\frac{h_s^2h_t}{\sqrt{n}}\right)+O_p\left(\frac{h_sh_t^2}{\sqrt{n}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{F}_N &= (\widehat{F}_{n20}\widehat{F}_{n02}-\widehat{F}_{n11}^2)-(\widehat{\mu}_{ks}-u_s)I_{\{|\widehat{\mu}_{ks}-u_s|\le h_s\}}(\widehat{F}_{n10}\widehat{F}_{n02}-\widehat{F}_{n01}\widehat{F}_{n11})\\
&\quad -(\widehat{\mu}_{kt}-u_t)I_{\{|\widehat{\mu}_{kt}-u_t|\le h_t\}}(\widehat{F}_{n10}\widehat{F}_{n11}-\widehat{F}_{n01}\widehat{F}_{n20})\\
&= F_N+O_p\left(\frac{h_s^2h_t}{\sqrt{n}}\right)+O_p\left(\frac{h_sh_t^2}{\sqrt{n}}\right) \qquad \square
\end{aligned}
$$

Lemma 7.5 *For $(u_s,u_t)\in D_{st}$, let W_{nk} and $\widehat{W}_{nk}$ are as defined as in* (4.9) *and* (4.18)*. Then*

$$
\sup_{(u_s,u_t)\in D}\max_{1\le k\le n}\left|\widehat{W}_{nk}-W_{nk}\right| = O_p\left(\frac{1}{n^2\sqrt{n}h_s^2h_t}\right)+O_p\left(\frac{1}{n^2\sqrt{n}h_sh_t^2}\right)
$$

Proof: Let

$$
H_N = K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)F_N, \qquad \widehat{H}_N = K\left(\frac{\widehat{\mu}_{ks}-u_s}{h_s},\frac{\widehat{\mu}_{kt}-u_t}{h_t}\right)\widehat{F}_N.
$$

Then,

$$
\begin{aligned}
\widehat{H}_N-H_N &= K\left(\frac{\widehat{\mu}_{ks}-u_s}{h_s},\frac{\widehat{\mu}_{kt}-u_t}{h_t}\right)\widehat{F}_N-K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)F_N\\
&= \left[K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)+O_p\left(\frac{1}{\sqrt{n}h_s}\right)+O_p\left(\frac{1}{\sqrt{n}h_t}\right)\right]\\
&\quad \times\left[F_N+O_p\left(\frac{h_s^2h_t}{\sqrt{n}}\right)+O_p\left(\frac{h_sh_t^2}{\sqrt{n}}\right)\right]\\
&\quad -K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)F_N\\
&= O_p\left(\frac{h_s^2h_t}{\sqrt{n}}\right)+O_p\left(\frac{h_sh_t^2}{\sqrt{n}}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sup_{(u_s,u_t)\in D} \max_{1\le k\le n} \left|\widehat{W}_{nk} - W_{nk}\right| &= \left| \frac{1}{n^2 h_s h_t} K\left(\frac{\widehat{\mu}_{ks} - u_s}{h_s}, \frac{\widehat{\mu}_{kt} - u_t}{h_t}\right) \frac{\widehat{F}_N}{\widehat{F}_D} \right. \\
&\quad \left. - \frac{1}{n^2 h_s h_t} K\left(\frac{\mu_{ks} - u_s}{h_s}, \frac{\mu_{kt} - u_t}{h_t}\right) \frac{F_N}{F_D} \right| \\
&= \frac{1}{n^2 h_s h_t} \left| \frac{\widehat{H}_N}{\widehat{F}_D} - \frac{H_N}{F_D} \right| \\
&= \frac{1}{n^2 h_s h_t} \left| \frac{H_N + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right)}{F_D + O_p\left(\frac{h_s^2 h_t}{\sqrt{n}}\right) + O_p\left(\frac{h_s h_t^2}{\sqrt{n}}\right)} - \frac{H_N}{F_D} \right| \\
&= O_p\left(\frac{1}{n^2 \sqrt{n} h_s^2 h_t}\right) + O_p\left(\frac{1}{n^2 \sqrt{n} h_s h_t^2}\right) \qquad \square
\end{aligned}
$$

Lemma 7.6 *Under* $(N1)-(N5)$, *if* $\widehat{\boldsymbol{\beta}}$ *is a* $\sqrt{n}$ *-consistent estimator of* $\boldsymbol{\beta}$ *in the sense that*

$$
\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| = O_p\left(\frac{1}{\sqrt{n}}\right),
$$

then

$$
\max_{1\le k\le n} |\widehat{\mu}_{ks} - \mu_{ks}| = O_p\left(\frac{1}{\sqrt{n}}\right),
$$

and

$$
\max_{1\le k\le n} |\widehat{\mu}_{kt} - \mu_{kt}| = O_p\left(\frac{1}{\sqrt{n}}\right).
$$

Proof: Since

$$
\widehat{\mu}_{ks} = g\left(\mathbf{x}_{ks}^T \widehat{\boldsymbol{\beta}}\right) \qquad \mu_{ks} = g\left(\mathbf{x}_{ks}^T \boldsymbol{\beta}\right), \tag{7.4}
$$

By the mean value theorem, we have

$$
\begin{aligned}
&\max_{1\le k\le n} |\widehat{\mu}_{ks} - \mu_{ks}| \\
&= |g\left(\mathbf{x}_{ks}^T \widehat{\boldsymbol{\beta}}\right) - g\left(\mathbf{x}_{ks}^T \boldsymbol{\beta}\right)| \\
&= O_p\left(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\right) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

Similarly, we have

$$\max_{1\le k\le n} |\widehat{\mu}_{kt} - \mu_{kt}| = O_p\left(\frac{1}{\sqrt{n}}\right). \qquad \square$$

The result of the following Lemma will be used in the proof of the asymptotic theorems.

Lemma 7.7 *Under* $(N1) - (N5)$,

$$\begin{aligned}\sup_{(u_s,u_t)\in D_{st}} \frac{1}{n^2h_sh_t} \sum_{k=1}^{n} \epsilon_{ks}\epsilon_{kt} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}} \\ = O_p\left(\left(\frac{\log n}{n^2h_sh_t}\right)^{1/2}\right) + O(1);\end{aligned}$$

Proof: Let

$$\begin{aligned} S_n & \\ &= \left\{\frac{1}{n^2h_sh_t}\sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}}\left[\epsilon_{ks}\epsilon_{kt} - \sigma_{st}(\mu_{ks},\mu_{kt})\right]\right\} \Big/ \left(\frac{\log n}{n^2h_sh_t}\right)^{1/2} \\ &= \left(\frac{n^2h_sh_t}{\log n}\right)^{1/2}\sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}}\left[\epsilon_{ks}\epsilon_{kt} - \sigma_{st}(\mu_{ks},\mu_{kt})\right]\end{aligned}$$

By (7.3) and $(N1)$ we have

$$\begin{aligned}\operatorname{Var} S_n &= \left(\frac{n^2h_sh_t}{\log n}\right)\left(\frac{1}{n^2h_sh_t}\right)^2\sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}}\operatorname{Cov}(\epsilon_{ks},\epsilon_{kt}) \\ &= O\left(\frac{1}{\log n}\right) \to 0 \quad as \quad n\to\infty\end{aligned}$$

By Chebyshev's Inequality, for any $M > 0$

$$P\left[S_n > M\right] \le \frac{1}{M^2}\operatorname{Var} S_n \to 0 \quad as \quad n\to\infty$$

Therefore

$$\begin{aligned}\frac{1}{n^2h_sh_t}\sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}}\left[\epsilon_{ks}\epsilon_{kt} - \sigma_{st}(\mu_{ks},\mu_{kt})\right] \\ = O_p\left(\left(\frac{\log n}{n^2h_sh_t}\right)^{1/2}\right)\end{aligned}$$

It is obvious that

$$\frac{1}{n^2 h_s h_t} \sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}} \sigma_{st}(\mu_{ks}, \mu_{kt}) = O(1)$$

Hence

$$\begin{aligned}
&\sup_{(u_s,u_t)\in D_{st}} \frac{1}{n^2 h_s h_t} \sum_{k=1}^{n} \epsilon_{ks}\epsilon_{kt} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}} \\
= & \frac{1}{n^2 h_s h_t} \sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}} \left[\epsilon_{ks}\epsilon_{kt} - \sigma_{st}(\mu_{ks}, \mu_{kt})\right] \\
& + \frac{1}{n^2 h_s h_t} \sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\le h_s, |\mu_{kt}-u_t|\le h_t\}} \sigma_{st}(\mu_{ks}, \mu_{kt}) \\
= & O_p\left(\left(\frac{\log n}{n^2 h_s h_t}\right)^{1/2}\right) + O(1). \qquad \square
\end{aligned}$$

Next, we are ready to prove the asymptotic theorems stated in Chapter 5.

Proof of Theorem 5.1: To prove (i), we use Taylor expansion as follows:

$$\begin{aligned}
&[\sigma_{st}(\mu_{ks}, \mu_{kt}) - \sigma_{st}(u_s, u_t)] \\
= & \frac{\partial \sigma_{st}}{\partial u_s}(u_s, u_t)(\mu_{ks} - u_s) + \frac{\partial \sigma_{st}}{\partial u_t}(u_s, u_t)(\mu_{kt} - u_t) \\
& + (\mu_{ks} - u_s)^2 \frac{\partial^2 \sigma_{st}}{\partial {u_s}^2}(\xi_k, \eta_k) \\
& + 2(\mu_{ks} - u_s)(\mu_{kt} - u_t) \frac{\partial^2 \sigma_{st}}{\partial u_s \partial u_t}(\xi_k, \eta_k) \\
& + (\mu_{kt} - u_t)^2 \frac{\partial^2 \sigma_{st}}{\partial {u_t}^2}(\xi_k, \eta_k).
\end{aligned}$$

Then

$$
\begin{aligned}
&|E\sigma_{nst}(u_s,u_t)-\sigma_{st}(u_s,u_t)| \\
&= \left|E\left(\sum_{k=1}^{n} W_{nk}(u_s,u_t)\epsilon_{ks}\epsilon_{kt}\right)-\sigma_{st}(u_s,u_t)\right| \\
&= \left|\sum_{k=1}^{n} W_{nk}(u_s,u_t)E\left(\epsilon_{ks}\epsilon_{kt}\right)-\sigma_{st}(u_s,u_t)\right| \\
&= \left|\sum_{k=1}^{n} W_{nk}(u_s,u_t)\left[\sigma_{st}(\mu_{ks},\mu_{kt})-\sigma_{st}(u_s,u_t)\right]\right| \\
&\leq \sum_{k=1}^{n} |W_{nk}(u_s,u_t)| \\
&\quad \times \left|\left[\frac{\partial^2\sigma}{\partial {u_s}^2}(\xi_k,\eta_k)-\frac{\partial^2\sigma}{\partial {u_s}^2}(u_s,u_t)\right](\mu_{ks}-u_s)^2+O(h_s^2)\right| \\
&\quad + \sum_{k=1}^{n} |W_{nk}(u_s,u_t)| \\
&\quad \times \left|\left[\frac{\partial^2\sigma}{\partial {u_t}^2}(\xi_k,\eta_k)-\frac{\partial^2\sigma}{\partial {u_t}^2}(u_s,u_t)\right](\mu_{kt}-u_t)^2+O(h_t^2)\right| \\
&= O(h_s^2)+O(h_t^2).
\end{aligned}
$$

Therefore, we have

$$
\sup_{(u_s,u_t)\in D_{st}} |E\sigma_{nst}(u_s,u_t)-\sigma_{st}(u_s,u_t)| = O\left(h_s^2\right)+O\left(h_t^2\right).
$$

Notice that

$$
\begin{aligned}
\mathrm{Var}\left(\sigma_{nst}\right) &= \sum_{k=1}^{n} W_{nk}^2 \mathrm{Cov}\left(\epsilon_{ks}\epsilon_{kt}\right) \\
&= \sum_{k=1}^{n} W_{nk}^2 \sigma_{st}(\mu_{ks},\mu_{kt}) \\
&= \sum_{k=1}^{n} \left\{\left[\frac{1}{n^2h_sh_t}K\left(\frac{\mu_{ks}-u_s}{h_s},\frac{\mu_{kt}-u_t}{h_t}\right)I_{\{|\mu_{ks}-u_s|\leq h_s,|\mu_{kt}-u_t|\leq h_t\}}/f_sf_t\right.\right. \\
&\quad \left.\left. +\,O\left(\frac{1}{n^2h_s}\right)+O\left(\frac{1}{n^2h_t}\right)\right]^2 \sigma_{st}(\mu_{ks},\mu_{kt})\right\} \\
&= O\left(\frac{1}{n^2h_sh_t}\right).
\end{aligned}
$$

By (K1)-(K2) and using the same process as in the proof in Lemma 7.7, we can prove (ii).

From (i), we can prove (iii) as follows:

$$\begin{aligned}
E((\sigma_{nst}-\sigma_{st})^2) &= E\left(\sigma_{nst}^2-2\sigma_{nst}\sigma_{st}+\sigma_{st}^2\right)\\
&= E\sigma_{nst}^2-2\sigma_{st}E\sigma_{nst}+\sigma_{st}^2\\
&= O\left(\frac{1}{n^2h_sh_t}+h_s^2h_t^2+h_s^4+h_t^4\right).
\end{aligned}$$

Using (i) and (ii), it follows that

$$\begin{aligned}
&\sup_{(u_s,u_t)\in D_{st}}|\sigma_{nst}-\sigma_{st}|\\
&\leq \sup_{(u_s,u_t)\in D_{st}}|\sigma_{nst}-E\sigma_{nst}|+\sup_{(u_s,u_t)\in D_{st}}|E\sigma_{nst}-\sigma_{st}|\\
&\leq \sup_{(u_s,u_t)\in D_{st}}|\sum_{k=1}^{n}W_{nk}\left(\epsilon_{ks}\epsilon_{kt}-E(\epsilon_{ks}\epsilon_{kt})\right)|+O\left(h_t^2\right)\\
&= O_p\left(\left[\frac{\log n}{n^2h_sh_t}\right]^{1/2}+h_s^2+h_t^2\right). \qquad \square
\end{aligned}$$

Proof of Theorem 5.2: First we have:

$$\begin{aligned}
&\sup_{(u_s,u_t)\in D_{st}} |\widehat{\sigma}_{nst} - \sigma_{nst}| \\
= &\sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left[W_{nk} - \left(\widehat{W_{nk}} - W_{nk} \right) \right] \left[\epsilon_{ks} + (\widehat{\epsilon}_{ks} - \epsilon_{ks}) \right] \left[\epsilon_{kt} + (\widehat{\epsilon}_{kt} - \epsilon_{kt}) \right] \right. \\
&\qquad \left. - \sigma_{nst} \right| \\
\le &\sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk} \right) \epsilon_{ks}\epsilon_{kt} \right| + \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} W_{nk} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks} \right) \epsilon_{kt} \right| \\
&+ \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk} \right) \left(\widehat{\epsilon}_{ks} - \epsilon_{ks} \right) \epsilon_{kt} \right| \\
&+ \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} W_{nk} \left(\widehat{\epsilon}_{kt} - \epsilon_{kt} \right) \epsilon_{ks} \right| \\
&+ \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk} \right) \left(\widehat{\epsilon}_{kt} - \epsilon_{kt} \right) \epsilon_{ks} \right| \\
&+ \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} W_{nk} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks} \right) \left(\widehat{\epsilon}_{kt} - \epsilon_{kt} \right) \right| \\
&+ \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk} \right) \left(\widehat{\epsilon}_{ks} - \epsilon_{ks} \right) \left(\widehat{\epsilon}_{kt} - \epsilon_{kt} \right) \right| \\
= &\; I + II + III + IV + V + VI + VII.
\end{aligned}$$

By (K3) and (7.3), we have

$$\begin{aligned}
&\sum_{k=1}^{n} |W_{nk}| \\
&\quad = \sum_{k=1}^{n} \frac{1}{n^2 h_s h_t} \left| K \left(\frac{\mu_{ks} - u_s}{h_s}, \frac{\mu_{kt} - u_t}{h_t} \right) I_{\{|\mu_{ks}-u_s| \le h_s, |\mu_{kt}-u_t| \le h_t\}} / f_s f_t \right. \\
&\qquad\qquad \left. + O\left(\frac{1}{n^2 h_s} \right) + O\left(\frac{1}{n^2 h_t} \right) \right| \\
&\quad = O(1).
\end{aligned}$$

Since

$$\begin{aligned} \widehat{\epsilon}_{ks} - \epsilon_{ks} &= y_{ks} - \widehat{\mu}_{ks} - (y_{ks} - \mu_{ks}) \\ &= \mu_{ks} - \widehat{\mu}_{ks}, \end{aligned}$$

then

$$|\widehat{\epsilon}_{ks} - \epsilon_{ks}| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$|\widehat{\epsilon}_{kt} - \epsilon_{kt}| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

By (K4) and Lemma 7.7,

$$\begin{aligned} I &= \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right) \epsilon_{ks}\epsilon_{kt} \right| \\ &\leq \sup_{(u_s,u_t)\in D_{st}} \max_{1\leq k\leq n} \left|\widehat{W}_{nk} - W_{nk}\right| \left| \sum_{k=1}^{n} \epsilon_{ks}\epsilon_{kt} I_{\{|\mu_{ks}-u_s|\leq h_s, |\mu_{kt}-u_t|\leq h_t\}} \right| \\ &= O\left(\frac{1}{\sqrt{n}h_s} + \frac{1}{\sqrt{n}h_t}\right)\left[O\left(\left(\frac{\log n}{n^2 h_s h_t}\right)\right) + O(1)\right] \\ &= O\left(\frac{1}{\sqrt{n}h_s} + \frac{1}{\sqrt{n}h_t}\right). \end{aligned}$$

Let $W_{nk}^{+} = \max\{0, W_{nk}\}$ and $W_{nk}^{-} = \max_{1\leq k\leq n}\{0, -W_{nk}\}$. Then $W_{nk} = W_{nk}^{+} + W_{nk}^{-}$.

By using the Cauchy-Schwarz Inequality, we have

$$
\begin{aligned}
II &= \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} W_{nk} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right) \epsilon_{kt} \right| \\
&= \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} \left(W_{nk}^{+} + W_{nk}^{-}\right) \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right) \epsilon_{kt} \right| \\
&\leq \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} W_{nk}^{+} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right)^2 \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} W_{nk}^{+} {\epsilon_{kt}}^2 \right)^{1/2} \\
&\quad + \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} W_{nk}^{-} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right)^2 \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} W_{nk}^{-} {\epsilon_{kt}}^2 \right)^{1/2} \\
&\leq \max \left(\left|\widehat{\epsilon}_{ks} - \epsilon_{ks}\right|^2 \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} W_{nk}^{+} I_{\{|\mu_{ks}-u_s|\leq h_s, |\mu_{kt}-u_t|\leq h_t\}} \right)^{1/2} \\
&\quad \times \sup_{(u_s,u_t)\in D_{st}} \left(\max_{1\leq k\leq n} W_{nk}^{+} \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} {\epsilon_{kt}}^2 I_{\{|\mu_{kt}-u_t|\leq h_t\}} \right)^{1/2} \\
&\quad + \max \left(\left|\widehat{\epsilon}_{ks} - \epsilon_{ks}\right|^2 \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} W_{nk}^{-} I_{\{|\mu_{ks}-u_s|\leq h_s, |\mu_{kt}-u_t|\leq h_t\}} \right)^{1/2} \\
&\quad \times \sup_{(u_s,u_t)\in D_{st}} \left(\max_{1\leq k\leq n} W_{nk}^{-} \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} {\epsilon_{kt}}^2 I_{\{|\mu_{kt}-u_t|\leq h_t\}} \right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{n}}\right) \sup_{(u_s,u_t)\in D_{st}} \left(\max_{1\leq k\leq n} W_{nk}^{+} \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} {\epsilon_{kt}}^2 I_{\{|\mu_{kt}-u_t|\leq h_t\}} \right)^{1/2} \\
&\quad + O_p\left(\frac{1}{\sqrt{n}}\right) \sup_{(u_s,u_t)\in D} \left(\max_{1\leq k\leq n} W_{nk}^{-} \right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} {\epsilon_{kt}}^2 I_{\{|\mu_{kt}-u_t|\leq h_t\}} \right)^{1/2} \\
&= O_p\left(\frac{1}{n\sqrt{h_t}}\right).
\end{aligned}
$$

Similarly,

$$IV = O_p\left(\frac{1}{n\sqrt{h_s}}\right),$$

since IV has the same structure as II.

By Lemma 7.6, we have

$$\begin{aligned} VI &= \sup_{(u_s,u_t)\in D_{st}} \left| \sum_{k=1}^{n} W_{nk} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right)\left(\widehat{\epsilon}_{kt} - \epsilon_{kt}\right) \right| \\ &\leq \max_{1\leq k} \left(\left|\widehat{\epsilon}_{ks} - \epsilon_{ks}\right| \left|\widehat{\epsilon}_{kt} - \epsilon_{kt}\right| \right) \sup_{(u_s,u_t)\in D} \sum_{k=1}^{n} \left|W_{nk}\right| \\ &= O_p\left(\frac{1}{n}\right). \end{aligned}$$

By Lemma 7.5, Lemma 7.7 and (7.3),

$$\begin{aligned} VII &\leq \sup_{(u_s,u_t)\in D_{st}} \sum_{k=1}^{n} \left|\widehat{W}_{nk} - W_{nk}\right| \left|\widehat{\epsilon}_{ks} - \epsilon_{ks}\right| \left|\widehat{\epsilon}_{kt} - \epsilon_{kt}\right| \\ &\leq \sup_{(u_s,u_t)\in D_{st}} \left|\widehat{W}_{nk} - W_{nk}\right| O_p\left(\frac{1}{n}\right) \sup_{(u_s,u_t)\in D_{st}} \sum_{k=1}^{n} I_{\{|\mu_{ks}-u_s|\leq h_s, |\mu_{kt}-u_t|\leq h_t\}} \\ &= O_p\left(\frac{1}{n\sqrt{n}h_s} + \frac{1}{n\sqrt{n}h_t}\right). \end{aligned}$$

By Lemma 7.5 and Lemma 7.7 and Chebyshev's Inequality,

$$
\begin{aligned}
III &= \sup_{(u_s,u_t)\in D_{st}} \sum_{k=1}^{n} \left|\widehat{W}_{nk} - W_{nk}\right| \left|\left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right)\epsilon_{kt}\right| \\
&\leq \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right)^{+} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right)^2\right)^{1/2} \\
&\quad \times \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right)^{+} \epsilon_{kt}^2\right)^{1/2} \\
&\quad + \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right)^{-} \left(\widehat{\epsilon}_{ks} - \epsilon_{ks}\right)^2\right)^{1/2} \\
&\quad \times \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right)^{-} \epsilon_{kt}^2\right)^{1/2} \\
&\leq \sup_{(u_s,u_t)\in D_{st}} \max_{1\leq k\leq n} \left(\left|\widehat{\epsilon}_{ks} - \epsilon_{ks}\right|^2\right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right)^{+} \right. \\
&\qquad \left. \times\, I_{\{|\mu_{ks}-u_s|\leq h_s, |\mu_{kt}-u_t|\leq h_t\}}\right)^{1/2} \\
&\quad \sup_{(u_s,u_t)\in D_{st}} \max_{1\leq k\leq n} \left|\left(\widehat{W}_{nk} - W_{nk}\right)^{+}\right|^{1/2} \left(\sum_{k=1}^{n} {\epsilon_{kt}}^2 I_{\{|\mu_{kt}-u_t|\leq h_t\}}\right)^{1/2} \\
&\quad + \max_{1\leq k\leq n} \left(\left|\widehat{\epsilon}_{ks} - \epsilon_{ks}\right|^2\right)^{1/2} \sup_{(u_s,u_t)\in D_{st}} \left(\sum_{k=1}^{n} \left(\widehat{W}_{nk} - W_{nk}\right)^{-} \right. \\
&\qquad \left. \times\, I_{\{|\mu_{ks}-u_s|\leq h_s, |\mu_{kt}-u_t|\leq h_t\}}\right)^{1/2} \\
&\quad \sup_{(u_s,u_t)\in D_{st}} \max_{1\leq k\leq n} \left|\left(\widehat{W}_{nk} - W_{nk}\right)^{-}\right|^{1/2} \left(\sum_{k=1}^{n} {\epsilon_{kt}}^2 I_{\{|\mu_{kt}-u_t|\leq h_t\}}\right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(\frac{1}{nh_s\sqrt{h_s}} + \frac{1}{n\sqrt{h_s}h_t}\right) \\
&\leq O_p\left(\frac{1}{nh_s} + \frac{1}{nh_t}\right).
\end{aligned}
$$

Similarly,

$$
V = O_p\left(\frac{1}{nh_s} + \frac{1}{nh_t}\right),
$$

since V has the same structure as III.

Therefore,

$$\sup_{(u_s,u_t)\in D_{st}} |\widehat{\sigma}_{nst} - \sigma_{nst}| = O_p\left(\frac{1}{\sqrt{n}h_s} + \frac{1}{\sqrt{n}h_t}\right).$$

Hence,

$$\begin{aligned}\sup_{(u_s,u_t)\in D_{st}} |\widehat{\sigma}_{nst} - \sigma_{st}| &\leq \sup_{(u_s,u_t)\in D_{st}} |\widehat{\sigma}_{nst} - \sigma_{nst}| + \sup_{(u_s,u_t)\in D_{st}} |\sigma_{nst} - \sigma_{st}| \\ &= O_p\left(\left[\frac{\log n}{n^2 h_s h_t}\right]^{1/2} + h_s^2 + h_t^2 + \frac{1}{\sqrt{n}h_s} + \frac{1}{\sqrt{n}h_t}\right). \qquad \square\end{aligned}$$

Proof of Theorem 5.3.

Let $\mathcal{L}(\boldsymbol{\mu};\mathbf{Y})$ and $U(\boldsymbol{\beta})$ be the log quasi-likelihood function and quasi-score function with known covariance matrix defined as in Chapter 3. Also, suppose that $\mathbf{I}_\beta$ is the "observed" quasi-information matrix of $\boldsymbol{\beta}$ with known covariance matrix $\mathbf{V}(\cdot)$ and $\mathbf{i}_\beta$ is the expected value of $\mathbf{I}_\beta$. Then we have

$$U(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\mu};\mathbf{Y})}{\partial \boldsymbol{\beta}} = \sum_{k=1}^{n} V^{-1}(\boldsymbol{\mu}_k)(\mathbf{Y}_k - \boldsymbol{\mu}_k)(g'(\bar{\eta}_k))^T \mathbf{X}, \tag{7.5}$$

and

$$\begin{aligned}\mathbf{I}_\beta &= -\left\{\frac{\partial^2 \mathcal{L}(\boldsymbol{\mu};\mathbf{Y})}{\partial \boldsymbol{\beta}^2}\right\} = -\left\{\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right\} \\ &= -\left\{\frac{\partial U}{\partial \eta}\frac{\partial \eta}{\partial \boldsymbol{\beta}}\right\} \\ &= \sum_{k=1}^{n} \mathbf{X}_k \mathbf{V}^{-1}(\boldsymbol{\mu}_k)(g'(\boldsymbol{\eta}_k))^2 \mathbf{X}_k^T + \sum_{k=1}^{n} \mathbf{X}_k \Big\{\left(\mathbf{V}^{-1}(\boldsymbol{\mu}_k)\right)^2 (\mathbf{V}(\boldsymbol{\mu}_k))' (g'(\boldsymbol{\eta}_k))^2 \\ &\quad - g''(\boldsymbol{\eta}_k)\mathbf{V}^{-1}(\boldsymbol{\mu}_k)\Big\}(\mathbf{Y}_k - \boldsymbol{\mu}_k)\mathbf{X}_k^T \\ &= \sum_{k=1}^{n} \mathbf{X}_k \Big\{\left(\mathbf{V}^{-1}(\boldsymbol{\mu}_k)(g'(\boldsymbol{\eta}_k))\right)^2 + \Big[\left(\mathbf{V}^{-1}(\boldsymbol{\mu}_k)\right)^2 (\mathbf{V}(\boldsymbol{\mu}_k))' (g'(\boldsymbol{\eta}_k))^2 \\ &\quad - g''(\boldsymbol{\eta}_k)\mathbf{V}^{-1}(\boldsymbol{\mu}_k)\Big](\mathbf{Y}_k - \boldsymbol{\mu}_k)\Big\}\mathbf{X}_k^T.\end{aligned}$$

Therefore,

$$\mathbf{i}_\beta = E(\mathbf{I}_\beta) = \sum_{k=1}^{n} \mathbf{D}_k^T \mathbf{V}^{-1}(\boldsymbol{\mu}_k)\mathbf{D}_k.$$

Suppose that $\widehat{\boldsymbol{\beta}}$ is a quasi-likelihood estimator of $\boldsymbol{\beta}$. Then $\widehat{\boldsymbol{\beta}}$ satisfies

$$U(\widehat{\boldsymbol{\beta}}) = 0.$$

Here $U(\boldsymbol{\beta})$ is quasi-score equation with known covariance matrix $\mathbf{V}(\cdot)$, which defined in Chapter 3.

By Taylor expansion,

$$U(\boldsymbol{\beta}) - \mathbf{I}_{\bar{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = 0$$

where $\bar{\boldsymbol{\beta}}$ is between $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$. Also, $\mathbf{I}_{\bar{\boldsymbol{\beta}}} = \mathbf{I}_{\beta}\Big|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}$, $\mathbf{I}_{\beta}$ is the "observed" quasi-information matrix of $\boldsymbol{\beta}$ with known covariance matrix $\mathbf{V}(\cdot)$. Now $\mathbf{I}_{\bar{\boldsymbol{\beta}}}$ satisfies that

$$\begin{aligned}\mathbf{I}_{\bar{\boldsymbol{\beta}}} &= \sum_{k=1}^{n} \mathbf{X}_k \Big\{ \left(\mathbf{V}^{-1}\left(\bar{\boldsymbol{\mu}}_k\right)\left(g'(\bar{\boldsymbol{\eta}}_k)\right)\right)^2 \\ &\quad + \Big[\left(\mathbf{V}^{-1}\left(\bar{\boldsymbol{\mu}}_k\right)\right)^2 \left(\mathbf{V}\left(\bar{\boldsymbol{\mu}}_k\right)\right)' \left(g'(\bar{\boldsymbol{\eta}}_k)\right)^2 - g''(\bar{\boldsymbol{\eta}}_k)\mathbf{V}^{-1}\left(\bar{\boldsymbol{\mu}}_k\right)\Big] \left(\mathbf{Y}_k - \bar{\boldsymbol{\mu}}_k\right)\Big\} \mathbf{X}_k^T,\end{aligned}$$

where

$$\bar{\boldsymbol{\mu}}_k = g\left(\mathbf{X}_k\bar{\boldsymbol{\beta}}\right), \qquad \bar{\boldsymbol{\eta}} = \mathbf{X}_k\bar{\boldsymbol{\beta}}.$$

By the results of McCullagh (1983), we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N\left(\mathbf{0}, \Sigma^{-1}\right).$$

Let $\widehat{\boldsymbol{\beta}}^*$ be the nonparametric estimator of $\boldsymbol{\beta}$ such that

$$U^*(\widehat{\boldsymbol{\beta}}^*) = 0$$

and $\mathbf{I}^*_{\beta}$ is the "observed" quasi-information matrix of $\boldsymbol{\beta}$ with nonparametric estimator of covariance matrix $\mathbf{V}_n(\cdot)$, where $U^*(\boldsymbol{\beta})$ is defined as the nonparametric quasi-score equation in chapter 4.

$$\mathbf{U}^*(\boldsymbol{\beta}) = \sum_{k=1}^{n} \mathbf{V}_n^{-1}(\boldsymbol{\mu}_k)(\mathbf{Y}_k - \boldsymbol{\mu}_k)(g'(\bar{\boldsymbol{\eta}}_k))^T\mathbf{X} \tag{7.6}$$

and

$$
\begin{aligned}
\mathbf{I}_{\beta}^{*} &= -\left\{\frac{\partial^{2}\mathcal{L}^{*}(\boldsymbol{\mu};\mathbf{Y})}{\partial\boldsymbol{\beta}^{2}}\right\} = -\left\{\frac{\partial U^{*}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right\} \\
&= \sum_{k=1}^{n}\mathbf{X}_{k}\left\{\left(\widehat{\mathbf{V}}_{n}^{-1}(\boldsymbol{\mu}_{k})(g'(\boldsymbol{\eta}_{k}))\right)^{2} + \left[\left(\widehat{\mathbf{V}}_{n}^{-1}(\boldsymbol{\mu}_{k})\right)^{2}\left(\widehat{\mathbf{V}}_{n}(\boldsymbol{\mu}_{k})\right)'(g'(\boldsymbol{\eta}_{k}))^{2}\right.\right. \\
&\quad \left.\left. -\ g''(\boldsymbol{\eta}_{k})\widehat{\mathbf{V}}_{n}^{-1}(\boldsymbol{\mu}_{k})\right](\mathbf{Y}_{k}-\boldsymbol{\mu}_{k})\right\}\mathbf{X}_{k}^{T}.
\end{aligned}
$$

By Taylor expansion,

$$
U^{*}(\widehat{\boldsymbol{\beta}}^{*}) - \mathbf{I}_{\bar{\boldsymbol{\beta}}}^{*}(\widehat{\boldsymbol{\beta}}^{*}-\boldsymbol{\beta}) = 0
$$

where $\bar{\boldsymbol{\beta}}$ is between $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$. Also, $\mathbf{I}_{\bar{\boldsymbol{\beta}}}^{*} = \mathbf{I}_{\beta}^{*}\Big|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}$, $\mathbf{I}_{\beta}^{*}$ is the "observed" nonparametric quasi-information matrix of $\boldsymbol{\beta}$ with nonparametric estimator of unknown covariance matrix $\mathbf{V}(\cdot)$, and $\mathbf{I}_{\bar{\beta}}^{*}$ satisfies

$$
\begin{aligned}
\mathbf{I}_{\bar{\boldsymbol{\beta}}}^{*} &= \sum_{k=1}^{n}\mathbf{X}_{k}\left\{\left(\widehat{\mathbf{V}}_{n}^{-1}(\bar{\boldsymbol{\mu}}_{k})(g'(\bar{\boldsymbol{\eta}}_{k}))\right)^{2}\right. \\
&\quad \left. + \left[\left(\widehat{\mathbf{V}}_{n}^{-1}(\bar{\boldsymbol{\mu}}_{k})\right)^{2}\left(\widehat{\mathbf{V}}_{n}(\bar{\boldsymbol{\mu}}_{k})\right)'(g'(\bar{\boldsymbol{\eta}}_{k}))^{2} - g''(\bar{\boldsymbol{\eta}}_{k})\widehat{\mathbf{V}}_{n}^{-1}(\bar{\boldsymbol{\mu}}_{k})\right](\mathbf{Y}_{k}-\bar{\boldsymbol{\mu}}_{k})\right\}\mathbf{X}_{k}^{T}
\end{aligned}
$$

where

$$
\bar{\boldsymbol{\mu}}_{k} = g\left(\mathbf{X}_{k}\bar{\boldsymbol{\beta}}\right), \qquad \bar{\boldsymbol{\eta}} = \mathbf{X}_{k}\bar{\boldsymbol{\beta}}.
$$

In order to show the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^{*}$ has the same asymptotic distribution as the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$ obtained from the quasi-score equation with known covariance matrix, it will be sufficient to show that

$$
\sqrt{n}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}^{*}\right) = \sqrt{n}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right) + o_{p}(1) \tag{7.7}
$$

Equation (7.7) will follow if we can show

$$
\frac{1}{n^{2}}\left(\mathbf{I}_{\bar{\boldsymbol{\beta}}}^{*} - \mathbf{I}_{\bar{\boldsymbol{\beta}}}\right) = o_{p}(1)
$$

and

$$\frac{1}{n}\left(\mathbf{U}^*_{\boldsymbol{\beta}} - \mathbf{U}_{\boldsymbol{\beta}}\right) = o_p(1).$$

First, we have

$$\begin{aligned}
&\frac{1}{n^2}\left(\mathbf{I}^*_{\bar{\boldsymbol{\beta}}} - \mathbf{I}_{\bar{\boldsymbol{\beta}}}\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} \mathbf{X}_k \mathcal{H}_k \left(g'(\boldsymbol{\eta}_k)\right)^2 \mathbf{X}_k^T \\
&\quad + \frac{1}{n^2}\sum_{k=1}^{n} \mathbf{X}_k \left\{\mathcal{F}_k \left(g'(\boldsymbol{\eta}_k)\right)^2 - g''(\boldsymbol{\eta}_k)\mathcal{H}_k(\bar{\boldsymbol{\mu}}_k)\right\} \left(\mathbf{Y}_k - \boldsymbol{\mu}_k\right) \mathbf{X}_k^T
\end{aligned}$$

where

$$\mathcal{H}_k(\bar{\boldsymbol{\mu}}_k) = \widehat{\mathbf{V}}_n(\bar{\boldsymbol{\mu}}_k)^{-1} - \mathbf{V}^{-1}(\bar{\boldsymbol{\mu}}_k)$$

and

$$\mathcal{F}_k(\bar{\boldsymbol{\mu}}_k) = \left(\widehat{\mathbf{V}}_n^{-1}\left(\bar{\boldsymbol{\mu}}_k\right)\right)^2 \left(\widehat{\mathbf{V}}_n\left(\bar{\boldsymbol{\mu}}_k\right)\right)' - \left(\mathbf{V}^{-1}\left(\bar{\boldsymbol{\mu}}_k\right)\right)^2 \left(\mathbf{V}\left(\bar{\boldsymbol{\mu}}_k\right)\right)'.$$

Throughout the rest of Chapter, $\|\cdot\|$ means $\|\cdot\|_\infty$. By Theorem 5.2, $(K5)$ and $(N2)$, we have

$$\begin{aligned}
\|\mathcal{H}_k\| &\leq \left\|\widehat{\mathbf{V}}_n(\bar{\boldsymbol{\mu}}_k)^{-1}\right\| \left\|\mathbf{V}(\bar{\boldsymbol{\mu}}_k) - \widehat{\mathbf{V}}_n(\bar{\boldsymbol{\mu}}_k)\right\| \left\|\mathbf{V}(\bar{\boldsymbol{\mu}}_k)^{-1}\right\| \\
&\leq \operatorname{cond}\left(\widehat{\mathbf{V}}_n(\bar{\boldsymbol{\mu}}_k)\right) \left(\left\|\widehat{\mathbf{V}}_n(\bar{\boldsymbol{\mu}}_k)\right\|\right)^{-1} O_p\left(\nu_n\right) \\
&= O_p\left(\nu_n/\lambda_n\right)
\end{aligned}$$

where

$$\nu_n = \left[\frac{\log n}{n^2 h_s h_t}\right]^{1/2} + h_s^2 + h_t^2 + \frac{1}{\sqrt{n}h_s} + \frac{1}{\sqrt{n}h_t}.$$

By the definition of derivative, we have the following:

$$\begin{aligned}
\|\mathcal{F}_k(\bar{\boldsymbol{\mu}}_k)\| &= \left\|\left(1/(\widehat{\mathbf{V}}_n(\bar{\boldsymbol{\mu}}_k)\right)' - (1/(\mathbf{V}(\bar{\boldsymbol{\mu}}_k))'\right\| \\
&= O_p\left(\nu_n/\left(h_s h_t \lambda_n\right)\right).
\end{aligned}$$

By (K5), as $n \to \infty$ we have

$$\begin{aligned}
&\left\| \frac{1}{n^2}\left(\mathbf{I}^*_{\bar{\boldsymbol{\beta}}} - \mathbf{I}_{\bar{\boldsymbol{\beta}}}\right) \right\| \\
&\quad \leq \frac{1}{n^2}\sum_{k=1}^{n} \|\mathcal{H}_k\| + \frac{1}{n^2}\sum_{k=1}^{n} \{[\|\mathcal{F}_k\| + \|\mathcal{H}_k\|]\, \|\mathbf{Y}_k - \boldsymbol{\mu}_k\|\} \\
&\quad = O_p\left((h_s h_t \nu_n)/\lambda_n\right) + O_p\left(\nu_n/\lambda_n\right) \to 0.
\end{aligned}$$

Hence

$$\frac{1}{n^2}\left(\mathbf{I}^*_{\bar{\boldsymbol{\beta}}} - \mathbf{I}_{\bar{\boldsymbol{\beta}}}\right) = o_p(1). \tag{7.8}$$

From (7.5) and (7.6) and (K5), we have the following as $n \to \infty$:

$$\begin{aligned}
\left\| \frac{1}{n}\left(\mathbf{U}^*_{\boldsymbol{\beta}} - \mathbf{U}_{\boldsymbol{\beta}}\right) \right\| &\leq \frac{1}{n}\left\| \sum_{k=1}^{n} \left[\mathbf{V}_n^{-1}(\boldsymbol{\mu}_k) - \mathbf{V}^{-1}(\boldsymbol{\mu}_k)\right] (\mathbf{Y}_k - \boldsymbol{\mu}_k) \left(g'(\bar{\eta}_k)\right)^T \mathbf{X} \right\| \\
&\leq \frac{1}{n}\sum_{k=1}^{n} \left[O_p\left(\left\|\mathbf{V}_n{}^{-1}(\boldsymbol{\mu}_k) - \mathbf{V}^{-1}(\boldsymbol{\mu}_k)\right\|\right) \|\mathbf{Y}_k - \boldsymbol{\mu}_k\|\right] \\
&= O_p\left(\nu_n/\lambda_n\right) \to 0.
\end{aligned}$$

Hence,

$$\frac{1}{n}\left(\mathbf{U}^*_{\boldsymbol{\beta}} - \mathbf{U}_{\boldsymbol{\beta}}\right) = o_p(1). \tag{7.9}$$

Next, we are going to show that if (7.8) and (7.9) are true, then the nonparametric quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}^*$ will have the same asymptotic distribution as the quasi-likelihood estimator $\widehat{\boldsymbol{\beta}}$ with known covariance matrix.

Since

$$\mathbf{U}^*_{\boldsymbol{\beta}} - \mathbf{I}^*_{\bar{\boldsymbol{\beta}}}\left(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right) = \mathbf{U}_{\boldsymbol{\beta}} - \mathbf{I}_{\bar{\boldsymbol{\beta}}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right),$$

then by (7.8), (7.9) and the results in Chapter 3, we have

$$\sqrt{n}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^*\right) = \sqrt{n}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + o_p(1).$$

Therefore, $\sqrt{n}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^*\right)$ will asymptotically have the same distribution as $\sqrt{n}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)$.

Chapter 8

Conclusions and Future Research

8.1 Summary and Conclusions

This proposed nonparametric quasi-likelihood approach is designed to estimate regression model parameters $\boldsymbol{\beta}$ in the class of generalized linear models for longitudinal data analysis where the covariance matrix is totally unknown but consists of smooth functions of means. It has been shown that the estimator obtained from nonparametric quasi-likelihood approach has an asymptotic normal distribution, the same as the estimator obtained from quasi-likelihood approach with true covariance matrix. Moreover, the rate of convergence has been established. With sample size $n = 200$, 300 and 800 separately, four simulation studies were run by using the automatic bandwidth sector (6.1). The results of the first two simulation studies show that with sample size $n = 200$, the relative efficiency of NQLE is close to 1 compared to GEE with correct working correlation structure. In some cases, NQLE is more efficient than GEE, compared to QL with true covariance matrix, such as the estimates of slope in the second simulation. The simulation suggests the following:

NQLE becomes more efficient as n is increasing.

When substantial correlation is present, NQLE seems more efficient than GEE with independent working correlation.

When the working correlation structure is badly misspecified, GEE is very inefficient, compared to NQLE.

NQLE performs as well as or better than GEE with completely unstructured working correlation.

We conclude that NQLE may be superior to GEE when we have no idea of how to choose a working correlation matrix. It is sometimes suggested that GEE with independent working correlation matrix may be used when we have no idea how to choose the working correlation. However, our simulations suggest that this strategy may not be effective and that NQLE is a better approach.

The simulation suggests that the efficiency of NQLE increase as sample size n increases. This agrees with our theoretical result. Presumably the explanation is that the nonparametric estimator of the variance function is more accurate when n is large.

8.2 Suggestions for Future Research

The approach of this thesis assumes a correctly specified link function and a correctly specified linear predictors, while the covariance structure is unspecified, except for smoothness conditions. However, in some cases, the true link function may not be specified correctly or that assumption of a linear predictor may be inaccurate.

A semiparametric or fully nonparametric specification of $E(Y|X)$ might address these problems. A fully nonparametric approach might call for estimating

$\mu(x) = E(Y|X)$ by a smooth function of unspecified form. This approach has already been studied for independent data. Semiparametric approaches include generalized additive models (GAMs) in which $E(Y|X) = g\left(\sum_{j=1}^{p} f_j(x_j)\right)$ and the $f_j(\cdot)$ are smooth function estimated nonparametrically from the data (Hastie and Tibshirani (1992)). Hybrid approaches might also be used, for example by assuming $E(Y|X) = \mathbf{X}_1\boldsymbol{\beta} + g\left(\sum_{j=q+1}^{p} f_j(x_j)\right)$.

A combination of a semiparametric or nonparametric model for $E(Y|X)$ and nonparametric estimation of the covariance functions is a nature extension of the results of this thesis. The principal difficulties will be establishing the various rates of convergence.

It is unlikely that the efficiency of such methods will be as high as QL with correctly specified covariance structures, but they may be appropriate for some studies where useful modeling information is unavailable.

BIBLIOGRAPHY

[1] CARROLL, R. J. and RUPPERT, D. (1982). "Robust Estimation in Heteroscedastic Linear Models." *Annals of Statistics*, **10**, 429-441.

[2] CHIOU, J.-M. and MÜLLER, H.-G. (1999). "Nonparametric Quasi-likelihood." *Annals of Statistics*, **27**, 36-64.

[3] CLEVELAND S. W. (1979). "Robust Local Weighted Regression and Smoothing Scatterplots: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* , **74**, 829-836.

[4] CLEVELAND S. W. and DEVLIN J. S. (1988). "Local Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* , **83**, 597-610.

[5] CROWDER, M. (1995) "On the Use of a Working Matrix in Using Generalised Linear Models for Repeated Measures." *Biometrika* , **82**, 407-410.

[6] DIGGLE, P. J.,HEAGERTY, P., LIANG, K. Y., and ZEGER, S. L. (2002). *Analysis of Longitudinal Data (2nd Edition)*, Oxford, U.K: Oxford, University Press.

[7] FAN, J and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.

[8] FAN, J and LI, R. (2004). "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis." *Journal of the American Statistical Association* , **99**, 710-723.

[9] HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[10] LIANG, K. Y. and ZEGER S. L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* , **73**, 13-22.

[11] McCULLAGH, P. (1983). "Quasi-likelihood Functions" *Annals of Statistics,* **11**, 59-67

[12] MÜLLER, H. -G. and STADTMÜLLER, U. (1987). "Estimation of Heteroscedasticity in Regression Analysis." *Annals of Statistics,* **15**, 610-625.

[13] MÜLLER, H.-G. and ZHAO, P. L. (1995). "On a Semiparametric Variance Function Model and a Test for Heteroscedasticity." *Annals of Statistics,* **23**, 946-967.

[14] McCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models (2nd Edition)*, London: Chapman and Hall

[15] NELDER, J. A. and PREGIBON, P. (1987). "An Extended Quasi-likelihood Function." *Biometrika,* **74**, 221-232.

[16] NELDER, J. A. and WEDDERBURN, R. W. M. (1972). "Generalized Linear Models." *J. Royal Statistics Society, Ser. A,* **135**, 370-384.

[17] RUPPERT, D. and WANG, M. P. (1997). "Local Polynomial Variance-Function Estimation." *Technometrics,* **39**, 262-273.

[18] STONE, C. J. (1977). "Consistent Nonparametric Regression" *Annals of Statistics,* **5**, 595-645.

[19] STONE, C. J. (1980). "Optimal Rates of Convergence for Nonparametric Estimators" *Annals of Statistics,* **8**, 1348-1360.

[20] STONE, C. J. (1982). "Optimal Global Rates of Convergence for Nonparametric Regression" *Annals of Statistics,* **10**, 1040-1053. " *Annals of Statistics,* **5**, 595-645

[21] SUTRADHAR, B.C. (1999). "On the Efficiency of Regression Estimators in Generalised Linear Models for Longitudinal Data." *Biometrika,* **86**, 459-465.

[22] WEDDERBURN, R. W. M. (1974). " Quasi-likelihood Functions, Generalized Linear Models and Gauss-Newton Method." *Biometrika,* **61,** 439-447.

[23] WANG, Y. G.and CAREY, J.V. (2004). "Unbiased Estimating Equations From Working Correlation Models For Irregularly Timed Repeated Measures." *Journal of the American Statistical Association,* **99**, 845-853.

[24] ZEGER S. L. and LIANG, K.Y. (1986). "Longitudinal Data Analysis for Discrete and Continuous Outcomes" *Biometrics,* **42**, 121-130.