



uOttawa

L'Université canadienne  
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES  
ET POSTDOCTORALES



FACULTY OF GRADUATE AND  
POSTDOCTORAL STUDIES

Zuojing Li

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Mathematics)

GRADE / DEGREE

The Department of Mathematics and Statistics

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Longitudinal Data Analysis Using Generalized Linear Models

TITRE DE LA THÈSE / TITLE OF THESIS

Dr. R. Balan

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Dr. A. Dabrowski

Dr. C.G. Park

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

LONGITUDINAL DATA ANALYSIS  
USING  
GENERALIZED LINEAR MODELS

By  
Zuojing Li, B.Sc.  
June 2006

A Thesis  
submitted to the Faculty of Graduate Studies and Research  
in partial fulfillment of the requirements  
for the degree of  
Master of Science in Mathematics<sup>1</sup>

© Copyright 2006  
by Zuojing Li, B.Sc., Ottawa, Canada

---

<sup>1</sup>The M.Sc. Program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-18440-0*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-18440-0*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

# Abstract

In this work we examine various conditions under which the usual asymptotic results (i.e. the weak consistency, the asymptotic normality and the strong consistency) hold for the regressor parameter  $\beta$  which arises in a linear model (Chapter 2), a generalized linear model (GLM) with a fully specified likelihood (Chapter 3) or as a root of the generalized estimating equation (GEE) associated with a sequence of longitudinal observations (Chapter 4). Our main references for each of these chapters are [12], [9], respectively [20].

We provide detailed proofs of the results found in the above-mentioned references, and we extend the results of [9] to the case of stochastic regressors (Section 3.4). Finally, in Chapter 5, we identify a fundamental mistake appearing in the recent article [4], which examines the strong consistency of the regressor parameter  $\beta$  in a GLM for which the likelihood of the density is not specified. In Section 5.2, we give a correction to the main theorem of [4], as well as some new results concerning the weak consistency and asymptotic normality of  $\beta$ .

# Acknowledgements

I would like to thank my supervisor Dr. Raluca Balan for all her support, encouragement and helpful explanations, as well as for guiding me through the difficult task of writing my Master's thesis.

# Dedication

I dedicate this work to my parents and my husband.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Dedication</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Linear Regression</b>	<b>4</b>
2.1 The Framework . . . . .	5
2.2 The Strong Consistency . . . . .	7
2.3 Necessity of Condition ( $\mathbf{D}^*$ ) . . . . .	12
2.4 An Alternative Result of Kaufmann . . . . .	13
<b>3 The Generalized Linear Model</b>	<b>16</b>
3.1 The Framework . . . . .	16
3.2 The General Results . . . . .	19
3.2.1 Asymptotic Existence and Weak Consistency . . . . .	19
3.2.2 Asymptotic Normality . . . . .	22
3.2.3 Asymptotic Existence and Strong Consistency . . . . .	26
3.3 Particular Cases and Examples . . . . .	28
3.3.1 Bounded Regressors . . . . .	28
3.3.2 Linear Regression . . . . .	30
3.3.3 Poisson Log-linear Regression . . . . .	30

3.3.4	Logistic Regression for Binary Data . . . . .	32
3.4	Stochastic Regressors . . . . .	33
3.4.1	Asymptotic Existence and Weak Consistency . . . . .	36
3.4.2	Asymptotic Normality . . . . .	38
3.4.3	Asymptotic Existence and Strong Consistency . . . . .	41
<b>4</b>	<b>Longitudinal Data</b>	<b>43</b>
4.1	The Framework . . . . .	44
4.2	The General Results . . . . .	50
4.2.1	Asymptotic Existence and Weak Consistency . . . . .	50
4.2.2	Asymptotic Normality . . . . .	56
4.2.3	Asymptotic Existence and Strong Consistency . . . . .	61
4.3	Verification of Condition (CC) . . . . .	64
4.4	Particular Cases and Examples . . . . .	65
4.4.1	Bounded Regressors . . . . .	65
4.4.2	Bounded responses . . . . .	69
4.4.3	Logistic Regression for Binary Data . . . . .	71
4.4.4	Linear Regression . . . . .	72
4.4.5	Poisson Log-linear Regression . . . . .	74
<b>5</b>	<b>The Quasi-likelihood Approach</b>	<b>76</b>
5.1	The Argument of Chen, Hu and Ying . . . . .	77
5.2	The Asymptotic Results . . . . .	80
5.2.1	The Weak Consistency . . . . .	80
5.2.2	The Asymptotic Normality . . . . .	80
5.2.3	The Strong Consistency . . . . .	82
<b>A</b>	<b>Background Material</b>	<b>85</b>
A.1	Matrix Analysis Results . . . . .	85
A.2	Limit Theorems . . . . .	88
A.3	Analysis Results . . . . .	91

<b>B</b>	<b>Analysis of Longitudinal Data Sets Using the SAS System</b>	<b>92</b>
B.1	An Example Using Logistic Regression for Binary Data . . . .	93
B.2	An Example Using Log-linear Poisson Regression For Count Data . .	101

# Chapter 1

## Introduction

Longitudinal data analysis plays an important role in biostatistics. The observations are measured repeatedly **through time**. For example, we are interested in 10 patients' blood pressure on 8 successive days. The repeated measurement of a response variable  $y_{ij}$  is related to a set of covariates  $x_{ij}$ . The correlation within each subject should be taken into account. As in [14], Liang and Zeger proposed that longitudinal data analysis be an extension of a generalized linear model (GLM). In this paper, they proposed to use the generalized estimating equation method to get the estimator  $\hat{\beta}$  of  $\beta$ . GLM is an extension of classical linear models [15]. In GLM, the relationship between the response  $y_i$  and the linear combination of covariates  $x_i$  is specified (in the classical linear model, the mean of the response is a combination of covariates). In the GLM case, we suppose each variable  $y_i$  belongs to an exponential family driven by a parameter  $\theta_i$  and  $\theta_i = x_i^T \beta$ . In the case of longitudinal data, we assume that response  $y_{ij}$  satisfies the following assumptions (see [7]):

- (1) the marginal expectation of the response  $\mu_{ij}$  depends on the covariates (i.e.  $\mu_{ij}$  is a function of the explanatory variable);
- (2) the marginal variance of the response depends on the marginal mean;
- (3) the correlation between  $y_{ij}$  and  $y_{ik}$  is a function of the marginal means and perhaps of additional parameter  $\alpha$ .

In this work, we focus on the question of existence of a consistent estimator for the regression parameter  $\beta$  and its asymptotic properties in a classical linear model

(Chapter 2), a generalized linear model (Chapter 3 and 5) and in the case of a set of longitudinal data (Chapter 4).

In Chapter 2, we study the classical linear model. We use the least squares method to get the estimator  $\hat{\beta}_n$  by minimizing the sum of the squared distance from the points  $(x_i, y_i)$ ,  $i = 1, \dots, n$  to the hyperplane  $y = x^T \beta$ . We get  $\hat{\beta}_n = (\sum_{i=1}^n x_i x_i^T)^{-1} (\sum_{i=1}^n x_i y_i)$ . We examine the question of strong consistency of  $\hat{\beta}_n$  and present the theory of Lai, Robbins and Wei mainly (our reference [12]) by examining the condition  $H_n^{-1} = (\sum_{i=1}^n x_i x_i^T)^{-1} \rightarrow 0$ .

In Chapter 3, we investigate the main properties of a generalized linear model. For this, we assume that each response  $y_i$  has a density which belongs to an exponential family with parameter  $\theta_i = x_i^T \beta$ . The maximum likelihood method is used to get the estimator  $\hat{\beta}_n$ . The estimator  $\hat{\beta}_n$  is the point where the log-likelihood function  $l_n(\beta)$  attains its maximum. We present the asymptotic properties of  $\hat{\beta}_n$ , under various sets of conditions involving  $F_n(\beta) = \sum_{i=1}^n x_i \sigma_i^2(\beta) x_i^T$ . (The main reference we use is in [9]). The main tools used are the central limit theorem for asymptotic normality and the strong law large numbers for the strong consistency. For weak consistency, **no** independence is required. If the density function of  $y_i$  is not specified, we use the quasi-likelihood method to get the estimator  $\hat{\beta}_n$ . The question of the strong consistency of  $\hat{\beta}_n$  is discussed in [4], but we believe that the main result of [4] is not true. Therefore, in Chapter 5, we give a correction to the theory presented in [4], by applying a method similar to the one developed in Chapter 4.

In Chapter 4, we give a complete treatment of the GLM approach, in the case of longitudinal data. The generalized estimating equation method is used to get the estimator  $\hat{\beta}_n$ . The “working” correlation matrix  $R_i(\alpha)$ , whose form may depend on an unknown parameter  $\alpha$ , is used to replace the unknown true correlation matrix  $\bar{R}_i$ . We present the asymptotic properties of  $\hat{\beta}_n$  under various set of conditions which do not depend on the unknown correlation matrix.

This work contains two appendices. Appendix A contains some background material regarding to matrix analysis (Section A.1), limit theorems (Section A.2) and real analysis (Section A.3). In appendix B, there are two examples: the first one

illustrates the use of the logistic regression for binary data (Section B.1), whereas the second one is an application of the log-linear Poisson regression for count data (Section B.2).

We give highly elaborated proofs of asymptotic consistency and normality for estimators arising from the above mentioned references. Two new results are presented in our work. One is that in a GLM we demonstrate the asymptotic properties of the estimator for the regressor parameters when considering the stochastic regressors. The other one is that we correct the strong consistency theorem of Chen, Hu and Ying (their proof is not valid) when the density function of the response variable is not specified.

# Chapter 2

## Linear Regression

In this chapter, we examine the question of strong consistency of the least square estimator (LSE) in a classical linear regression model (with  $p$ -dimensional regressors,  $p \geq 1$ ):  $y_i = x_i^T \beta + \varepsilon_i$ . The major reference that we are using is the fundamental work of Lai, Robbins and Wei, which was published back in 1979 (our reference [12]). Their result is of tremendous importance since it represents the first instance in the literature when the exact rate of convergence of the LSE to the true parameter  $\beta_0$  is specified in terms of the design matrix  $H_n = \sum_{i=1}^n x_i x_i^T$ , under very weak dependence conditions on the regressors. The main condition for the strong consistency of the LSE is the “divergence” of  $H_n$ , in the sense that the diagonal elements of  $H_n^{-1}$  have to converge to 0.

This chapter is organized as follows. In Section 2.1 we introduce the main notation and conditions, and we develop the formula for the LSE. In Section 2.2 we provide the details of the proof of the main theorem in [12], as well as some corollaries. The proof uses heavily the algebraic form of the LSE. We omit many of the algebraic manipulations, as well as the proof of a deeper result (stated here as Theorem 2.2.2). In Section 2.3 we discuss the necessity of the divergence condition for the strong consistency of the LSE. Finally, an alternative result of Kaufmann is presented in Section 2.4.

## 2.1 The Framework

Let  $(y_i)_{i \geq 1}$  be a sequence of response variables, which are recorded together with some explanatory variables  $(x_i)_{i \geq 1}$ . We assume that each  $y_i$  is a random variable defined on a probability space  $(\Omega, \mathcal{F}, P_\beta)$  and each  $x_i$  is a (non-random)  $p$ -dimensional vector such that

$$E_\beta(y_i) = x_i^T \beta, \quad \forall i \geq 1,$$

where  $E_\beta$  denotes the expectation with respect to  $P_\beta$ . Here  $\beta$  denotes an unknown  $p$ -dimensional parameter which has to be estimated, and  $\beta_0$  is the true parameter.

We introduce the **errors**

$$\varepsilon_i(\beta) = y_i - x_i^T \beta, \quad \forall i \geq 1.$$

We will employ the usual convention of omitting  $\beta_0$  in writing, i.e. we will denote  $\varepsilon_i = \varepsilon_i(\beta_0)$ ,  $P = P_{\beta_0}$ , etc. Hence under  $P$ , our model becomes

$$y_i = x_i^T \beta_0 + \varepsilon_i, \quad \text{where } E(\varepsilon_i) = 0, \quad \forall i \geq 1. \quad (1)$$

In matrix notation, this can be written as follows: for every  $n \geq 1$

$$Y_n = X_n \beta + E_n,$$

$$\text{where } Y_n = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X_n = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix}, \quad E_n = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

We denote

$$H_n = X_n^T X_n = \sum_{i=1}^n x_i x_i^T,$$

and we suppose that  $H_n$  is nonsingular for large  $n$ .

We will assume that (under  $P$ ) the sequence  $\{\varepsilon_i\}_{i \geq 1}$  satisfies the following condition:

(C<sub>1</sub>)  $\sum_{i=1}^{\infty} C_i \varepsilon_i$  converges almost surely for any sequence of real numbers  $\{C_i\}_{i \geq 1}$  such that  $\sum_{i=1}^{\infty} C_i^2 < \infty$ .

We can also consider the following stronger condition:

**(C<sub>2</sub>)**  $\{\varepsilon_i\}_{i \geq 1}$  is martingale difference sequence ( i.e.  $E(\varepsilon_i | \varepsilon_1, \dots, \varepsilon_{i-1}) = 0, \forall i \geq 1$ ) and  $\sup_i E\varepsilon_i^2 < \infty$ .

**Remark:** In particular **(C<sub>2</sub>)** holds if  $\{\varepsilon_i\}_{i \geq 1}$  is a sequence of independent random variables with  $E(\varepsilon_i) = 0$  and  $\sup_i E\varepsilon_i^2 < \infty$ .

The next result shows that condition **(C<sub>1</sub>)** imposes a weaker dependence assumption on the errors  $(\varepsilon_i)_{i \geq 1}$  than condition **(C<sub>2</sub>)**

**Lemma 2.1.1** **(C<sub>2</sub>)** implies **(C<sub>1</sub>)**.

**Proof:** Let  $S_n = \sum_{i=1}^n C_i \varepsilon_i; n \geq 1$ . Note that under **(C<sub>2</sub>)**,  $\{S_n\}_{n \geq 1}$  is a zero-mean square-integrable martingale and

$$\begin{aligned} E(S_n^2) &= E\left(\sum_{i=1}^n C_i \varepsilon_i\right)^2 = E\left(\sum_{i=1}^n C_i^2 \varepsilon_i^2 + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} C_i \varepsilon_i C_j \varepsilon_j\right) = \\ &= \sum_{i=1}^n C_i^2 E\varepsilon_i^2 + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} C_i C_j E[E(\varepsilon_i \varepsilon_j | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{i-1})] = \\ &= \sum_{i=1}^n C_i^2 E\varepsilon_i^2 + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} C_i C_j E[\varepsilon_j (E(\varepsilon_i | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{i-1}))] = \sum_{i=1}^n C_i^2 E\varepsilon_i^2 \leq \\ &\leq (\sup_i E\varepsilon_i^2) \sum_{i=1}^n C_i^2 < C \text{ for every } n. \end{aligned}$$

From the martingale convergence theorem (Theorem A.2.9, Appendix A.2), we obtain that  $(S_n)_n$  converges almost surely, as  $n \rightarrow \infty$ .

□

The least square estimator of  $\beta$  is obtained by minimizing the sum of the squared distances from the points  $(x_i, y_i), i = 1, \dots, n$  to the hyperplane  $y = x^T \beta$ . More precisely,  $L_n(\hat{\beta}_n) = \min_{\beta \in R^p} L_n(\beta)$ , where  $L_n(\beta) = \sum_{i=1}^n (y_i - x_i^T \beta)^2$ . From

$$\frac{\partial L_n(\beta)}{\partial \beta_j} = 0, \quad j = 1, \dots, p,$$

we obtain that  $\hat{\beta}_n$  is the solution of the equation  $\sum_{i=1}^n x_i(y_i - x_i^T \beta) = 0$ , i.e.

$$\hat{\beta}_n = \left( \sum_{i=1}^n x_i x_i^T \right)^{-1} \left( \sum_{i=1}^n x_i y_i \right) = (X_n^T X_n)^{-1} X_n^T Y_n.$$

Note that

$$\hat{\beta}_n - \beta_0 = (X_n^T X_n)^{-1} (X_n^T Y_n - X_n^T X_n \beta_0) = H_n^{-1} (X_n^T E_n) = H_n^{-1} \sum_{i=1}^n x_i \varepsilon_i.$$

## 2.2 The Strong Consistency

We begin with the fundamental result of Lai, Robbins and Wei, whose proof we will explain in detail in this section. We should point out that we fill in most of the technical details which are not specifically stated in their paper.

**Theorem 2.2.1 (Theorem 1, p. 345, [12])** *Suppose that  $(\varepsilon_i)_i$  satisfies  $(C_1)$  and  $H_m$  is nonsingular for some  $m$ . Let  $V_n = H_n^{-1} = (v_{ij}^{(n)})_{i,j=1,\dots,p}$ . If  $(v_{jj}^{(n)}) \rightarrow 0$ , for each  $j = 1, \dots, p$ , then for every  $\delta > 0$ ,*

$$\hat{\beta}_{nj} - \beta_{0j} = o(\{v_{jj}^{(n)} |\log v_{jj}^{(n)}|^{1+\delta}\}^{1/2}) \quad a.s. \quad (2)$$

To prove the theorem, we will use the following two auxiliary results whose proofs are omitted. (The interested reader can find the proofs in [12]).

To fix ideas, we will give the proof only in the case  $j = 1$ . The case of an arbitrary  $j \in \{2, \dots, p\}$  is treated similarly.

**Theorem 2.2.2 (Theorem 2, p. 345, [12])** *Let  $\{\varepsilon_i\}_i$  be a sequence of random variables satisfying  $(C_1)$ . Let  $k$  be a positive integer. For each  $n \geq 1$ , let  $T_n$  be a  $k$ -dimensional vector of constants and let  $H_n = \sum_{i=1}^n T_i T_i^T$ . Assume that  $H_m$  is positive definite for some  $m$ .*

*If  $\{C_n\}_n$  is a sequence of constants, such that  $\sum_{i=m+1}^{\infty} C_i^2 (1 + T_i^T H_{i-1}^{-1} T_i) < \infty$ , then*

$$\sum_{i=m+1}^{\infty} C_i T_i^T H_{i-1}^{-1} \left( \sum_{j=1}^{i-1} T_j \varepsilon_j \right) \text{ converges a.s.}$$

**Lemma 2.2.3 (Lemma 3, p. 349, [12])** *Let  $p \geq 2$ . Assume that  $X_n^T X_n$  is positive definite for any  $n \geq m$ , where  $m \geq p$ . Let  $T_n = (x_{n2}, x_{n3}, \dots, x_{np})^T$ . Partition the matrix  $X_n^T X_n$  as*

$$X_n^T X_n = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & K_n \\ K_n^T & H_n \end{bmatrix}$$

(a) *Then for any  $n \geq m$ , we have*

$$\hat{\beta}_{n1} = \beta_{01} + \frac{\sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i) \varepsilon_i}{\sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i)^2} \quad (3)$$

(b) *Define for  $n \geq m$ ,*

$$u_n = \sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i) \varepsilon_i, \quad w_n = u_n - u_{n-1}, \quad (4)$$

$$d_n = x_{n1} - K_n H_n^{-1} T_n. \quad (5)$$

*Then for  $n \geq m$ , we have*

$$\sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i)^2 = \sum_{i=1}^{n-1} (x_{i1} - K_{n-1} H_{n-1}^{-1} T_i)^2 + d_n^2 (1 + T_n^T H_{n-1}^{-1} T_n), \quad (6)$$

$$w_n = d_n [\varepsilon_n - T_n^T H_{n-1}^{-1} (\sum_{i=1}^{n-1} T_i \varepsilon_i)]. \quad (7)$$

**Proof of Theorem 2.2.1:** (pages 355-356 of [12]) We treat first the case  $p \geq 2$ .

Note that  $H_n = H_m + \sum_{i=m+1}^n x_i x_i^T$ . Hence the fact that  $H_m$  is nonsingular implies that  $H_n$  is nonsingular for all  $n \geq m$ . From (3) and (4), we get

$$\hat{\beta}_{n1} - \beta_{01} = \frac{u_n}{s_n} = \frac{u_m + \sum_{j=m+1}^n w_j}{s_n}, \quad (8)$$

where  $s_n = \sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i)^2$ . An elementary matrix calculation shows that: (see (2.24) on page 351 of [12])

$$s_n = \frac{1}{v_{11}^{(n)}} \quad (9)$$

and therefore  $s_n \rightarrow \infty$  and  $|\log s_n| = |\log v_{11}^{(n)}|$ . It follows that

$$\begin{aligned} \frac{\hat{\beta}_{n1} - \beta_{01}}{[v_{11}^{(n)} |\log v_{11}^{(n)}|^{1+\delta}]^{1/2}} &= \frac{u_m}{s_n [v_{11}^{(n)} |\log v_{11}^{(n)}|^{1+\delta}]^{1/2}} + \frac{\sum_{j=m+1}^n w_j}{s_n [v_{11}^{(n)} |\log v_{11}^{(n)}|^{1+\delta}]^{1/2}} \\ &= \frac{u_m}{[s_n |\log s_n|^{1+\delta}]^{1/2}} + \frac{\sum_{j=m+1}^n w_j}{[s_n |\log s_n|^{1+\delta}]^{1/2}}. \end{aligned}$$

For the first term, since  $s_n \rightarrow \infty$  and  $m$  is fixed, we have immediately

$$\frac{u_m}{[s_n |\log s_n|^{1+\delta}]^{1/2}} \rightarrow 0 \text{ a.s.}$$

For the second term, we will use Kronecker's lemma. Hence if we want to show that

$$\frac{\sum_{j=m+1}^n w_j}{[s_n |\log s_n|^{1+\delta}]^{1/2}} \rightarrow 0, \text{ a.s.}$$

it is sufficient to show that

$$\sum_{i=m+1}^{\infty} \frac{w_i}{[s_i |\log s_i|^{1+\delta}]^{1/2}} \text{ converges a.s.} \quad (10)$$

From (7), for  $i > m$ ,

$$w_i = d_i \varepsilon_i - d_i T_i^T H_{i-1}^{-1} \left( \sum_{j=1}^{i-1} T_j \varepsilon_j \right).$$

Hence

$$\sum_{i=m+1}^{\infty} \frac{w_i}{[s_i |\log s_i|^{1+\delta}]^{1/2}} = \sum_{i=m+1}^{\infty} \frac{d_i \varepsilon_i}{[s_i |\log s_i|^{1+\delta}]^{1/2}} - \sum_{i=m+1}^{\infty} \frac{d_i T_i^T H_{i-1}^{-1} \left( \sum_{j=1}^{i-1} T_j \varepsilon_j \right)}{[s_i |\log s_i|^{1+\delta}]^{1/2}}. \quad (11)$$

We treat separately the two terms.

For the first term in (11), we will use condition  $(C_1)$  with  $C_i = d_i / [s_i |\log s_i|^{1+\delta}]^{1/2}$ .

Therefore, in order to prove that

$$\sum_{i=m+1}^{\infty} C_i \varepsilon_i = \sum_{i=m+1}^{\infty} \frac{d_i \varepsilon_i}{[s_i |\log s_i|^{1+\delta}]^{1/2}} \text{ converges a.s.} \quad (12)$$

it suffices to show that

$$\sum_{i=m+1}^{\infty} C_i^2 = \sum_{i=m+1}^{\infty} \frac{d_i^2}{s_i |\log s_i|^{1+\delta}} < \infty. \quad (13)$$

Note that (see (6) and the definition of  $s_n$ ):

$$s_n - s_{n-1} = d_n^2(1 + T_n^T H_{n-1}^{-1} T_n) := a_n, \quad \forall n \geq m.$$

Hence  $(s_n)_{n \geq m}$  is the partial sum sequence of  $(a_i)_{i \geq m}$ , i.e.

$$s_n = s_m + \sum_{i=m+1}^n d_i^2(1 + T_i^T H_{i-1}^{-1} T_i).$$

Clearly

$$d_i^2 \leq d_i^2(1 + T_i^T H_{i-1}^{-1} T_i), \quad \forall i \geq m$$

by the nonnegative-definiteness of  $H_{i-1}^{-1}$ . It follows that

$$\sum_{i=m+1}^{\infty} \frac{d_i^2}{s_i |\log s_i|^{1+\delta}} \leq \sum_{i=m+1}^{\infty} \frac{d_i^2(1 + T_i^T H_{i-1}^{-1} T_i)}{s_i |\log s_i|^{1+\delta}} \leq \int_{s_{m+1}}^{\infty} \frac{1}{x |\log x|^{1+\delta}} dx < \infty, \quad (14)$$

where we assume that  $s_1 > 1$ . Relation (13) is proved and (12) follows.

For the second term in (11), we will apply Theorem 2.2.2 (with the same  $C_i$  as above). Therefore, in order to prove that

$$\sum_{i=m+1}^{\infty} C_i T_i^T H_{i-1}^{-1} \left( \sum_{j=1}^{i-1} T_j \varepsilon_j \right) = \sum_{i=m+1}^{\infty} \frac{d_i T_i^T H_{i-1}^{-1} \left( \sum_{j=1}^{i-1} T_j \varepsilon_j \right)}{[s_i |\log s_i|^{1+\delta}]^{1/2}} \text{ converges a.s.} \quad (15)$$

it suffices to show that

$$\sum_{i=m+1}^{\infty} C_i^2 (1 + T_i^T H_{i-1}^{-1} T_i) = \sum_{i=m+1}^{\infty} \frac{d_i^2 (1 + T_i^T H_{i-1}^{-1} T_i)}{s_i |\log s_i|^{1+\delta}} < \infty.$$

But this is exactly (14). From (11), (12) and (15), we obtain (10). This concludes the proof in the case  $p \geq 2$ .

In the case  $p = 1$ , we have

$$X_n^T X_n = (x_1, \dots, x_n)(x_1, \dots, x_n)^T = \sum_{i=1}^n x_i^2 = s_n, \quad V_n = H_n^{-1} = v_{11}^{(n)} = \frac{1}{s_n},$$

$$X_n^T Y_n = (x_1, \dots, x_n)(y_1, \dots, y_n)^T = \sum_{i=1}^n x_i y_i, \quad \hat{\beta}_n = \left( \sum_{i=1}^n x_i^2 \right)^{-1} \sum_{i=1}^n x_i y_i,$$

$$\hat{\beta}_n - \beta_0 = \frac{\sum_{i=1}^n x_i \varepsilon_i}{s_n}$$

and hence

$$\frac{\hat{\beta}_n - \beta_0}{\{v_{11} |\log v_{11}^{(n)}|^{1+\delta}\}^{1/2}} = \frac{\sum_{i=1}^n x_i \varepsilon_i}{s_n \{v_{11} |\log v_{11}^{(n)}|^{1+\delta}\}^{1/2}} = \frac{\sum_{i=1}^n x_i \varepsilon_i}{\{s_n |\log s_n|^{1+\delta}\}^{1/2}}. \quad (16)$$

We will use the integral test for convergence of sums with  $f(x) = 1/(x|\log x|^{1+\delta})$ .

Since  $s_n = \sum_{i=1}^n x_i^2$ , we obtain

$$\sum_{i=1}^{\infty} \frac{x_i^2}{s_i |\log s_i|^{1+\delta}} \leq \int_{s_1}^{\infty} \frac{1}{x |\log x|^{1+\delta}} dx < \infty,$$

where we assume that  $s_1 > 1$ . By condition  $(\mathbf{C}_1)$  with  $C_i = x_i/[s_i |\log s_i|^{1+\delta}]^{1/2}$ , this implies

$$\sum_{i=1}^{\infty} C_i \varepsilon_i = \sum_{i=1}^{\infty} \frac{x_i \varepsilon_i}{\{s_i |\log s_i|^{1+\delta}\}^{1/2}} \text{ converges a.s.}$$

By Kronecker's Lemma, we get

$$\frac{\sum_{i=1}^n x_i \varepsilon_i}{\{s_n |\log s_n|^{1+\delta}\}^{1/2}} \longrightarrow 0 \text{ a.s.} \quad (17)$$

The desired conclusion follows from (16) and (17).

□

**Corollary 2.2.4** *Suppose that  $(\varepsilon_i)_{i \geq 1}$  satisfies  $(\mathbf{C}_1)$  and the following condition holds*

$$(\mathbf{D}^*) \quad \lambda_{\min}(H_n) \rightarrow \infty.$$

Then

$$\hat{\beta}_n - \beta_0 = o\left(\left[\frac{|\log \|H_n\||^{1+\delta}}{\lambda_{\min}(H_n)}\right]^{1/2}\right) \text{ a.s.}$$

**Remark:** Note that  $(\mathbf{D}^*)$  is equivalent to  $\|H_n^{-1}\| = \lambda_{\max}(H_n^{-1}) \rightarrow 0$ .

**Proof:** We have

$$\frac{1}{\|H_n\|} = \frac{1}{\lambda_{\max}(H_n)} = \lambda_{\min}(V_n) \leq v_{jj}^{(n)} \leq \lambda_{\max}(V_n) = \frac{1}{\lambda_{\min}(H_n)}.$$

Note that  $\lambda_{\min}(H_n) \rightarrow \infty$  implies that  $v_{jj}^{(n)} \rightarrow 0$  for all  $j = 1, \dots, p$ , and hence Theorem 2.2.1 can be applied. For  $n$  large enough, we have

$$-\log \|H_n\| \leq \log v_{jj}^{(n)} \leq -\log \lambda_{\min}(H_n) < 0.$$

It follows that

$$|\log \lambda_{\min}(H_n)| \leq |\log v_{jj}^{(n)}| \leq |\log \|H_n\||$$

and

$$v_{jj}^{(n)} |\log v_{jj}^{(n)}|^{1+\delta} \leq \frac{|\log \|H_n\||^{1+\delta}}{\lambda_{\min}(H_n)}.$$

□

## 2.3 Necessity of Condition (D\*)

The following results show that the divergence condition (D\*) is also necessary for the strong consistency of the least squares estimator in the case  $p = 1$ . This is obtained under condition (C<sub>2</sub>).

**Theorem 2.3.1 (Lemma 4.1, p. 125, [8])** *Let  $p = 1$ . Suppose that  $(\varepsilon_i)_{i \geq 1}$  satisfies (C<sub>2</sub>) and in addition  $\inf_{i \geq 1} E\varepsilon_i^2 > 0$ . Then*

$$\hat{\beta}_n - \beta_0 \rightarrow 0 \quad \text{a.s.} \quad (18)$$

*if and only if (D\*) holds, i.e.*

$$\sum_{i=1}^{\infty} x_i^2 = \infty \quad (19)$$

**Proof:** For (19)  $\implies$  (18) see the proof of Theorem 2.2.1, case  $p = 1$ .

(18)  $\implies$  (19) : Suppose that  $x = \sum_{i=1}^{\infty} x_i^2 \in (0, \infty)$ .

Because (C<sub>2</sub>) implies (C<sub>1</sub>), we know that there exists a random variable  $V$  such that  $\sum_{i=1}^n x_i \varepsilon_i \rightarrow V$  a.s.

Hence

$$\hat{\beta}_n - \beta_0 = \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2} \rightarrow \frac{V}{x} \quad \text{a.s.}$$

Since  $\hat{\beta}_n - \beta_0 \rightarrow 0$  a.s., we must have  $V = 0$  a.s., that is  $S_n := \sum_{i=1}^n x_i \varepsilon_i \rightarrow 0$  a.s.

On the other hand, since  $(S_n)_{n \geq 1}$  is a zero-mean martingale with

$$E(S_n^2) = \sum_{i=1}^n x_i^2 E(\varepsilon_i^2) \leq c \sum_{i=1}^n x_i^2 \leq cx, \quad \forall n \geq 1$$

by the martingale convergence theorem (Theorem A.2.9, Appendix A.2), we obtain that  $S_n \rightarrow 0$  a.s. and in  $L^2$ . Hence  $E(S_n^2) \rightarrow 0$ . But

$$E(S_n^2) \geq (\inf_{i \geq 1} E\varepsilon_i^2) \sum_{i=1}^n x_i^2.$$

Taking  $n \rightarrow \infty$ , we get  $0 \geq (\inf_{i \geq 1} E\varepsilon_i^2)x > 0$ , which is a contradiction.

□

## 2.4 An Alternative Result of Kaufmann

When errors  $(\varepsilon_i)_i$  satisfy  $(\mathbf{C}_2)$ , the strong consistency of the LSE  $\hat{\beta}_n$  can be treated by martingale methods, since  $\hat{\beta}_n - \beta_0 = H_n^{-1} S_n$ , where  $S_n = \sum_{i=1}^n x_i \varepsilon_i$ ,  $n \geq 1$  is a zero-mean martingale with

$$F_n := \text{Cov}(S_n) = \sum_{i=1}^n E(\varepsilon_i^2) x_i x_i^T \leq c H_n, \quad \forall n \geq 1. \quad (20)$$

This approach is considered in [16] and we present in this section.

Our first result (which is weaker than Corollary 2.2.4) is obtained as a direct application of strong law of large numbers for multivariate martingales normalized by the covariance matrix.

**Proposition 2.4.1** *Suppose that in the linear model (1), the errors  $(\varepsilon_i)_i$  satisfy condition  $(\mathbf{C}_2)$  and  $\inf_i E(\varepsilon_i^2) > 0$ . If*

$$\frac{(\log \|H_n\|)^{1+\delta}}{\lambda_{\min}(H_n)} \rightarrow 0 \text{ for some } \delta > 0, \quad (21)$$

then  $\hat{\beta}_n \rightarrow \beta_0$  a.s. (and in  $L^2$ ).

**Proof:** Note that (21) implies  $\|H_n\| \rightarrow \infty$ , which in turn implies  $\|F_n\| \rightarrow \infty$  by (20). By applying Theorem A.2.2 (Appendix A.2) and using (20), (21), we get

$$\begin{aligned} \|\hat{\beta}_n - \beta_0\| &= \|H_n^{-1}S_n\| \leq \frac{1}{[\lambda_{\min}(H_n)]^{1/2}} \|H_n^{-1/2}S_n\| \leq \frac{1}{[\lambda_{\min}(H_n)]^{1/2}} c_1 \|F_n^{-1/2}S_n\| \\ &\leq c_1 \frac{(\log \|F_n\|)^{\frac{1+\delta}{2}}}{[\lambda_{\min}(H_n)]^{1/2}} \cdot \frac{\|F_n^{-1/2}S_n\|}{(\log \|F_n\|)^{\frac{1+\delta}{2}}} \leq c_2 \left[ \frac{(\log \|H_n\|)^{1+\delta}}{\lambda_{\min}(H_n)} \right]^{1/2} \cdot \frac{\|F_n^{-1/2}S_n\|}{(\log \|F_n\|)^{\frac{1+\delta}{2}}} \rightarrow 0, \end{aligned}$$

where  $c_1, c_2 > 0$  are constants.

□

In order to improve this result, we need to employ a different martingale strong law of large numbers.

**Theorem 2.4.2 (Theorem 4, p. 82, [16])** *Suppose that in the linear model (1), the errors  $(\varepsilon_i)$  form a martingale difference sequence and  $\sup_{i \geq 1} E|\varepsilon_i|^p < \infty$  for some  $p \in [1, 2]$ . If*

$$\sum_{i=1}^{\infty} \left[ \frac{x_i^T H_i^{-1} x_i}{\lambda_{\min}(H_i)} \right]^{p/2} < \infty, \quad (22)$$

then  $\hat{\beta}_n \rightarrow \beta_0$  a.s. (and in  $L^p$ ).

**Proof:** We will apply Theorem A.2.3 (Appendix A.2) with  $A_n = [\lambda_{\min}(H_n)]^{1/2} H_n^{1/2}$  and  $S_n = \sum_{i=1}^n x_i \varepsilon_i$ . Note that

$$A_n A_n^T = [\lambda_{\min}(H_n)] H_n \leq [\lambda_{\min}(H_{n+1})] H_{n+1} = A_{n+1} A_{n+1}^T, \quad \forall n \geq 1,$$

$X_i = S_i - S_{i-1} = x_i \varepsilon_i$ , and

$$\sum_{i=1}^{\infty} E \|A_i^{-1} X_i\|^p \leq \sum_{i=1}^{\infty} \|A_i^{-1} x_i\|^p E |\varepsilon_i|^p \leq c \sum_{i=1}^{\infty} (x_i^T A_i^{-2} x_i)^{p/2} = c \sum_{i=1}^{\infty} \left[ \frac{x_i^T H_i^{-1} x_i}{\lambda_{\min}(H_i)} \right]^{p/2} < \infty.$$

We obtain that  $A_n^{-1} S_n \rightarrow 0$  a.s. (and in  $L^p$ ) and hence

$$\|\hat{\beta}_n - \beta_0\| = \|H_n^{-1} S_n\| \leq \|H_n^{-1/2}\| \|H_n^{-1/2} S_n\| = \|A_n^{-1} S_n\| \rightarrow 0 \text{ a.s. and (in } L^p\text{).}$$

□

**Concluding Remarks:** (a) The main problem of Kaufmann's approach is the fact that the norming matrices  $(A_n)_n$  should satisfy the rather stringent monotonicity condition  $A_n A_n^T \leq A_{n+1} A_{n+1}^T$ ,  $\forall n \geq 1$ , which is not a consequence of the "natural" monotonicity condition  $A_n \leq A_{n+1}$ ,  $n \geq 1$ . In particular, in the linear model (1), the normalizing matrices  $H_n = \sum_{i=1}^n x_i x_i^T$ ,  $n \geq 1$  do not satisfy condition  $A_n A_n^T \leq A_{n+1} A_{n+1}^T$ ,  $\forall n \geq 1$ , and had to be replaced by  $A_n = [\lambda_{\min}(H_n)]^{1/2} H_n^{1/2}$ , which lead to condition (22).

(b) Using Theorem A.2.2 (Appendix A.2) we obtain a very quick proof of Theorem 2.2.1, in the case when  $(C_2)$  holds. More precisely, recall that by (3)

$$\hat{\beta}_{n1} - \beta_{01} = \frac{1}{s_n} \sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i) \varepsilon_i = \frac{1}{s_n} u_n.$$

Note that under  $(C_2)$ ,  $(u_n)_{n \geq 1}$  is a zero-mean martingale with

$$\sigma_n^2 := E(u_n^2) = \sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i)^2 E(\varepsilon_i^2) \leq c \sum_{i=1}^n (x_{i1} - K_n H_n^{-1} T_i)^2 = c s_n.$$

By Theorem A.2.2 (Appendix A.2) (case  $p = 1$ ), we have

$$\frac{\hat{\beta}_{n1} - \beta_{01}}{[v_{11}^{(n)} |\log v_{11}^{(n)}|^{1+\delta}]^{1/2}} = \frac{u_n}{[s_n |\log s_n|^{1+\delta}]^{1/2}} \leq c_1 \frac{u_n}{[\sigma_n^2 |\log \sigma_n^2|^{1+\delta}]^{1/2}} \longrightarrow 0 \text{ a.s. (and in } L^2).$$

□

# Chapter 3

## The Generalized Linear Model

In this chapter we consider a generalized regression model, in which there is a non-linear dependence between the responses  $y_i$  and the covariates  $x_i$ . Moreover, we will assume that each variable  $y_i$  has a density which belongs to an exponential family driven by a parameter  $\theta_i$ , where  $\theta_i = x_i^T \beta$ , and  $\beta$  is a  $p$ -dimensional unknown parameter.

This chapter is organized as follows. In Section 3.1, we introduce the framework and the main notation. In Section 3.2, we examine the question of existence of the maximum likelihood estimator (MLE)  $\hat{\beta}_n$ , as well as its asymptotic properties, under various sets of conditions involving the information matrix  $F_n(\beta)$ . In Section 3.3, we examine some particular cases of interest (e.g. the case of bounded regressors), as well as some examples. In Section 3.4, we extend the results of Section 3.2 to the case of the stochastic regressors. Our main reference for Sections 3.1-3.3 is [9], whereas the results we present in Section 3.4 are slightly more general than those of [9].

### 3.1 The Framework

As in linear regression, let  $(y_i)_{i \geq 1}$  be a sequence of response variables, which are recorded together with some explanatory variable  $(x_i)_{i \geq 1}$ . We assume that each  $y_i$  is

a random variable defined on a probability space  $(\Omega, \mathcal{F}, P_\beta)$  and each  $x_i$  is a (non-random)  $p$ -dimensional vector such that

$$\mu_i(\beta) := E_\beta(y_i) = \mu(x_i^T \beta) \text{ and } \sigma_i^2(\beta) := \text{Var}_\beta(y_i) = \phi \mu'(x_i^T \beta), \quad \forall i \geq 1, \quad (23)$$

where  $E_\beta$  denotes the expectation with respect to  $P_\beta$ ,  $\mu(\cdot)$  is a differentiable function,  $\phi$  is a nuisance scale parameter (which we will normally take as  $\phi = 1$ ) and  $\beta$  is an unknown  $p$ -dimensional parameter which has to be estimated.

We define the **errors**

$$\varepsilon_i(\beta) = y_i - \mu(x_i^T \beta), \quad \forall i \geq 1$$

and hence under  $P$ , our model becomes

$$y_i = \mu(x_i^T \beta_0) + \varepsilon_i, \text{ where } E(\varepsilon_i) = 0, \quad \forall i \geq 1.$$

Unlike the linear model, in this chapter we focus only on the case of *independent* errors (i.e. independent measurements  $(y_i)_i$ ).

A generalized linear model which satisfy (23) arises in the following situation: assume that each  $y_i$  has a density function which belongs to the following exponential family:

$$f(y_i | \theta_i) = c(y_i) \exp \left\{ \frac{\theta_i y_i - a(\theta_i)}{\phi} \right\}, \quad (24)$$

where  $c(\cdot)$  is a positive function,  $a(\cdot)$  is an arbitrary function and  $\theta_i$  is a 1-dimensional parameter.

By denoting  $\mu = a'$  we obtain

$$E(y_i) = \mu(\theta_i) \text{ and } \text{Var}(y_i) = \phi \mu'(\theta_i). \quad (25)$$

Moreover, we suppose that there exists an injection function (called the **link function**) which relates the mean  $\mu(\theta_i)$  of  $y_i$  with the linear combination  $x_i^T \beta$  by the following relation:

$$x_i^T \beta = g(\mu(\theta_i)), \quad \forall i \geq 1,$$

or equivalently

$$\theta_i = u(x_i^T \beta), \text{ where } u = (g \circ \mu)^{-1}.$$

In the present work, we consider only the case of the **natural link function**, i.e.  $\mu = g^{-1}$ , which leads to the relation

$$\theta_i = x_i^T \beta, \quad \forall i \geq 1. \quad (26)$$

From (25) and (26), we see immediately that (23) holds. For the remaining part of this chapter, we will work under the assumption that (24) and (26) hold, and hence

$$f(y_i|\beta) = c(y_i) \exp \left\{ \frac{(x_i^T \beta)y_i - a(x_i^T \beta)}{\phi} \right\}.$$

Therefore, we can write down the likelihood function of  $\beta$  as:

$$L_n(\beta) = \left[ \prod_{i=1}^n c(y_i) \right] \exp \left\{ \frac{1}{\phi} \sum_{i=1}^n [(x_i^T \beta)y_i - a(x_i^T \beta)] \right\}.$$

The log-likelihood function is

$$l_n(\beta) = \log L_n(\beta) = \sum_{i=1}^n [\log c(y_i)] + \frac{1}{\phi} \sum_{i=1}^n [x_i^T \beta y_i - a(x_i^T \beta)].$$

The score function is

$$s_n(\beta) = \frac{\partial l_n}{\partial \beta}(\beta) = \frac{1}{\phi} \sum_{i=1}^n [x_i y_i - x_i a'(x_i^T \beta)] = \frac{1}{\phi} \sum_{i=1}^n x_i (y_i - \mu(x_i^T \beta)).$$

The MLE of  $\beta$  is denoted by  $\hat{\beta}_n$  and it is the point where the function  $l_n(\cdot)$  attains its maximum. Note that the function  $l_n$  is strictly concave, since

$$\frac{\partial^2 l_n}{\partial \beta^2}(\beta) = \frac{\partial s_n}{\partial \beta}(\beta) = -\frac{1}{\phi} \sum_{i=1}^n x_i \mu'(x_i^T \beta) x_i^T = -\frac{1}{\phi^2} \sum_{i=1}^n x_i \sigma_i^2(\beta) x_i^T < 0, \quad \forall \beta.$$

Therefore  $\hat{\beta}_n$  is unique, if it exists, and it is the solution of the equation  $s_n(\beta) = 0$ .

Let

$$F_n(\beta) = \sum_{i=1}^n x_i \sigma_i^2(\beta) x_i^T$$

be the **information matrix**, where we assume that  $\Phi = 1$ . Note that

$$F_n(\beta) = -\frac{\partial s_n}{\partial \beta}(\beta).$$

**Lemma 3.1.1** *We have  $F_n(\beta) = \text{Cov}_\beta(s_n(\beta))$ .*

**Proof:** Note that

$$E(s_n(\beta)) = E\left(\sum_{i=1}^n x_i \varepsilon_i(\beta)\right) = \sum_{i=1}^n x_i E(\varepsilon_i(\beta)) = 0.$$

and

$$\begin{aligned} \text{Cov}(s_n(\beta)) &= E(s_n(\beta)s_n(\beta)^T) = E\left[\left(\sum_{i=1}^n x_i \varepsilon_i(\beta)\right)\left(\sum_{j=1}^n x_j^T \varepsilon_j(\beta)\right)\right] = \\ &= E\left[\sum_{j=1}^n \sum_{i=1}^n x_i (\varepsilon_i(\beta)\varepsilon_j(\beta)) x_j^T\right] = \sum_{i=1}^n \sum_{j=1}^n [x_i E(\varepsilon_i(\beta)\varepsilon_j(\beta)) x_j^T] = \\ &= \sum_{i=1}^n x_i E(\varepsilon_i^2(\beta)) x_i^T = \sum_{i=1}^n x_i \sigma_i^2(\beta) x_i^T = F_n(\beta). \end{aligned}$$

□

Let  $\Theta = \{\theta \in R \mid 0 < \int c(y) \exp(\theta y) dy < \infty\}$  be the natural parameter space and  $\Theta^0$  be the interior of  $\Theta$ . Let  $M$  denote the image  $\mu(\Theta^0)$  of  $\Theta^0$ . We impose the following regularity assumptions:

- (1)  $\beta$  lies in an open set  $B \subseteq R^p$ .
- (2)  $x_i^T \beta \in g(M)$  for all  $\beta \in B$  and for all  $i \geq 1$ .
- (3)  $\mu$  is twice continuously differentiable and  $\mu'(\theta) > 0$  for all  $\theta \in \Theta^0$ .
- (4)  $F_n$  is positive definite for all  $n \geq n_0$ .

## 3.2 The General Results

In this section we present the general asymptotic results which can be obtained for a generalized linear model, under the assumption that the likelihood function of the response variables  $y_i$  is fully specified (and it belongs to an exponential family).

### 3.2.1 Asymptotic Existence and Weak Consistency

We define the normed information matrix as

$$V_n(\beta) = F_n^{-1/2} F_n(\beta) F_n^{-T/2}.$$

The existence of a sequence of weakly consistent estimators of  $\beta$  is obtained under the following conditions:

(D)  $\lambda_{\min}(F_n) \longrightarrow \infty$ .

(C) There exists  $c > 0$  such that for any  $r > 0$ , there exists  $n_1 = n_1(r)$  with

$$F_n(\beta) \geq cF_n, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1,$$

where  $N_n(r) = \{\beta \in B; \|F_n^{T/2}(\beta - \beta_0)\| < r\}$ .

**Theorem 3.2.1 (Theorem 1, p. 349, [9])** *Under (D) and (C), there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that*

(a)  $P(s_n(\hat{\beta}_n) = 0) \longrightarrow 1$ .

(b)  $\hat{\beta}_n \longrightarrow \beta_0$  in probability.

**Proof:** (p. 351, [9]) (a) Let  $c > 0$  be the constant given by condition (C). For every  $r > 0$ ,  $n \geq 1$ , we consider the event

$$E_n(r) = \{l_n(\beta) - l_n(\beta_0) < 0 \text{ for all } \beta \in \partial N_n(r)\}.$$

By Theorem A.3.1 (Appendix A.3), on the event  $E_n(r)$ , the local maximum of  $l_n(\cdot)$  exists and we denote it by  $\hat{\beta}_n$ . Hence

$$E_n(r) \subseteq \{\text{there exists } \hat{\beta}_n \in N_n(r) \text{ such that } s_n(\hat{\beta}_n) = 0\}.$$

To prove part (a) of the theorem, we will show that for any  $\varepsilon > 0$  there exists  $r = r(\varepsilon)$  such that

$$P(E_n(r)) \geq 1 - \varepsilon, \quad \forall n \geq n_1, \quad (27)$$

where  $n_1 = n_1(r)$  is given by condition (C).

To prove (27), let  $\varepsilon > 0$  be arbitrary and  $r = r(\varepsilon)$  be a constant to be specified later. Let  $\beta \in \partial N_n(r)$  be arbitrary. Using the Taylor's expansion of  $l_n(\cdot)$  around  $\beta_0$ , we get

$$l_n(\beta) - l_n(\beta_0) = (\beta - \beta_0)^T l_n'(\beta_0) + (1/2)(\beta - \beta_0)^T l_n''(\tilde{\beta}_n)(\beta - \beta_0),$$

where  $\tilde{\beta}_n$  is a random vector which lies between  $\beta$  and  $\beta_0$ . Denote  $\lambda = (1/r)F_n^{T/2}(\beta - \beta_0)$  and note that  $\|\lambda\| = 1$ , since  $\beta \in \partial N_n(r)$ . Hence

$$\begin{aligned} l_n(\beta) - l_n(\beta_0) &= (\beta - \beta_0)^T s_n - (1/2)(\beta - \beta_0)^T F_n(\tilde{\beta}_n)(\beta - \beta_0) = \\ &= \frac{(\beta - \beta_0)^T F_n^{1/2}}{r} r F_n^{-1/2} s_n - (1/2) \frac{(\beta - \beta_0)^T F_n^{1/2}}{r} r^2 F_n^{-1/2} F_n(\tilde{\beta}_n) F_n^{-T/2} \frac{F_n^{T/2}(\beta - \beta_0)}{r} = \\ &= r \lambda^T F_n^{-1/2} s_n - (1/2) r^2 \lambda^T V_n(\tilde{\beta}_n) \lambda, \quad \forall \beta \in \partial N_n(r). \end{aligned} \quad (28)$$

For the first term,

$$r \lambda^T F_n^{-1/2} s_n \leq r \|F_n^{-1/2} s_n\|. \quad (29)$$

For the second term, with  $n_1 = n_1(r)$  given by (C) we have:

$$r^2 \lambda^T V_n(\tilde{\beta}_n) \lambda \geq r^2 \lambda_{\min}(V_n(\tilde{\beta}_n)) \lambda^T \lambda \geq r^2 c, \quad \forall n \geq n_1, \quad (30)$$

since  $V_n(\tilde{\beta}_n) = F_n^{-1/2} F_n(\tilde{\beta}_n) F_n^{-T/2} \geq c F_n^{-1/2} F_n F_n^{-T/2} = cI$  by condition (C).

Using (28), (29) and (30), we get

$$l_n(\beta) - l_n(\beta_0) \leq r \|F_n^{-1/2} s_n\| - (1/2) r^2 c, \quad \forall \beta \in \partial N_n(r), \quad \forall n \geq n_1.$$

Hence

$$P(E_n(r)) \geq P\{\|F_n^{-1/2} s_n\| < (1/2)rc\}, \quad \forall n \geq n_1. \quad (31)$$

By the Chebyshev's inequality, we have:

$$\begin{aligned} P\{\|F_n^{-1/2} s_n\| < (1/2)rc\} &= 1 - P\{\|F_n^{-1/2} s_n\| \geq (1/2)rc\} \geq 1 - \frac{1}{[(1/2)rc]^2} E\|F_n^{-1/2} s_n\|^2 \\ &= 1 - \frac{4}{r^2 c^2} E[\text{tr}(F_n^{-1/2} s_n s_n^T F_n^{-T/2})] = 1 - \frac{4}{r^2 c^2} \text{tr}[F_n^{-1/2} E(s_n s_n^T) F_n^{-T/2}] \\ &= 1 - \frac{4}{r^2 c^2} \text{tr}(I) = 1 - \frac{4}{r^2 c^2} p > 1 - \varepsilon, \quad \forall n \geq 1, \end{aligned} \quad (32)$$

by choosing  $r = \sqrt{(4p)/(c^2\varepsilon)}$ . Relation (27) follows from (31) and (32).

(b) Let  $\eta, \varepsilon > 0$  be arbitrary. We need to prove that there exists an  $N = N_{\eta, \varepsilon}$  such that

$$P\left(\|\hat{\beta}_n - \beta_0\| \leq \eta\right) \geq 1 - \varepsilon \quad \forall n \geq N. \quad (33)$$

In part (a) we proved the existence of  $\hat{\beta}_n$  on the event  $E_n(r)$ , where  $r = \sqrt{(4p)/(c^2\varepsilon)}$ . Moreover,  $\hat{\beta}_n \in N_n(r)$ .

By (D), there exists  $N = N_\eta > n_1$  such that  $\|F_n\| \geq (r/\eta)^2$ ,  $\forall n \geq N$ . Since  $\hat{\beta}_n \in N_n(r)$ , on the event  $E_n(r)$  we have:

$$\|\hat{\beta}_n - \beta_0\| \leq \|F_n^{-T/2}\| \|F_n^{T/2}(\hat{\beta}_n - \beta_0)\| \leq \frac{r}{\|F_n\|^{1/2}} \leq \eta, \quad \forall n \geq N$$

and therefore, by (27)

$$1 - \varepsilon \leq P(E_n(r)) \leq P\left(\|\hat{\beta}_n - \beta_0\| \leq \eta\right), \quad \forall n \geq N.$$

This concludes the proof of (33).

□

### 3.2.2 Asymptotic Normality

We consider the condition:

(N) For all  $r > 0$ , as  $n \rightarrow \infty$

$$\max_{\beta \in N_n(r)} \|V_n(\beta) - I\| \rightarrow 0,$$

where  $N_n(r) = \{\beta \in B; \|F_n^{T/2}(\beta - \beta_0)\| < r\}$ .

**Lemma 3.2.2** (p. 349, [9]) (N) *implies* (C).

**Proof:** (N) is equivalent to the following assertion: for any  $r > 0$  and  $\eta > 0$ , there exists  $n_1 = n_1(r, \eta)$  such that

$$\|V_n(\beta) - I\| \leq \eta, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1,$$

i.e.

$$\|F_n^{-1/2}(F_n(\beta) - F_n)F_n^{-T/2}\| \leq \eta, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1.$$

By Corollary A.1.17 (Appendix A.1), this becomes

$$(\mathbf{N}') \quad \left| \lambda^T F_n^{-1/2}(F_n(\beta) - F_n)F_n^{-T/2} \lambda \right| \leq \eta(\lambda^T \lambda) \quad \forall \lambda \in R^p, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1.$$

By choosing  $x = F_n^{-T/2}\lambda$ , the previous relation can be written as

$$(\mathbf{N}'') \quad |x^T F_n(\beta)x - x^T F_n x| \leq \eta(x^T F_n x), \quad \forall x \in R^p, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1.$$

In particular, this implies:

$$x^T F_n(\beta)x \geq (1 - \eta)x^T F_n x, \quad \forall x \in R^p, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1$$

i.e.

$$F_n(\beta) \geq (1 - \eta)F_n, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1.$$

By fixing  $\eta \in (0, 1)$  and letting  $c = 1 - \eta$ , we obtain exactly **(C)**.

□

**Lemma 3.2.3 (Lemma 1, p. 349, [9])** *Under (D) and (N)*

$$F_n^{-1/2}s_n \longrightarrow N(0, I) \text{ in distribution.}$$

**Proof:** (p. 352, [9]) By the Cràmer-Wold theorem (Theorem A.2.4, Appendix A.2 ) it suffices to show that for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ , we have

$$\lambda^T F_n^{-1/2}s_n \longrightarrow N(0, 1) \text{ in distribution.} \quad (34)$$

Let  $\lambda \in R^p$  with  $\|\lambda\| = 1$  be arbitrary. By the continuity theorem of the moment generating function (Theorem A.2.6, Appendix A.2), (34) is equivalent to:

$$E[\exp(r\lambda^T F_n^{-1/2}s_n)] \longrightarrow \exp(r^2/2), \quad \forall r \in R, \quad (35)$$

since  $\phi(r) = e^{r^2/2}$  is the moment generating function of the  $N(0, 1)$  distribution.

Suppose first that  $r > 0$  is arbitrary. Let  $\beta_n = \beta_0 + rF_n^{-T/2}\lambda$  and note that using (28), there exists  $\tilde{\beta}_n$  between  $\beta_n$  and  $\beta_0$  such that :

$$l_n(\beta_n) = l_n(\beta_0) + r\lambda^T F_n^{-1/2}s_n - (1/2)r^2\lambda^T V_n(\tilde{\beta}_n)\lambda$$

i.e.

$$(1/2)r^2\lambda^T V_n(\tilde{\beta}_n)\lambda + l_n(\beta_n) = r\lambda^T F_n^{-1/2}s_n + l_n(\beta_0). \quad (36)$$

Taking the exponential in both sides of (36), we get:

$$\exp\left((r^2/2)\lambda^T V_n(\tilde{\beta}_n)\lambda\right) L_n(\beta_n) = \exp(r\lambda^T F_n^{-1/2} s_n) L_n(\beta_0). \quad (37)$$

Recall that

$$L_n(\beta) \text{ is the density function of } (y_1, \dots, y_n) \text{ under } P_\beta. \quad (38)$$

Integrating both sides of (37) with respect to  $dy_1 \dots dy_n$ , and using (38) we get

$$E_{\beta_n} \left[ \exp\left(\frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda\right) \right] = E \left[ \exp(r\lambda^T F_n^{-1/2} s_n) \right].$$

Therefore, in order to prove (35), it suffices to show that

$$E_{\beta_n} \left[ \exp\left(\frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda\right) \right] \longrightarrow \exp(r^2/2). \quad (39)$$

Using **(N)** (in its equivalent form **(N')**) and the fact  $\tilde{\beta}_n \in N_n(r)$ , we know that there exists  $n_1(\eta)$  such that

$$|\lambda^T V_n(\tilde{\beta}_n)\lambda - 1| < \eta, \quad \forall n \geq n_1(\eta).$$

Multiplying by  $r^2/2$ , we get

$$\left| \frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda - \frac{r^2}{2} \right| < \frac{r^2}{2}\eta, \quad \forall n \geq n_1(\eta). \quad (40)$$

Let  $\varepsilon > 0$  be arbitrary. By the continuity of the function  $f(x) = \exp(x)$  at  $r^2/2$ , there exists  $\delta_\varepsilon > 0$  such that for all  $x$  with  $|x - (r^2/2)| < \delta_\varepsilon$  we have

$$|\exp(x) - \exp(r^2/2)| < \varepsilon. \quad (41)$$

Choose  $\eta_\varepsilon$  such that  $\eta_\varepsilon(r^2/2) < \delta_\varepsilon$ . From (40) and (41), we obtain

$$\left| \exp\left(\frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda\right) - \exp\left(\frac{1}{2}r^2\right) \right| \leq \varepsilon, \quad \forall n \geq n_1(\eta_\varepsilon).$$

Multiplying by  $L_n(\beta_n)$  and integrating with respect to  $dy_1 \dots dy_n$ , we obtain:

$$\left| \int_{\mathbb{R}^n} \exp\left(\frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda\right) L_n(\beta_n) dy_1 \dots dy_n - \int_{\mathbb{R}^n} \exp(r^2/2) L_n(\beta_n) dy_1 \dots dy_n \right| \leq$$

$$\leq \int_{R^n} \left| \exp\left(\frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda\right) - \exp(r^2/2) \right| L_n(\beta_n) dy_1 \dots dy_n \leq \int_{R^n} \varepsilon L_n(\beta_n) dy_1 \dots dy_n = \varepsilon,$$

for any  $n \geq n_1(\eta_\varepsilon)$ . Using (38), this becomes:

$$\left| E_{\beta_n} \left[ \exp\left(\frac{1}{2}r^2\lambda^T V_n(\tilde{\beta}_n)\lambda\right) \right] - \exp(r^2/2) \right| \leq \varepsilon, \quad \forall n \geq n_1(\eta_\varepsilon),$$

which concludes the proof of (39).

□

**Theorem 3.2.4 (Theorem 3, p. 349, [9])** *Under (D) and (N), we have*

$$F_n^{T/2}(\hat{\beta}_n - \beta_0) \longrightarrow N(0, I) \text{ in distribution.}$$

**Proof:** We focus on the event  $\{s_n(\hat{\beta}_n) = 0\}$  whose probability converges to 1, by Theorem 3.2.1. Using Taylor's expansion for the function  $F_n^{-1/2}s_n(\beta)$  around  $\beta_0$ , we get

$$F_n^{-1/2}s_n = -F_n^{-1/2}\left(s_n(\hat{\beta}_n) - s_n\right) = F_n^{-1/2}F_n(\tilde{\beta}_n)(\hat{\beta}_n - \beta_0) = V_n(\tilde{\beta}_n)F_n^{T/2}(\hat{\beta}_n - \beta_0), \quad (42)$$

where  $\tilde{\beta}_n$  is a random vector which lies between  $\hat{\beta}_n$  and  $\beta_0$ .

Note that

$$E\left(\|F_n^{-1/2}s_n\|^2\right) = \text{tr}\left(F_n^{-1/2}E(s_n s_n^T)F_n^{-T/2}\right) = \text{tr}(I) = p, \quad \forall n \geq 1. \quad (43)$$

We claim that

$$\|V_n(\tilde{\beta}_n) - I\| \longrightarrow 0 \text{ in probability.} \quad (44)$$

Using Lemma A.2.7 (Appendix A.2), (42), (43) and (44) yields

$$F_n^{T/2}(\hat{\beta}_n - \beta_0) = F_n^{-1/2}s_n + o_p(1).$$

The conclusion follows using Lemma 3.2.3 and Slutsky's theorem.

To prove (44), let  $\varepsilon > 0$  and  $\eta > 0$  be arbitrary. In the proof of part (a) of Theorem 3.2.1, we showed that there exists  $r = r(\varepsilon)$  and  $n_0 = n_0(\varepsilon)$  such that

$$P\left(\hat{\beta}_n \in N_n(r)\right) \geq 1 - \varepsilon, \quad \forall n \geq n_0.$$

We have

$$\|V_n(\beta) - I\| \leq \eta, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_1. \quad (45)$$

Therefore, the event  $\{\tilde{\beta}_n \in N_n(r)\}$  is contained in the event  $\{\|V_n(\tilde{\beta}_n) - I\| \leq \eta\}$ , for every  $n \geq n_1$  and hence

$$P\left(\tilde{\beta}_n \in N_n(r)\right) \leq P\left(\|V_n(\tilde{\beta}_n) - I\| \leq \eta\right), \quad \forall n \geq n_1. \quad (46)$$

From (45) and (46) we get

$$P\left(\|V_n(\tilde{\beta}_n) - I\| \leq \eta\right) \geq 1 - \varepsilon, \quad \forall n \geq n_1,$$

which concludes the proof of (44).

□

### 3.2.3 Asymptotic Existence and Strong Consistency

Before we present the strong consistency result, we look at the following lemma.

**Lemma 3.2.5** *Under (D), for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ , we have*

$$\frac{\lambda^T s_n}{[\lambda_{\max}(F_n)]^{1/2+\delta}} \longrightarrow 0 \text{ a.s.}$$

**Proof:** This follows immediately from Theorem A.2.1 (Appendix A.2), since  $(\lambda^T s_n)_{n \geq 1}$  is a sum of independent zero-mean random variables with

$$E[(\lambda^T s_n)^2] = \lambda^T F_n \lambda \leq \lambda_{\max}(F_n).$$

□

The existence of a sequence of strongly consistent estimators is obtained under (D) and the following condition:

(S<sub>δ</sub>) There exist some constants  $c > 0$ ,  $\delta > 0$ ,  $n_1 > 1$  and a neighborhood  $N \subset B$  of  $\beta_0$  such that

$$\lambda_{\min}[F_n(\beta)] \geq c[\lambda_{\max}(F_n)]^{(1/2)+\delta}, \quad \forall \beta \in N, \quad \forall n \geq n_1.$$

**Remark:** Note that (S<sub>1/2</sub>) implies (C): for any  $p$ -dimensional vector  $\lambda$ , we have

$$\lambda^T F_n(\beta) \lambda \geq \lambda_{\min}[F_n(\beta)] \lambda^T \lambda \geq c \lambda_{\max}(F_n) \lambda^T \lambda \geq c \lambda^T F_n \lambda.$$

**Theorem 3.2.6 (Theorem 2, p. 349, [9])** Under **(D)** and **(S<sub>δ</sub>)**, there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables and a random number  $n_2$  with

(a)  $P\{s_n(\hat{\beta}_n) = 0 \text{ for all } n \geq n_2\} = 1.$

(b)  $\hat{\beta}_n \longrightarrow \beta_0$  a.s.

**Proof:** (p. 351, [9]) Choose  $\varepsilon_0 > 0$  such that the neighborhood  $B_{\varepsilon_0}(\beta_0) = \{\beta : \|\beta - \beta_0\| < \varepsilon_0\} \subseteq N$ . We will show that with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$  there exists a random number  $n_2 = n_2(\varepsilon)$  such that

$$l_n(\beta) - l_n(\beta_0) < 0, \quad \forall \beta \in \partial B_\varepsilon(\beta_0), \quad \forall n \geq n_2. \quad (47)$$

This will imply the conclusion of our theorem, in the following format: with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$ , there exists  $n_2 = n_2(\varepsilon)$  (random) such that for any  $n \geq n_2$ ,

$$\text{there exists } \hat{\beta}_n \in B_\varepsilon(\beta_0) \text{ with } s_n(\hat{\beta}_n) = 0.$$

(The fact that  $\hat{\beta}_n$  does not depend on  $\varepsilon$ , follows from the uniqueness of the zeros' of the function  $s_n(\cdot)$ . It is also clear that  $\hat{\beta}_n \longrightarrow \beta_0$  a.s.)

To prove (47), we use Taylor's expansion for  $l_n(\cdot)$  around  $\beta_0$ : for  $\beta \in \partial B_\varepsilon(\beta_0)$ ,  $\varepsilon \in (0, \varepsilon_0)$  and  $\lambda = (\beta - \beta_0)/\varepsilon$ , we have

$$\begin{aligned} l_n(\beta) - l_n(\beta_0) &= (\beta - \beta_0)^T l'_n(\beta_0) + (1/2)(\beta - \beta_0)^T l''_n(\tilde{\beta}_n)(\beta - \beta_0) = \\ &= (\beta - \beta_0)^T s_n - (1/2)(\beta - \beta_0)^T F_n(\tilde{\beta}_n)(\beta - \beta_0) = \varepsilon \lambda^T s_n - (1/2)\varepsilon^2 \lambda^T F_n(\tilde{\beta}_n)\lambda, \end{aligned}$$

where  $\tilde{\beta}_n$  is a random vector which lies between  $\beta$  and  $\beta_0$ . It follows that

$$l_n(\beta) - l_n(\beta_0) < \varepsilon \lambda^T s_n - \frac{1}{2}\varepsilon^2 c[\lambda_{\max}(F_n)]^{1/2+\delta}, \quad \forall \beta \in \partial B_\varepsilon(\beta_0), \quad \forall n \geq n_1, \quad (48)$$

since by condition **(S<sub>δ</sub>)**

$$\lambda^T F_n(\tilde{\beta}_n)\lambda \geq \lambda_{\min}[F_n(\tilde{\beta}_n)] \geq c[\lambda_{\max}(F_n)]^{1/2+\delta}, \quad \forall n \geq n_1.$$

From Lemma 3.2.5, it follows that with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$ , there exists a random number  $n_2 = n_2(\varepsilon) > n_1$  such that

$$\frac{\lambda^T s_n}{[\lambda_{\max}(F_n)]^{1/2+\delta}} < \frac{\varepsilon}{2}c, \quad \forall n \geq n_2. \quad (49)$$

From (48) and (49), we obtain (47).

□

### 3.3 Particular Cases and Examples

In this section, we will introduce some particular cases and examples.

#### 3.3.1 Bounded Regressors

In this subsection we suppose that the regressors  $(x_i)_{i \geq 1}$  satisfy

$$K := \sup_{i \geq 1} \|x_i\| < \infty. \quad (50)$$

Let  $H_n = \sum_{i=1}^n x_i x_i^T$ . We consider the following condition:

(D\*)  $\lambda_{\min}(H_n) \rightarrow \infty$ .

(S\* $_{\delta}$ ) There exist  $c > 0$  and  $n_1 \geq 1$  such that

$$\lambda_{\min}(H_n) \geq c[\lambda_{\max}(H_n)]^{\frac{1}{2} + \delta}, \quad \forall n \geq n_1.$$

In addition, we introduce the following assumption:

$$(AH) \quad \sup_{\beta \in N_n(r)} \max_{i \leq n} \left| \frac{\mu''(x_i^T \beta)}{\mu'(x_i^T \beta)} \right| \leq c_0, \quad \forall n \geq 1.$$

We have the following results:

**Corollary 3.3.1 (Corollary 1, p. 355, [9])** *Suppose that  $(x_i)_{i \geq 1}$  satisfy (50). Then (D\*) is equivalent to (D), and (S\* $_{\delta}$ ) is equivalent to (S $_{\delta}$ ). Moreover, if (D) and (AH) hold, then (N) holds.*

**Proof:** (p. 358-359, [9]) We have

$$|x_i^T \beta_0| \leq \|x_i\| \|\beta_0\| \leq K \|\beta_0\| \quad \forall i \geq 1.$$

Since  $\mu'$  is continuous and a continuous function maps a compact set into a compact, there exist  $c_1, c_2 > 0$  such that

$$0 < c_1 \leq \mu'(x_i^T \beta_0) \leq c_2, \quad \forall i \geq 1.$$

Hence

$$c_1 H_n \leq F_n \leq c_2 H_n, \quad c_1 \lambda_{\min}(H_n) \leq \lambda_{\min}(F_n) \leq c_2 \lambda_{\min}(H_n).$$

and we see that  $(\mathbf{D}^*)$  is equivalent to  $(\mathbf{D})$ .

To prove that  $(\mathbf{S}_\delta^*)$  is equivalent to  $(\mathbf{S}_\delta)$ , let  $N$  be an arbitrary neighborhood of  $\beta_0$ , say  $N = \{\beta : \|\beta - \beta_0\| < \varepsilon_0\}$ . For any  $\beta \in N$ , we have

$$|x_i^T \beta| = |x_i^T (\beta - \beta_0 + \beta_0)| \leq K(\|\beta - \beta_0\| + \|\beta_0\|) \leq K(\varepsilon_0 + \|\beta_0\|), \quad \forall i \geq 1$$

and hence  $c'_1 \leq \mu'(x_i^T \beta) \leq c'_2, \forall i \geq 1, \forall \beta \in N$ .

From here, we conclude that  $c'_1 H_n \leq F_n(\beta) \leq c'_2 H_n$ . This forces

$$c'_1 \lambda_{\min}(H_n) \leq \lambda_{\min}[F_n(\beta)] < \lambda_{\max}[F_n(\beta)] \leq c'_2 \lambda_{\max}(H_n).$$

and hence  $(\mathbf{S}_\delta^*)$  implies  $(\mathbf{S}_\delta)$ :

$$\frac{\lambda_{\min}[F_n(\beta)]}{[\lambda_{\max}(F_n)]^{1/2+\delta}} \geq \frac{c'_1 \lambda_{\min}(H_n)}{(c'_2)^{1/2+\delta} [\lambda_{\max}(H_n)]^{1/2+\delta}} \geq \frac{c'_1}{(c'_2)^{1/2+\delta}} c > 0.$$

Similarly  $(\mathbf{S}_\delta)$  implies  $(\mathbf{S}_\delta^*)$ .

To prove  $(\mathbf{N})$ , let  $r > 0$  and  $\varepsilon > 0$  be arbitrary. We want to prove that there exists  $n_0 = n_0(r, \varepsilon)$  such that

$$|\lambda^T F_n(\beta) \lambda - \lambda^T F_n \lambda| \leq \varepsilon \lambda^T F_n \lambda, \quad \forall \lambda \in R^p, \quad \forall \beta \in N_n(r), \quad \forall n \geq n_0. \quad (51)$$

Since  $(\mathbf{D})$  holds, there exists  $n_1 = n_1(\varepsilon, r)$ , such that  $N_n(r) \subseteq \{\beta : \|\beta - \beta_0\| < \varepsilon\}$ ,  $\forall n \geq n_1$ .

For each  $i \leq n$ , using the Taylor's expansion for the function  $\sigma_i(\cdot)$  around  $\beta_0$ , we get: for  $\beta \in N_n(r)$  and  $n \geq n_1$

$$\sigma_i^2(\beta) - \sigma_i^2(\beta_0) = \frac{d\sigma_i^2}{d\beta}(\tilde{\beta}_i)(\beta - \beta_0)^T = \mu''(x_i^T \tilde{\beta}_i) x_i (\beta - \beta_0)^T,$$

where  $\tilde{\beta}_i$  lies between  $\beta$  and  $\beta_0$ . Using  $(\mathbf{AH})$ , we get

$$|\sigma_i^2(\beta) - \sigma_i^2(\beta_0)| \leq |\mu''(x_i^T \tilde{\beta}_i)| \|x_i\| \|\beta - \beta_0\| \leq c_0 \mu'(x_i^T \beta_0) c \varepsilon = c_0 c \varepsilon \sigma_i^2(\beta_0).$$

Multiplying with  $\lambda^T x_i x_i^T \lambda$  and taking the sum over  $i = 1, \dots, n$ , we get

$$\left| \lambda^T \sum_{i=1}^n [x_i \sigma_i^2(\beta) - \sigma_i^2(\beta_0) x_i^T] \lambda \right| \leq \sum_{i=1}^n |\lambda^T [x_i \sigma_i^2(\beta) - \sigma_i^2(\beta_0) x_i^T] \lambda| \leq c_0 c \varepsilon \sum_{i=1}^n \lambda^T x_i \sigma_i^2(\beta_0) x_i \lambda,$$

which is exactly (51).

□

### 3.3.2 Linear Regression

In this subsection, we suppose that  $y_i$  is normally distributed with mean  $\theta_i = x_i^T \beta$  and variance 1. Hence

$$f(y_i|\theta_i) = c(y_i) \exp \left\{ \theta_i y_i - \frac{\theta_i^2}{2} \right\}$$

In this case  $\mu(x) = x$  and  $F_n(\beta) = \sum_{i=1}^n x_i x_i^T = H_n$  for any  $\beta$ . The score function is

$$s_n(\beta) = \sum_{i=1}^n x_i (y_i - x_i^T \beta),$$

and the MLE  $\hat{\beta}_n$  coincides with the LSE. Since  $V_n(\beta) = I$ , condition **(N)** holds automatically. Hence **(C)** holds. Condition **(S $_{\delta}$ )** becomes: there exists  $c > 0$ ,  $\delta > 0$  and  $n_1 \geq 1$  such that

$$\lambda_{\min}(H_n) \geq c[\lambda_{\max}(H_n)]^{1/2+\delta}, \forall n \geq n_1.$$

From Chapter 2, we know that the strong consistency of  $\hat{\beta}$  is obtained under **(D)** alone. Therefore, condition **(S $_{\delta}$ )** is not needed.

### 3.3.3 Poisson Log-linear Regression

In this subsection, we suppose that  $y_i$  is a count measurement with the following Poisson density:

$$f(y_i|\lambda_i) = e^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!} = (1/y_i!) \exp(y_i \ln \lambda_i - \lambda_i)$$

From here we see that the natural link function for this model is

$$\ln \lambda_i = \theta_i = x_i^T \beta.$$

Since  $\mu(\theta_i) = E(y_i) = \lambda_i = e^{\theta_i}$ , it follows that  $\mu(\theta) = e^{\theta}$ . We have

$$\mu_i(\beta) = \sigma_i^2(\beta) = e^{x_i^T \beta} \text{ and } F_n(\beta) = \sum_{i=1}^n (e^{x_i^T \beta}) x_i x_i^T.$$

We have the following results:

**Proposition 3.3.2** (p. 353, [9]) *We suppose that (D) holds and*

$$x_n^T F_n^{-1} x_n \longrightarrow 0. \quad (52)$$

*Then (N) holds.*

**Proof:** (p. 357, [9]) By the Cauchy-Schwarz inequality, we have: for any  $\beta \in N_n(r)$

$$\begin{aligned} |x_n^T(\beta - \beta_0)|^2 &= |x_n^T F_n^{-T/2} F_n^{T/2}(\beta - \beta_0)|^2 \leq \|x_n^T F_n^{-T/2}\|^2 \|F_n^{T/2}(\beta - \beta_0)\|^2 \\ &\leq x_n^T F_n^{-T/2} F_n^{-1/2} x_n r^2 = r^2 (x_n^T F_n^{-1} x_n), \forall r > 0. \end{aligned}$$

From (52), we get that for any  $r > 0$ ,

$$\max_{\beta \in N_n(r)} |\exp[x_n^T(\beta - \beta_0)] - 1| \longrightarrow 0.$$

Hence for any  $\varepsilon > 0$ , there exists  $N_\varepsilon$  such that

$$\left| \frac{\sigma_n^2(\beta)}{\sigma_n^2(\beta_0)} - 1 \right| \leq \varepsilon, \forall \beta \in N_n(r), \forall n \geq N_\varepsilon,$$

that is

$$|\sigma_n^2(\beta) - \sigma_n^2(\beta_0)| \leq \varepsilon \sigma_n^2(\beta_0), \forall \beta \in N_n(r), \forall n \geq N_\varepsilon.$$

From the continuity of  $\sigma_i^2(\beta)$ , the divergence and monotony of  $F_n$ , we get

$$|\sigma_i^2(\beta) - \sigma_i^2(\beta_0)| \leq \varepsilon \sigma_i^2(\beta_0), \forall i \leq n, \forall \beta \in N_n(r), \forall n \geq N_\varepsilon.$$

Multiplying by  $\lambda^T x_i x_i^T \lambda$ , and taking the sum over  $i = 1, \dots, n$ , we obtain

$$\begin{aligned} |\lambda^T F_n(\beta) \lambda - \lambda^T F_n \lambda| &= \left| \sum_{i=1}^n \lambda^T x_i \sigma_i^2(\beta) x_i^T \lambda - \sum_{i=1}^n \lambda^T x_i \sigma_i^2(\beta_0) x_i^T \lambda \right| \\ &\leq \sum_{i=1}^n |\sigma_i^2(\beta) - \sigma_i^2(\beta_0)| \|\lambda^T x_i\|^2 \leq \varepsilon \sum_{i=1}^n \sigma_i^2(\beta_0) \|\lambda^T x_i\|^2 = \varepsilon \lambda^T F_n \lambda, \end{aligned}$$

which is equivalent to (N).

□

**Proposition 3.3.3** (p. 354, [9]) *(For  $p = 1$ )*

(a) *If  $x_n \geq c > 0$  for all  $n$  and  $x_n = o(\ln n)$ , then (D) and (52) hold.*

(b) *If  $x_n \geq c \ln n$  for all  $n \geq n_1$ , then condition (D) does not hold for  $\beta_0 < -1/c$ .*

**Proof:** (p. 357, [9]) **(a)**  $x_n/\ln n \rightarrow 0$  is equivalent to  $(x_n\beta)/\ln n \rightarrow 0$ ,  $\beta \in R$ .

This means that  $\forall \delta > 0$ , there exists  $N_\delta$  such that  $|x_n\beta| \leq \delta \ln n$ ,  $\forall n \geq N_\delta$ .

In particular,  $x_n\beta \geq -\delta \ln n$ ,  $\forall n \geq N_\delta$ . Hence  $\exp(x_n\beta) \geq n^{-\delta}$ ,  $\forall n \geq N_\delta$ .

To prove **(D)**, we note that

$$\begin{aligned} F_n &= \sum_{i=1}^n \exp(x_i\beta_0)x_i^2 = \sum_{i=1}^{N_\delta-1} \exp(x_i\beta_0)x_i^2 + \sum_{i=N_\delta}^n \exp(x_i\beta_0)x_i^2 \\ &\geq \sum_{i=N_\delta}^n \exp(x_i\beta_0)x_i^2 \geq \sum_{i=N_\delta}^n i^{-\delta}x_i^2 \geq c^2 \sum_{i=N_\delta}^n i^{-\delta} \geq c_1 n^{-\delta+1} \rightarrow \infty, \end{aligned} \quad (53)$$

by choosing  $\delta \in (0, 1)$ . Since  $n^{-\delta+1} \rightarrow \infty$ , if  $\delta \in (0, 1)$ , it follows that  $F_n \rightarrow \infty$  i.e.

**(D)** holds.

To prove (52), by  $x_n = o(\ln n)$  and (53), we get

$$\frac{x_n^2}{F_n} \leq \frac{(c_2 \ln n)^2}{c_1 n^{-\delta+1}} = \frac{c_2^2 (\ln n)^2}{c_1 n^{-\delta+1}} \rightarrow 0.$$

This concludes the proof of **(a)**.

**(b)** Note that if  $\beta < 0$ , then the function  $f(x) = \exp(x\beta)x^2$  is decreasing, if  $x$  is large enough. Since  $x_i \geq c \ln i$ ,  $\forall i \geq n_1$  we get

$$\exp(x_i\beta)x_i^2 = f(x_i) \leq f(c \ln i) = \exp[(c \ln i)\beta](c \ln i)^2 = c^2 i^{c\beta} (\ln i)^2 \leq c^2 i^{c\beta} c_1 i^\varepsilon.$$

Hence by choosing  $\varepsilon > 0$ , such that  $-c\beta - \varepsilon > 1$ , we get

$$F_n = \sum_{i=1}^{n_1-1} \exp(x_i\beta_0)x_i^2 + \sum_{i=n_1}^n \exp(x_i\beta_0)x_i^2 \leq c_0 + c^2 c_1 \sum_{i=n_1}^n i^{c\beta+\varepsilon} = c_0 + c^2 c_1 \sum_{i=n_1}^n \frac{1}{i^{-c\beta-\varepsilon}},$$

which converges to a finite limit. Note that the choice of  $\varepsilon \in (0, -c\beta - 1)$  is possible if  $-c\beta - 1 > 0$ , i.e.  $\beta < -1/c$ .

□

### 3.3.4 Logistic Regression for Binary Data

In this subsection we assume that the responses  $y_i$  assume only the values 0 and 1, say

$$P(y_i = 1) = p_i \text{ and } P(y_i = 0) = 1 - p_i.$$

Then the density function of  $y_i$  is

$$\begin{aligned} f(y_i|p_i) &= p_i^{y_i}(1-p_i)^{1-y_i} = \exp\{y_i \ln p_i + [(1-y_i) \ln(1-p_i)]\} = \\ &= \exp\left[y_i \ln \frac{p_i}{1-p_i} + \ln(1-p_i)\right] = \exp[y_i \text{logit}(p_i) + \ln(1-p_i)]. \end{aligned}$$

From here we see that the natural link function is:

$$\text{logit}(p_i) = \theta_i = x_i^T \beta.$$

Hence

$$E(y_i) = p_i = \frac{e^{\theta_i}}{1+e^{\theta_i}} := \mu(\theta_i), \quad \text{Var}(y_i) = p_i(1-p_i) = \frac{e^{\theta_i}}{(1+e^{\theta_i})^2} = \mu'(\theta_i) \text{ and } \mu(x) = \frac{e^x}{1+e^x}.$$

We have

$$F_n(\beta) = \sum_{i=1}^n \sigma_i^2(\beta) x_i x_i^T = \sum_{i=1}^n \frac{e^{x_i^T \beta}}{(1+e^{x_i^T \beta})^2} x_i x_i^T.$$

**Proposition 3.3.4** *If (D) holds and  $x_n^T F_n^{-1} x_n \rightarrow 0$ , then (N) holds.*

**Proof:** Note that

$$\begin{aligned} \left| \frac{\sigma_n^2(\beta)}{\sigma_n^2(\beta_0)} - 1 \right| &= \left| e^{x_n^T(\beta-\beta_0)} \left( \frac{1+e^{x_n^T \beta_0}}{1+e^{x_n^T \beta}} \right)^2 - 1 \right| = \left| e^{x_n^T(\beta-\beta_0)} \left( 1 + \frac{e^{x_n^T \beta_0} - e^{x_n^T \beta}}{1+e^{x_n^T \beta}} \right)^2 - 1 \right| \\ &= \left| e^{x_n^T(\beta-\beta_0)} \left[ 1 + \frac{e^{x_n^T \beta}}{1+e^{x_n^T \beta}} \left( e^{x_n^T(\beta_0-\beta)} - 1 \right) \right]^2 - 1 \right| \rightarrow 0, \end{aligned}$$

since  $|x_n^T(\beta-\beta_0)| \rightarrow 0$  and  $|e^{x_n^T \beta} / (1+e^{x_n^T \beta})| < 1$ . The remaining part of the proof is the same as the proof of Proposition 3.3.2.

□

### 3.4 Stochastic Regressors

In this section we extend the results of Section 3.2, to the case of stochastic regressors  $(x_i)_{i \geq 1}$ . More precisely, we suppose that  $(y_i, x_i^T)_{i \geq 1}$  is a sequence of random vectors

defined on the probability space  $(\Omega, \mathcal{F}, P_\beta)$ , which satisfies the following assumptions:

(A) The conditional density of  $y_i$  given  $(x_1^T, x_2^T, \dots, x_i^T, y_1, y_2, \dots, y_{i-1})$  is

$$f_\beta(y_i | x_1^T, x_2^T, \dots, x_i^T, y_1, y_2, \dots, y_{i-1}) = c(y_i) \exp \left\{ \frac{y_i x_i^T \beta - a(x_i^T \beta)}{\phi} \right\},$$

where  $\phi$  is a scale parameter.

(B) The conditional density of  $x_i^T$  given  $(x_1^T, y_1, \dots, x_{i-1}^T, y_{i-1})$  does not depend on  $\beta$ ; we denote this density by  $g_i(x_i^T | x_1^T, y_1, \dots, x_{i-1}^T, y_{i-1})$ .

Let  $\mathcal{F}_{i-1}$  be the  $\sigma$ -field generated by  $(x_1^T, x_2^T, \dots, x_i^T, y_1, y_2, \dots, y_{i-1})$ . Note that  $y_i$  is  $\mathcal{F}_i$ -measurable and  $x_i$  is  $\mathcal{F}_{i-1}$ -measurable for any  $i$ . Moreover

$$E_\beta(y_i | \mathcal{F}_{i-1}) =: \mu(x_i^T \beta), \quad \text{Var}_\beta(y_i | \mathcal{F}_{i-1}) = \phi \mu'(x_i^T \beta),$$

where  $\mu = a'$ .

We define the **errors**

$$\varepsilon_i(\beta) = y_i - \mu(x_i^T \beta), \quad \forall i \geq 1.$$

In what follows, we denote  $\mathbf{y}_n = (y_1, \dots, y_n)$ ,  $\mathbf{x}_n = (x_1^T, x_2^T, \dots, x_n^T)$  and  $f_\beta(x_i^T, y_i | \mathbf{x}_{i-1}, \mathbf{y}_{i-1})$  the conditional density of  $(x_i^T, y_i)$  given  $(\mathbf{x}_{i-1}, \mathbf{y}_{i-1})$ . The likelihood function of  $(\mathbf{x}_n, \mathbf{y}_n)$  under  $P_\beta$  is:

$$\begin{aligned} L_n(\beta | \mathbf{x}_n, \mathbf{y}_n) &= f_\beta(\mathbf{x}_n, \mathbf{y}_n) = f(x_1^T, y_1) \prod_{i=2}^n f_\beta(x_i^T, y_i | \mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \\ &= g_1(x_1^T) f(y_1 | x_1^T) \prod_{i=2}^n g_i(x_i^T | \mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \prod_{i=2}^n f_\beta(y_i | \mathbf{x}_{i-1}, \mathbf{y}_{i-1}, x_i^T) \\ &= g_1(x_1^T) \prod_{i=2}^n g_i(x_i^T | \mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \prod_{i=1}^n c(y_i) \exp \left\{ \frac{y_i x_i^T \beta - a(x_i^T \beta)}{\phi} \right\}. \end{aligned}$$

The log-likelihood function is

$$l_n(\beta) = \log L_n(\beta) = \log g_1(x_1^T) + \sum_{i=2}^n \log g_i(x_i^T | \mathbf{x}_{i-1}, \mathbf{y}_{i-1}) + \sum_{i=1}^n \log c(y_i) + \sum_{i=1}^n [y_i x_i^T \beta - a(x_i^T \beta)]$$

The score function is

$$s_n(\beta) = \frac{\partial l_n}{\partial \beta}(\beta) = \sum_{i=1}^n [y_i - \mu(x_i^T \beta)] x_i = \sum_{i=1}^n \varepsilon_i(\beta) x_i.$$

The derivative of the score function is

$$\frac{\partial s_n}{\partial \beta}(\beta) = - \sum_{i=1}^n \mu'(x_i^T \beta) x_i x_i^T := -F_n(\beta)$$

Note that  $F_n(\beta)$  is now a  $p \times p$  random matrix.

For sake of completeness we include the following lemma providing the necessary calculations.

**Lemma 3.4.1** *Under  $P_\beta$   $\{s_n(\beta)\}_{n \geq 1}$  is a zero-mean  $p$ -dimensional martingale with*

$$\text{Cov}_\beta[s_n(\beta)] := E_\beta[s_n(\beta)s_n^T(\beta)] = E_\beta[F_n(\beta)].$$

**Proof:** Note that  $\{\varepsilon_i(\beta)\}_{i \geq 1}$  is martingale difference under  $P_\beta$ :

$$E_\beta(\varepsilon_i(\beta)|\mathcal{F}_{i-1}) = E_\beta[y_i - E_\beta(y_i|\mathcal{F}_{i-1})|\mathcal{F}_{i-1}] = E_\beta(y_i|\mathcal{F}_{i-1}) - E_\beta(y_i|\mathcal{F}_{i-1}) = 0.$$

We have

$$E_\beta[s_n(\beta)|\mathcal{F}_{n-1}] = E_\beta[s_{n-1}(\beta)|\mathcal{F}_{n-1}] + x_n E_\beta[\varepsilon_n(\beta)|\mathcal{F}_{n-1}] = s_{n-1}(\beta),$$

i.e.  $(s_n(\beta))_{n \geq 1}$  is a martingale.

$$\begin{aligned} \text{To prove that } E_\beta[s_n(\beta)] &= 0, \text{ we proceed as follows: } E_\beta[s_n(\beta)] = E_\beta\left[\sum_{i=1}^n \varepsilon_i(\beta)x_i\right] \\ &= \sum_{i=1}^n E_\beta\{E_\beta[\varepsilon_i(\beta)x_i|\mathcal{F}_{i-1}]\} = \sum_{i=1}^n E_\beta\{x_i E_\beta[\varepsilon_i(\beta)|\mathcal{F}_{i-1}]\} = 0. \end{aligned}$$

To prove that  $\text{Cov}_\beta[s_n(\beta)] = E_\beta[F_n(\beta)]$ , we write

$$\begin{aligned} E_\beta[s_n(\beta)s_n^T(\beta)] &= E_\beta\left\{\left[\sum_{i=1}^n x_i \varepsilon_i(\beta)\right] \left[\sum_{i=1}^n x_i \varepsilon_i(\beta)\right]^T\right\} = E_\beta\left[\sum_{i,j=1}^n x_i \varepsilon_i(\beta) \varepsilon_j(\beta) x_j^T\right] \\ &= E_\beta\left[\sum_{i=1}^n x_i \varepsilon_i^2(\beta) x_i^T\right] + 2E_\beta\left[\sum_{i=1}^n \sum_{j=i+1}^n x_i \varepsilon_i(\beta) \varepsilon_j(\beta) x_j^T\right] \\ &= \sum_{i=1}^n E_\beta\{x_i E_\beta[\varepsilon_i^2(\beta)|\mathcal{F}_{i-1}] x_i^T\} + 2 \sum_{i=1}^n \sum_{j=i+1}^n E_\beta\{x_i \varepsilon_i(\beta) E_\beta[\varepsilon_j(\beta)|\mathcal{F}_{j-1}] x_j^T\} \\ &= E_\beta\left[\sum_{i=1}^n x_i \mu'(x_i^T \beta) x_i^T\right] = E_\beta[F_n(\beta)]. \quad \square \end{aligned}$$

### 3.4.1 Asymptotic Existence and Weak Consistency

We define

$$V_n^{(s)}(\beta) = (EF_n)^{-1/2} F_n(\beta) (EF_n)^{-T/2}.$$

We consider the following conditions:

(**D<sub>s</sub>**)  $\lambda_{\min}(EF_n) \longrightarrow \infty$ .

(**C<sub>s</sub>**) There exists  $c > 0$  such that for any  $r > 0$  and for any  $\varepsilon > 0$ , there exists  $n_1 = n_1(r, \varepsilon)$  with

$$P(F_n(\beta) \geq c(EF_n), \forall \beta \in N_n^{(s)}(r)) \geq 1 - \varepsilon, \forall n \geq n_1,$$

where  $N_n^{(s)}(r) = \{\beta : \|(EF_n)^{T/2}(\beta - \beta_0)\| < r\}$ .

**Theorem 3.4.2** *Under (**D<sub>s</sub>**) and (**C<sub>s</sub>**), there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that*

(**a**)  $P(s_n(\hat{\beta}_n) = 0) \longrightarrow 1$

(**b**)  $\hat{\beta}_n \longrightarrow \beta_0$  in probability.

**Proof:** The proof is similar to the proof of Theorem 3.2.1. The details are quite different though, because now  $F_n$  is a random matrix and we work with  $E(F_n)$  instead of  $F_n$ .

(**a**) Let  $c > 0$  be the constant given by condition (**C<sub>s</sub>**). For every  $r > 0$ ,  $n \geq 1$ , we consider the following events:

$$\Omega_n^{(s)}(r) = \{F_n(\beta) \geq c(EF_n), \forall \beta \in N_n^{(s)}(r),$$

$$E_n^{(s)}(r) = \{l_n(\beta) - l_n(\beta_0) < 0, \forall \beta \in \partial N_n^{(s)}(r)\}.$$

Note that on the event  $\Omega_n^{(s)}$ , the function  $l_n(\cdot)$  is strictly concave. Hence on the event  $E_n^{(s)}(r) \cap \Omega_n^{(s)}$ , the local maximum of  $l_n(\cdot)$  exists and we denote it by  $\hat{\beta}_n$ , that is

$$E_n^{(s)}(r) \cap \Omega_n^{(s)}(r) \subseteq \{\text{there exist } \hat{\beta} \in N_n^{(s)}(r), \text{ such that } s_n(\hat{\beta}_n) = 0\}.$$

To prove part (**a**) of the theorem, we will show that for any  $\varepsilon > 0$ , there exists  $r = r_\varepsilon$  such that

$$P(E_n^{(s)}(r) \cap \Omega_n^{(s)}(r)) \geq 1 - \varepsilon, \forall n \geq n_1, \quad (54)$$

where  $n_1 = n_1(r)$  is given by condition  $(\mathbf{C}_s)$ .

To prove (54), let  $\varepsilon > 0$  be arbitrary and  $r = r_\varepsilon$  be a constant to be specified later. Let  $\beta \in \partial N_n^{(s)}(r)$  be arbitrary. Using Taylor's expansion of  $l_n(\cdot)$  around  $\beta_0$  (as in the proof of Theorem 3.2.1), we get

$$\begin{aligned} l_n(\beta) - l_n(\beta_0) &= (\beta - \beta_0)^T s_n - \frac{1}{2}(\beta - \beta_0)^T F_n(\tilde{\beta}_n)(\beta - \beta_0) \\ &= \frac{(\beta - \beta_0)^T (EF_n)^{1/2}}{r} r (EF_n)^{-1/2} s_n \\ &\quad - \frac{1}{2} \frac{(\beta - \beta_0)^T (EF_n)^{1/2}}{r} r^2 (EF_n)^{-1/2} F_n(\tilde{\beta}_n) (EF_n)^{-T/2} \frac{(EF_n)^{T/2} (\beta - \beta_0)}{r}, \end{aligned}$$

where  $\tilde{\beta}_n$  lies between  $\beta$  and  $\beta_0$  and  $\beta \in \partial N_n^{(s)}(r)$ . Denote  $\lambda = (1/r)(EF_n)^{T/2}(\beta - \beta_0)$  and note that  $\|\lambda\| = 1$ , because  $\beta \in \partial N_n^{(s)}(r)$ . Hence

$$l_n(\beta) - l_n(\beta_0) = r\lambda^T (EF_n)^{-1/2} s_n - \frac{1}{2} r^2 \lambda^T V_n^{(s)}(\tilde{\beta}_n) \lambda, \quad \forall \beta \in \partial N_n^{(s)}(r). \quad (55)$$

For the first term,

$$r\lambda^T (EF_n)^{-1/2} s_n \leq r \|(EF_n)^{-1/2} s_n\|. \quad (56)$$

For the second term, on the event  $\Omega_n^{(s)}(r)$ ,

$$V_n^{(s)}(\tilde{\beta}_n) = (EF_n)^{-1/2} F_n(\tilde{\beta}_n) (EF_n)^{-T/2} \geq c(EF_n)^{-1/2} (EF_n) (EF_n)^{-T/2} = cI.$$

Hence:

$$\frac{1}{2} r^2 \lambda^T V_n^{(s)}(\tilde{\beta}_n) \lambda \geq \frac{1}{2} r^2 \lambda_{\min}(V_n^{(s)}(\tilde{\beta}_n)) \lambda^T \lambda \geq \frac{1}{2} r^2 c. \quad (57)$$

Using (55), (56) and (57), we get that on the event  $\Omega_n^{(s)}(r)$ ,

$$l_n(\beta) - l_n(\beta_0) \leq r \|(EF_n)^{-1/2} s_n\| - (1/2) r^2 c.$$

Hence

$$\begin{aligned} P(E_n^{(s)}(r) \cap \Omega_n^{(s)}(r)) &\geq P(\{\|(EF_n)^{-1/2} s_n\| < (1/2)rc\} \cap \Omega_n^{(s)}(r)) \geq \\ &\geq P(\|(EF_n)^{-1/2} s_n\| < (1/2)rc) + P(\Omega_n^{(s)}(r)) - 1, \end{aligned} \quad (58)$$

where we used Bonferroni's inequality. By Chebyshev's inequality, we have

$$P(\|(EF_n)^{-1/2} s_n\| < (1/2)rc) \geq 1 - \frac{1}{[(1/2)rc]^2} E\|(EF_n)^{-1/2} s_n\|^2$$

$$= 1 - \frac{4}{r^2 c^2} \text{tr}\{(EF_n)^{-1/2} E(s_n s_n^T) (EF_n)^{-T/2}\} = 1 - \frac{4}{r^2 c^2} \text{tr}(I) = 1 - \frac{4}{r^2 c^2} p = 1 - \frac{\varepsilon}{2}, \quad (59)$$

by choosing  $r = \sqrt{8p/(c^2\varepsilon)}$ .

From condition  $(\mathbf{C}_s)$ , there exists  $n_1 = n_1(r_\varepsilon, \varepsilon)$ , such that

$$P(\Omega_n^{(s)}(r)) > 1 - \frac{\varepsilon}{2}, \quad \forall n \geq n_1. \quad (60)$$

From (58), (59) and (60), we get

$$P(E_n^{(s)}(r) \cap \Omega_n^{(s)}(r)) \geq (1 - \frac{\varepsilon}{2}) + (1 - \frac{\varepsilon}{2}) - 1 = 1 - \varepsilon, \quad \forall n \geq n_1.$$

This concludes the proof of (54).

**(b)** Let  $\eta, \varepsilon > 0$  be arbitrary. We need to prove that there exists  $N_{\eta, \varepsilon} = N$  such that

$$P(\|\hat{\beta}_n - \beta_0\| \leq \eta) \geq 1 - \varepsilon \quad \forall n \geq N. \quad (61)$$

In part **(a)**, we proved the existence of  $\hat{\beta}_n$  on the event  $E_n^{(s)}(r) \cap \Omega_n^{(s)}(r)$ , where  $r = \sqrt{8p/(c^2\varepsilon)}$ . Moreover,  $\hat{\beta}_n \in N_n^{(s)}(r)$ . By  $(\mathbf{D}_s)$ , there exists  $N = N_\eta > n_1$  such that  $\lambda_{\min}(EF_n) > (r/\eta)^2, \forall n \geq N$ . Since  $\hat{\beta} \in N_n^{(s)}(r)$ , on the event  $E_n^{(s)}(r) \cap \Omega_n^{(s)}(r)$ , we have

$$\|\hat{\beta}_n - \beta_0\| \leq \|(EF_n)^{-1/2}\| \cdot \|(EF_n)^{1/2}(\hat{\beta} - \beta_0)\| \leq \frac{1}{[\lambda_{\min}(EF_n)]^{1/2}} r \leq \eta, \quad \forall n \geq N$$

and therefore, by (54)

$$1 - \varepsilon \leq P(E_n^{(s)}(r) \cap \Omega_n^{(s)}(r)) \leq P(\|\hat{\beta}_n - \beta_0\| \leq \eta), \quad \forall n \geq N.$$

This concludes the proof of (61).

□

### 3.4.2 Asymptotic Normality

We consider the condition:

$(\mathbf{N}_s)$  For all  $r > 0$ , we have

$$\sup_{\beta \in N_n^{(s)}(r)} \|V_n^{(s)}(\beta) - I\| \longrightarrow 0, \quad \text{in probability.}$$

**Lemma 3.4.3**  $(\mathbf{N}_s)$  implies  $(\mathbf{C}_s)$ .

**Proof:**  $(\mathbf{N}_s)$  is equivalent to the following assertion: for any  $r > 0$ ,  $\varepsilon > 0$  and  $\eta > 0$ , there exists  $n_1 = n_1(r, \varepsilon, \eta)$  such that

$$P(\|(EF_n)^{-1/2}F_n(\beta)(EF_n)^{-T/2} - I\| \leq \eta, \forall \lambda \in R^p, \forall \beta \in N_n^{(s)}(r)) \geq 1 - \varepsilon, \forall n \geq n_1.$$

By taking  $x = (EF_n)^{-T/2}\lambda$ , the previous relation can be written as:

$$(\mathbf{N}_s'') P\left(|x^T F_n(\beta)x - x^T (EF_n)x| \leq \eta x^T (EF_n)x, \forall x \in R^p, \forall \beta \in N_n^{(s)}(r)\right) \geq 1 - \varepsilon, \forall n \geq n_1.$$

In particular, this implies

$$P(x^T F_n(\beta)x \geq (1 - \eta)x^T (EF_n)x, \forall x \in R^p, \forall \beta \in N_n^{(s)}(r)) \geq 1 - \varepsilon, \forall n \geq n_1,$$

i.e.

$$P(F_n(\beta) \geq (1 - \eta)(EF_n), \forall \beta \in N_n^{(s)}(r)) \geq 1 - \varepsilon, \forall n \geq n_1.$$

By fixing  $\eta \in (0, 1)$  and letting  $c = 1 - \eta$ , we obtain exactly  $(\mathbf{C}_s)$ .

□

**Lemma 3.4.4** Under  $(\mathbf{D}_s)$  and  $(\mathbf{N}_s)$ ,

$$(EF_n)^{-1/2}s_n \longrightarrow N(0, I) \text{ in distribution.}$$

**Proof:** The proof is similar to the proof of Lemma 3.2.2. By the Cràmer-Wold theorem (Theorem A.2.4, Appendix A.2) it suffices to show that for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ , we have

$$\lambda^T (EF_n)^{-1/2}s_n \longrightarrow N(0, 1) \text{ in distribution.} \quad (62)$$

Let  $\lambda \in R^p$  with  $\|\lambda\| = 1$  be arbitrary. By the continuity theorem of the moment generating function (Theorem A.2.6, Appendix A.2), (62) is equivalent to:

$$E[\exp(r\lambda^T (EF_n)^{-1/2}s_n)] \longrightarrow \exp(r^2/2), \forall r \in R. \quad (63)$$

Let  $r > 0$  be arbitrary. Let  $\beta_n = \beta_0 + r(EF_n)^{-T/2}\lambda$  and note that using (55), there exists  $\tilde{\beta}_n$  between  $\beta_n$  and  $\beta_0$  such that

$$l_n(\beta_n) = l_n(\beta_0) + r\lambda^T (EF_n)^{-1/2}s_n - \frac{1}{2}r^2\lambda^T V_n^{(s)}(\tilde{\beta}_n)\lambda,$$

i.e.

$$\frac{1}{2}r^2\lambda^T V_n^{(s)}(\tilde{\beta}_n)\lambda + l_n(\tilde{\beta}_n) = r\lambda^T (EF_n)^{-1/2}s_n + l_n(\beta_0). \quad (64)$$

Taking the exponential for both sides of (64), we get:

$$\exp[(r^2/2)\lambda^T V_n^{(s)}(\tilde{\beta}_n)\lambda]L_n(\beta_n) = \exp(r\lambda^T (EF_n)^{-1/2}s_n)L_n(\beta_0). \quad (65)$$

Recall that

$$L_n(\beta) \text{ is the density function of } (x_1, y_1, \dots, x_n, y_n) \text{ under } P_\beta. \quad (66)$$

Integrating both sides of (65) with respect to  $dx_1 dy_1 \dots dx_n dy_n$ , and using (66) we get

$$E_{\beta_n} \left[ \exp[(r^2/2)\lambda^T V_n^{(s)}(\tilde{\beta}_n)\lambda] \right] = E[\exp(r\lambda^T (EF_n)^{-1/2}s_n)]. \quad (67)$$

Therefore, in order to prove (63), it suffices to show that

$$E_{\beta_n} \left[ \exp \left( \frac{1}{2}r^2\lambda^T V_n^{(s)}(\tilde{\beta}_n)\lambda \right) \right] \longrightarrow \exp(r^2/2). \quad (68)$$

The proof of (68) relies on condition  $(\mathbf{N}'_s)$ , and is similar to the proof of relation (39).

Details are omitted.

□

**Theorem 3.4.5** Under  $(\mathbf{D}_s)$  and  $(\mathbf{N}_s)$ ,

$$(EF_n)^{T/2}(\hat{\beta}_n - \beta_0) \longrightarrow N(0, I) \text{ in distribution.}$$

**Proof:** The proof is similar to the Theorem 3.2.4. We focus on the event  $\{s_n(\hat{\beta}_n) = 0\}$  whose probability converges to 1, by Theorem 3.4.1. Using the Taylor's expansion for the function  $(EF_n)^{-1/2}s_n(\cdot)$  around  $\beta_0$ , we get

$$\begin{aligned} (EF_n)^{-1/2}s_n &= -(EF_n)^{-1/2} \left( s_n(\hat{\beta}_n) - s_n \right) = \\ (EF_n)^{-1/2}F_n(\tilde{\beta}_n)(\hat{\beta}_n - \beta_0) &= V_n(\tilde{\beta}_n)(EF_n)^{T/2}(\hat{\beta}_n - \beta_0), \end{aligned} \quad (69)$$

where  $\tilde{\beta}_n$  is a random vector which lies between  $\hat{\beta}_n$  and  $\beta_0$ . Note that

$$E \left( \|(EF_n)^{-1/2}s_n\|^2 \right) = \text{tr} \left[ (EF_n)^{-1/2}E(s_n s_n^T)(EF_n)^{-T/2} \right] = \text{tr}(I) = p, \quad \forall n \geq 1. \quad (70)$$

We claim that

$$\|V_n(\tilde{\beta}_n) - I\| \longrightarrow 0 \text{ in probability.} \quad (71)$$

Using Lemma A.2.7 (Appendix A.2), (69), (70) and (71) yield

$$(EF_n)^{T/2}(\hat{\beta}_n - \beta_0) = (EF_n)^{-1/2}s_n + o_p(1).$$

The conclusion follows using Lemma 3.4.4 and Slutsky's theorem. The proof of (71) is similar to the proof of (44), and therefore is omitted.

□

### 3.4.3 Asymptotic Existence and Strong Consistency

We consider the following condition:

$(\mathbf{S}_\delta^{(s)})$  There exist constants  $c > 0$ ,  $\delta > 0$  and a neighborhood  $N$  of  $\beta_0$  such that with probability 1, there exists a random number  $n_1$  such that

$$\lambda_{\min}[F_n(\beta)] \geq c[\lambda_{\max}(EF_n)]^{(1/2)+\delta}, \quad \forall \beta \in N, \quad \forall n \geq n_1.$$

**Lemma 3.4.6** *Under  $(\mathbf{D}_s)$ , for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ , we have*

$$\frac{\lambda^T s_n}{[\lambda_{\max}(EF_n)]^{1/2+\delta}} \longrightarrow 0 \text{ a.s.}$$

**Proof:** By Lemma 3.4.1,  $\{\lambda^T s_n\}_{n \geq 1}$  is a zero-mean 1-dimensional martingale with

$$E[(\lambda^T s_n)^2] = \lambda^T (EF_n) \lambda \leq \lambda_{\max}(EF_n) \lambda^T \lambda = \lambda_{\max}(EF_n).$$

The conclusion follows by Theorem A.2.2 (ii) (Appendix A.2).

□

**Theorem 3.4.7** *Under  $(\mathbf{D}_s)$  and  $(\mathbf{S}_\delta^{(s)})$ , there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables and a random number  $n_2$  with*

(a)  $P\left(s_n(\hat{\beta}_n) = 0, \text{ for all } n \geq n_2\right) = 1.$

(b)  $\hat{\beta}_n \longrightarrow \beta_0 \text{ a.s.}$

**Proof:** The proof is similar to Theorem 3.2.6. Choose  $\varepsilon_0 > 0$  such that the neighborhood  $B_{\varepsilon_0}(\beta_0) = \{\beta : \|\beta - \beta_0\| \leq \varepsilon_0\} \subseteq N$ . We will show that with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$ , there exists a random number  $n_2 = n_2(\varepsilon)$  such that

$$l_n(\beta) - l_n(\beta_0) < 0, \quad \forall \beta \in \partial B_\varepsilon(\beta_0), \quad \forall n \geq n_2. \quad (72)$$

By Theorem A.3.1 (Appendix A.3), this will imply the conclusion of our theorem in the following format: with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$ , there exists  $n_2 = n_2(\varepsilon)$  (random) such that for all  $n \geq n_2$ , there exists  $\hat{\beta}_n \in B_\varepsilon(\beta_0)$  with  $s_n(\hat{\beta}_n) = 0$ .

To prove (72), we use the Taylor's expansion of  $l_n(\cdot)$  around  $\beta_0$ : for  $\beta \in \partial B_\varepsilon(\beta_0)$ ,  $\varepsilon \in (0, \varepsilon_0)$ , we have

$$\begin{aligned} l_n(\beta) - l_n(\beta_0) &= (\beta - \beta_0)^T s_n - \frac{1}{2}(\beta - \beta_0)^T F_n(\tilde{\beta}_n)(\beta - \beta_0) \\ &= \frac{(\beta - \beta_0)^T}{\varepsilon} \varepsilon s_n - \frac{1}{2} \frac{(\beta - \beta_0)^T}{\varepsilon} F_n(\tilde{\beta}_n) \varepsilon^2 \frac{\beta - \beta_0}{\varepsilon} = \varepsilon \lambda^T s_n - \frac{1}{2} \varepsilon^2 \lambda^T F_n(\tilde{\beta}_n) \lambda, \end{aligned} \quad (73)$$

where  $\tilde{\beta}_n$  lies between  $\beta$  and  $\beta_0$  and  $\lambda = (\beta - \beta_0)/\varepsilon$ . Note that  $\lambda^T \lambda = 1$  since  $\beta \in \partial B_\varepsilon(\beta_0)$ . By condition  $(\mathbf{S}_\delta^{(s)})$ , we get that with probability 1, there exists a random number  $n_1$  such that for any  $\varepsilon \in (0, \varepsilon_0)$ ,  $\beta \in B_\varepsilon(\beta_0)$ , we have

$$\lambda^T F_n(\tilde{\beta}) \lambda \geq [\lambda_{\max}(EF_n)]^{(1/2)+\delta}, \quad \forall n \geq n_1. \quad (74)$$

From Lemma 3.4.6, it follows that with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$  there exists a random number  $n_2 = n_2(\varepsilon) > n_1$  such that

$$\lambda^T s_n \leq (\varepsilon/2)c[(\lambda_{\max}E(F_n))^{(1/2)+\delta}], \quad \forall n > n_2. \quad (75)$$

From (73), (74) and (75), we obtain (72) holds.

□

# Chapter 4

## Longitudinal Data

In biostatistics and life-time testing problems, longitudinal data sets comprise a series of repeated measurements of a response variable, together with a set of covariates, on each subject observed chronologically over time. Assuming that the mean of the response variables can be specified in term of some regression parameter  $\beta$ , Liang and Zeger proposed an approach based on a generalized linear model (GLM) for each marginal response. This approach has been fruitfully exploited recently by Xie and Yang, who proved that most of the asymptotic results which are valid in the GLM case continue to hold in the longitudinal case. The estimator  $\hat{\beta}_n$  of  $\beta$  is defined as a root of the so-called **generalized estimating equation** (GEE), which reduces to a maximum likelihood equation if there is only one observation performed on each individual, and this observation has a density function which belongs to an exponential family.

This chapter is organized as follows. In Section 4.1, we introduce the framework, the (rather complicated) matrix notation and the construction of the GEE. In Section 4.2, we develop the asymptotic theory (in parallel with the theory that was presented in Section 3.2). The major difference between the two theories is the use of an injection lemma for the existence of solution of the GEE, in the longitudinal case, instead of the typical argument based on the concavity of the likelihood function, in the GLM case.

## 4.1 The Framework

For each  $i \geq 1$ , let  $y_i = (y_{i1}, \dots, y_{im})^T$  be a vector of  $m$  response variables recorded on the same individual at  $m$  different moments of time. Suppose that each  $y_{ij}$  variable is recorded together with an explanatory variable  $x_{ij}$ . We assume that each  $y_{ij}$  is a random variable defined on probability space  $(\Omega, \mathcal{F}, P_\beta)$  and each  $x_{ij}$  is a (non-random)  $p$ -dimensional vector such that: for any  $i \geq 1$  and for any  $j = 1, \dots, m$

$$\mu_{ij}(\beta) = E_\beta(y_{ij}) = \mu(x_{ij}^T \beta) \text{ and } \sigma_{ij}^2(\beta) = Var_\beta(y_{ij}) = \phi \mu'(x_{ij}^T \beta). \quad (76)$$

where  $E_\beta$  denotes the expectation with respect to  $P_\beta$ ,  $\mu(\cdot)$  is a differentiable function,  $\phi$  is a nuisance scale parameter (which we will normally take as  $\phi = 1$ ) and  $\beta$  is an unknown  $p$ -dimensional parameter which has to be estimated. We define the **errors**

$$\varepsilon_{ij}(\beta) = y_{ij} - \mu(x_{ij}^T \beta) \text{ and } \varepsilon_i(\beta) = (\varepsilon_{i1}(\beta), \dots, \varepsilon_{im}(\beta)), \quad \forall i \geq 1,$$

and we assume that the random vectors  $(\varepsilon_i)_{i \geq 1}$  are independent (under  $P$ ).

As in the GLM case, condition (76) is satisfied if we suppose that each  $y_{ij}$  has a density function which belongs to the following exponential family

$$f(y_{ij} | \theta_{ij}) = c(y_{ij}) \exp \left\{ \frac{y_{ij} \theta_{ij} - a(\theta_{ij})}{\phi} \right\}, \quad (77)$$

where  $c(\cdot)$  is a positive function and  $\theta_{ij}$  is a 1-dimensional parameter. Letting  $\mu = a'$ , we conclude that

$$E(y_{ij}) = \mu(\theta_{ij}) \text{ and } Var(y_{ij}) = \phi \mu''(\theta_{ij}).$$

Moreover, we suppose that there exists an injective function  $g$  (called the **link function**) such that

$$x_{ij}^T \beta = g(\mu(\theta_{ij})), \quad \forall i \geq 1, \quad \forall j = 1, \dots, m$$

or equivalently  $\theta_{ij} = u(x_{ij}^T \beta)$ , where  $u = (g \circ \mu)^{-1}$ .

In the present work, we consider only the case of *natural* link functions i.e.  $\mu = g^{-1}$ . Hence  $\theta_{ij} = x_{ij}^T \beta$  and (76) holds. Moreover, in what follows, we will suppose that  $\phi = 1$ . We denote

$$\mu_i(\beta) = E_\beta(y_i) = (\mu_{i1}(\beta), \dots, \mu_{im}(\beta))^T, \quad \Sigma_i(\beta) = Cov_\beta(y_i),$$

$$A_i(\beta) = \text{diag}(\sigma_{i1}^2(\beta), \dots, \sigma_{im}^2(\beta)).$$

We work under the assumption that the true correlation matrix  $\bar{R}_i(\beta)$  does not depend on  $\beta$ , i.e.  $\bar{R}_i(\beta) = \bar{R}_i$  for all  $\beta$ . Let  $\bar{r}_{i,j,k}$  be the  $(j, k)$  element of the matrix  $\bar{R}_i$ . If  $\bar{r}_{i,j,k} = 0$  then  $y_{ij}$  and  $y_{ik}$  are independent. If  $|\bar{r}_{i,j,k}| = 1$  then  $y_{ij}$  and  $y_{ik}$  are linearly dependent.

For sake of completeness we include the following lemmas providing the necessary calculations.

**Lemma 4.1.1** *We have  $\Sigma_i(\beta) = A_i(\beta)^{1/2} \bar{R}_i A_i(\beta)^{1/2}$ .*

**Proof:** The  $(j, k)$ -element of the matrix  $\Sigma_i(\beta)$  is

$$\text{Cov}_\beta(y_{ij}, y_{ik}) = \sqrt{\text{Var}_\beta(y_{ij})} \text{Corr}(y_{ij}, y_{ik}) \sqrt{\text{Var}_\beta(y_{ik})} = \sigma_{ij}(\beta) \bar{r}_{i,jk} \sigma_{ik}(\beta),$$

which is exactly the  $(j, k)$ -element of matrix  $A_i(\beta)^{1/2} \bar{R}_i A_i(\beta)^{1/2}$ .

□

### *Construction of the generalized estimating equation (GEE)*

The main idea behind the development of the GEE comes from examining the case  $m = 1$ . Then, as we have seen in Chapter 3 the likelihood equation is

$$s_n(\beta) = \sum_{i=1}^n x_i (y_i - \mu(x_i^T \beta)) = 0. \quad (78)$$

Denote  $\mu_i(\beta) = E_\beta(y_i) = \mu(x_i^T \beta)$  and  $\sigma_i^2(\beta) = \text{Var}_\beta(y_i) = \mu'(x_i^T \beta)$ , since we supposed that  $\phi = 1$ . Let us denote

$$D_i(\beta) := \frac{\partial}{\partial \beta} \mu_i(\beta), \quad i = 1, \dots, n,$$

and note that

$$D_i(\beta) = x_i^T \mu'(x_i^T \beta) = \sigma_i^2(\beta) x_i^T. \quad (79)$$

Hence  $x_i = D_i(\beta)^T (\sigma_i^2(\beta))^{-1}$  and the likelihood equation (78) can be written as

$$\sum_{i=1}^n D_i(\beta)^T (\sigma_i^2(\beta))^{-1} (y_i - \mu_i(\beta)) = 0. \quad (80)$$

This equation is called the **quasi-likelihood equation** and its generalization to the case  $m > 1$ , leads us to the GEE.

More precisely in the case of an arbitrary  $m > 1$ , let  $D_i(\beta) = \frac{\partial}{\partial \beta} \mu_i(\beta)$ . In this case,  $D_i(\beta)$  is an  $m \times p$  matrix whose exact formula is similar to (79), and is given by the lemma below.

**Lemma 4.1.2** We have  $D_i(\beta) = A_i(\beta)X_i$ , where  $X_i = \begin{bmatrix} x_{i1}^T \\ \vdots \\ x_{im}^T \end{bmatrix}$ .

**Proof:** Note that the  $j$ -th row of  $D_i(\beta)$  is

$$\frac{\partial}{\partial \beta} \mu_{ij}(\beta) = \frac{\partial}{\partial \beta} \mu(x_{ij}^T \beta) = x_{ij}^T \mu'(x_{ij}^T \beta) = \sigma_{ij}^2(\beta) x_{ij}^T,$$

which exactly the  $j$ -th row of  $A_i(\beta)X_i$ .

□

The main difficulty in dealing with a longitudinal data set comes from the fact that the time correlation matrix  $\bar{R}_i$  is generally unknown. The normal approach which was suggested by Liang and Zeger, is to replace  $\bar{R}_i$  by another correlation matrix  $R_i(\alpha)$  called the “working” correlation matrix, whose form may depend on an known parameter  $\alpha$ . Typical examples of working correlation matrices are (p. 59-73, [10]):

(1) Exchangeable correlation

$$(r_{i,jk}) = \begin{cases} 1, & \text{if } j = k \\ \alpha, & \text{otherwise} \end{cases}$$

(2) Autoregressive correlation  $(r_{i,jk}) = \alpha^{|j-k|}$

(3) Stationary correlation

$$(r_{i,jk}) = \begin{cases} \alpha_{|j-k|}, & \text{if } |j-k| \leq l \\ 0, & \text{otherwise} \end{cases}$$

(4) Non-stationary correlation

$$(r_{i,jk}) = \begin{cases} 1, & \text{if } j = k \\ \alpha_{jk}, & \text{if } 0 < |j-k| \leq l \\ 0, & \text{otherwise} \end{cases},$$

where  $R_i(\alpha) := (r_{i,jk})_{jk} = (\text{corr}(y_{ij}, y_{ik}))_{jk}$ . By analogy with Lemma 4.1.1, we denote

$$V_i(\beta, \alpha) := A_i(\beta)^{1/2} R_i(\alpha) A_i(\beta)^{1/2}. \quad (81)$$

The  $m$ -dimensional analogue of equation (80), in which  $\Sigma_i(\beta)$  is replaced by  $V_i(\beta, \alpha)$  is:

$$g_n(\beta) := \sum_{i=1}^n D_i(\beta)^T V_i(\beta, \alpha)^{-1} (y_i - \mu_i(\beta)) = 0. \quad (82)$$

Equation (82) is called the **generalized estimating equation (GEE)**. Its solution  $\hat{\beta}_n$  (if it exists) is called the **GEE estimator**. Note that  $g_n(\beta)$  is not a score function, and therefore the existence of  $\hat{\beta}_n$  can not be examined using likelihood methods.

**Lemma 4.1.3** *We have  $g_n(\beta) = \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1} A_i(\beta)^{-1/2} \varepsilon_i(\beta)$ .*

**Proof:** Using Lemma 4.1.2 and (81) we have  $g_n(\beta) = \sum_{i=1}^n D_i(\beta)^T V_i(\beta, \alpha)^{-1} \varepsilon_i(\beta) = \sum_{i=1}^n X_i^T A_i(\beta) A_i(\beta)^{-1/2} R_i(\alpha)^{-1} A_i(\beta)^{-1/2} \varepsilon_i(\beta) = \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1} A_i(\beta)^{-1/2} \varepsilon_i(\beta)$ .  $\square$

**Lemma 4.1.4** *Let  $M_n := \text{Cov}(g_n)$ . Then*

$$M_n = \sum_{i=1}^n X_i^T A_i^{1/2} R_i(\alpha)^{-1} \bar{R}_i R_i(\alpha)^{-1} A_i^{1/2} X_i.$$

**Proof:** By Lemma 4.1.3 and Lemma 4.1.1 and using the independence of  $(\varepsilon_i)_{i \geq 1}$ , we have

$$\begin{aligned} \text{Cov}(g_n) &= E(g_n g_n^T) \\ &= E \left[ \left( \sum_{i=1}^n X_i^T A_i^{1/2} R_i(\alpha)^{-1} A_i^{-1/2} \varepsilon_i \right) \left( \sum_{j=1}^n \varepsilon_j^T A_j^{-1/2} R_j(\alpha)^{-1} A_j^{1/2} X_j \right) \right] \\ &= \sum_{i=1}^n X_i^T A_i^{1/2} R_i(\alpha)^{-1} A_i^{-1/2} \cdot E(\varepsilon_i \varepsilon_i^T) \cdot A_i^{-1/2} R_i^{-1}(\alpha) A_i^{1/2} X_i \\ &= \sum_{i=1}^n X_i^T A_i^{1/2} R_i(\alpha)^{-1} A_i^{-1/2} A_i^{1/2} \bar{R}_i A_i^{1/2} A_i^{-1/2} R_i(\alpha)^{-1} A_i^{1/2} X_i \\ &= \sum_{i=1}^n X_i^T A_i^{1/2} R_i(\alpha)^{-1} \bar{R}_i R_i(\alpha)^{-1} A_i^{1/2} X_i. \end{aligned}$$

□

The following quantities will play an important role in our theory:

$$\mathcal{D}_n(\beta) = -\frac{\partial}{\partial \beta} g_n(\beta), \quad H_n(\beta) = \sum_{i=1}^n D_i(\beta)^T V_i(\beta, \alpha)^{-1} D_i(\beta),$$

$$F_n = H_n M_n^{-1} H_n, \quad \tau_n = \max_{1 \leq i \leq n} \{ \lambda_{\max}(R_i(\alpha)^{-1} \bar{R}_i) \}.$$

Note that by Corollary A.1.19 (Appendix A.1)

$$\tau_n = \max_{1 \leq i \leq n} \{ \lambda_{\max}(R_i(\alpha)^{-1/2} \bar{R}_i R_i(\alpha)^{-1/2}) \}.$$

Let  $\Theta = \{ \theta \in R \mid 0 < \int c(y) \exp(\theta y) dy < \infty \}$  be the natural parameter space and  $\Theta^0$  be the interior of  $\Theta$ . Let  $M$  denote the image  $\mu(\Theta^0)$  of  $\Theta^0$ . We impose the following regularity assumptions:

- (1)  $\beta$  lies in an open set  $B \subseteq R^p$ .
- (2)  $x_i^T \beta \in g(M)$  for all  $\beta \in B$  and for all  $i \geq 1, j = 1, \dots, m$ .
- (3)  $\mu$  is three times continuously differentiable and  $\mu'(\theta) > 0$  for all  $\theta \in \Theta^0$ .
- (4)  $H_n$  and  $M_n$  are positive definite for all  $n \geq n_0$ .

**Lemma 4.1.5** *We have  $H_n(\beta) = \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1} A_i(\beta)^{1/2} X_i$ .*

**Proof:** Using Lemma 4.1.2 and (81), we get

$$\begin{aligned} H_n(\beta) &= \sum_{i=1}^n D_i(\beta)^T V_i(\beta, \alpha)^{-1} D_i(\beta) = \sum_{i=1}^n X_i^T A_i(\beta) A_i(\beta)^{-1/2} R_i(\alpha)^{-1} A_i(\beta)^{-1/2} A_i(\beta) X_i \\ &= \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1} A_i(\beta)^{1/2} X_i. \end{aligned}$$

□

In order to give a simplified expression for  $\mathcal{D}_n(\beta)$ , we need to introduce the following matrices.

$$\begin{aligned} B_n^{[1]}(\beta) &= \sum_{i=1}^n X_i^T \text{diag}[R_i(\alpha)^{-1} A_i(\beta)^{-1/2} (\mu_i - \mu_i(\beta))] G_i^{[1]}(\beta) X_i, \\ B_n^{[2]}(\beta) &= \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1} \text{diag}[(\mu_i - \mu_i(\beta))] G_i^{[2]}(\beta) X_i, \\ B_n(\beta) &= B_n^{[1]}(\beta) + B_n^{[2]}(\beta), \end{aligned}$$

$$\begin{aligned}\mathcal{E}^{[1]}(\beta) &= \sum_{i=1}^n X_i^T \text{diag} [R_i(\alpha)^{-1} A_i(\beta)^{-1/2} \varepsilon_i] G_i^{[1]}(\beta) X_i, \\ \mathcal{E}^{[2]}(\beta) &= \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1} \text{diag} [\varepsilon_i] G_i^{[2]}(\beta) X_i, \\ \mathcal{E}_n(\beta) &= \mathcal{E}_n^{[1]}(\beta) + \mathcal{E}_n^{[2]}(\beta), \\ G^{[1]}(\beta) &= \text{diag} \left[ \frac{\mu''(x_i^T \beta)}{2\sqrt{\mu'(x_i^T \beta)}} \right], \quad G^{[2]}(\beta) = \text{diag} \left[ \frac{-\mu''(x_i^T \beta)}{2\sqrt{(\mu'(x_i^T \beta))^3}} \right].\end{aligned}$$

Here the notation  $\text{diag}[v]$  represents the diagonal matrix with elements  $v_1, v_2, \dots, v_m$  on the diagonal.

Note that  $H_n(\beta)$ ,  $B_n(\beta)$  are non-random matrix, but  $\mathcal{E}_n(\beta)$  is a random matrix.

**Lemma 4.1.6 (Remark 1, p. 314, [20])** *We have*

$$\mathcal{D}_n(\beta) = H_n(\beta) - B_n(\beta) - \mathcal{E}_n(\beta).$$

**Proof:** To simplify the notation, we let  $R_i = R_i(\alpha)$ . By writing

$$\varepsilon_i(\beta) = \varepsilon_i + (\varepsilon_i(\beta) - \varepsilon_i) = \varepsilon_i + (\mu_i - \mu_i(\beta)),$$

we get

$$\begin{aligned}g_n(\beta) &= \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} A_i(\beta)^{-1/2} (\varepsilon_i + (\mu_i - \mu_i(\beta))) \\ &= \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} A_i(\beta)^{-1/2} \varepsilon_i + \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} A_i(\beta)^{-1/2} (\mu_i - \mu_i(\beta)).\end{aligned}$$

Differentiation with respect to  $\beta$  yields

$$\begin{aligned}\mathcal{D}_n(\beta) &= -\frac{\partial}{\partial \beta} g_n(\beta) \\ &= -\sum_{i=1}^n X_i^T G_i^{[1]}(\beta) X_i R_i^{-1} A_i(\beta)^{-1/2} \varepsilon_i - \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} G^{[2]}(\beta) X_i \varepsilon_i \\ &\quad - \sum_{i=1}^n X_i^T G_i^{[1]}(\beta) X_i R_i^{-1} A_i(\beta)^{-1/2} (\mu_i - \mu_i(\beta)) \\ &\quad - \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} G_i^{[2]}(\beta) X_i (\mu_i - \mu_i(\beta)) + \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} A_i(\beta)^{-1/2} \frac{\partial \mu_i}{\partial \beta}(\beta) \\ &= -\sum_{i=1}^n X_i^T \text{diag} [R_i^{-1} A_i(\beta)^{-1/2} \varepsilon_i] G_i^{[1]}(\beta) X_i - \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} \text{diag} [\varepsilon_i] G^{[2]}(\beta) X_i \\ &\quad - \sum_{i=1}^n X_i^T \text{diag} [R_i^{-1} A_i(\beta)^{-1/2} (\mu_i - \mu_i(\beta))] G_i^{[1]}(\beta) X_i\end{aligned}$$

$$\begin{aligned}
& - \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} \text{diag} [(\mu_i - \mu_i(\beta))] G_i^{[2]}(\beta) X_i + \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i^{-1} A_i(\beta)^{1/2} X_i \\
& = -\mathcal{E}_n^{[1]}(\beta) - \mathcal{E}_n^{[2]}(\beta) - B_n^{[1]}(\beta) - B_n^{[2]}(\beta) + H_n(\beta) = H_n(\beta) - B_n(\beta) - \mathcal{E}_n(\beta).
\end{aligned}$$

□

## 4.2 The General Results

### 4.2.1 Asymptotic Existence and Weak Consistency

We begin now to investigate the existence of a sequence of weakly consistent GEE estimators. The main tool used for proving the asymptotic existence of a solution  $\hat{\beta}_n$  of the equation  $g_n(\beta) = 0$  is an analytic result (Corollary A.3.3, Appendix A.3). We consider the following conditions, similar to conditions (D) and (C) in the GLM case (Subsection 3.2.1).

(**I<sub>w</sub>**)  $\lambda_{\min}(F_n) \longrightarrow \infty$ .

(**L<sub>w</sub>**) There exists  $c > 0$  such that for any  $r > 0$  and for any  $\varepsilon > 0$ , there exists  $n_1 = n_1(r, \varepsilon)$  with

$$P(\mathcal{D}_n(\beta)^T M_n^{-1} \mathcal{D}_n(\beta) \geq cF_n, \mathcal{D}_n(\beta) \text{ is nonsingular } \forall \beta \in N_n(r)) \geq 1 - \varepsilon, \forall n \geq n_1,$$

where  $N_n(r) = \{\beta : \|M_n^{-1/2} H_n(\beta - \beta_0)\| < r\}$ .

**Theorem 4.2.1 (Theorem 1, p. 315, [20])** *Under (**I<sub>w</sub>**) and (**L<sub>w</sub>**), there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables, such that*

(a)  $P(g_n(\hat{\beta}_n) = 0) \longrightarrow 1$ .

(b)  $\hat{\beta}_n \longrightarrow \beta_0$  in probability.

**Proof:** (p. 315, [20]) (a) Let  $c > 0$  be constant given by condition (**L<sub>w</sub>**). For any  $r > 0$ ,  $n \geq 1$ , we consider the following events:

$$\Omega_n(r) = \{\mathcal{D}_n(\beta)^T M_n^{-1} \mathcal{D}_n(\beta) \geq cF_n, \mathcal{D}_n(\beta) \text{ is nonsingular}, \forall \beta \in N_n(r)\},$$

$$E_n(r) = \{\|T_n(\beta_0)\| \leq \inf_{\beta \in \partial N_n(r)} \|T_n(\beta) - T_n(\beta_0)\|\}.$$

Note that on the event  $\Omega_n(r)$ , the function  $T_n(\beta) = M_n^{-1/2} g_n(\beta)$ ,  $\beta \in N_n(r)$  is a one to one function, because  $\dot{T}_n(\beta) = -M_n^{-1/2} \mathcal{D}_n(\beta)$  is nonsingular.

By Corollary A.3.3 (Appendix A.3), we know that on the event  $E_n(r) \cap \Omega_n(r)$ , there exists  $\hat{\beta}_n \in N_n(r)$  such that  $T_n(\hat{\beta}_n) = 0$  (i.e.  $g_n(\hat{\beta}_n) = 0$ ), that is

$$E_n(r) \cap \Omega_n(r) \subseteq \{\text{there exists } \hat{\beta}_n \in N_n(r), \text{ such that } g_n(\hat{\beta}_n) = 0\}.$$

To prove part (a) of the theorem, we will show that for any  $\varepsilon > 0$ , there exists  $r = r(\varepsilon)$  such that

$$P(E_n(r) \cap \Omega_n(r)) \geq 1 - \varepsilon, \quad \forall n \geq n_1, \quad (83)$$

where  $n_1 = n_1(r, \varepsilon)$  is given by condition ( $\mathbf{L}_w$ ).

To prove (83), let  $\varepsilon > 0$  be arbitrary and  $r = r(\varepsilon)$  be a constant to be specified later. Let  $\beta \in \partial N_n(r)$  be arbitrary. Using the Taylor's expansion of  $T_n(\cdot)$  around  $\beta_0$ , we get

$$T_n(\beta) - T_n(\beta_0) = -M_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n)(\beta - \beta_0) = -M_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-1} M_n^{1/2} \cdot M_n^{-1/2} H_n(\beta - \beta_0),$$

where  $\tilde{\beta}_n$  is a random vector which lies between  $\beta$  and  $\beta_0$ .

Using Lemma A.1.22 (Appendix A.1) with  $A = M_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-1} M_n^{1/2}$ ,  $v = M_n^{-1/2} H_n(\beta - \beta_0)$  and  $x$  an eigenvector of norm 1 of the matrix  $A^T A$  corresponding to  $\lambda_{\min}(A^T A)$ , we get that on the event  $\Omega_n(r)$  for any  $\beta \in \partial N_n(r)$

$$\begin{aligned} \|T_n(\beta) - T_n(\beta_0)\|^2 &\geq \left(x^T M_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-1} M_n^{1/2} x\right)^2 \|M_n^{-1/2} H_n(\beta - \beta_0)\|^2 = \\ &= \left(x^T M_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-1} M_n^{1/2} x\right)^2 r^2 = x^T M_n^{1/2} H_n^{-1} \left(\mathcal{D}_n(\tilde{\beta}_n)^T M_n^{-1} \mathcal{D}_n(\tilde{\beta}_n)\right) H_n^{-1} M_n^{1/2} x \cdot r^2 \\ &\geq x^T M_n^{1/2} H_n^{-1} (cF_n) H_n^{-1} M_n^{1/2} x r^2 = x^T M_n^{1/2} H_n^{-1} (cH_n M_n^{-1} H_n) H_n^{-1} M_n^{1/2} x r^2 = cr^2. \end{aligned}$$

Hence

$$\Omega_n(r) \subseteq \left\{ \inf_{\beta \in \partial N_n(r)} \|T_n(\beta) - T_n(\beta_0)\| \geq c^{1/2} r \right\}. \quad (84)$$

Using Bonferroni's inequality

$$\begin{aligned} P(E_n(r) \cap \Omega_n(r)) &\geq P(\{\|T_n(\beta_0)\| \leq c^{1/2} r \leq \inf_{\beta \in \partial N_n(r)} \|T_n(\beta) - T_n(\beta_0)\|\} \cap \Omega_n(r)) \\ &\geq P(\|T_n(\beta_0)\| \leq c^{1/2} r) + P(\{c^{1/2} r \leq \inf_{\beta \in \partial N_n(r)} \|T_n(\beta) - T_n(\beta_0)\|\} \cap \Omega_n(r)) - 1 \\ &= P(\|T_n(\beta_0)\| \leq c^{1/2} r) + P(\Omega_n(r)) - 1, \quad \forall n \geq 1, \end{aligned} \quad (85)$$

where we used (84) for the last equality above. For the first term, by the Chebyshev's inequality we have:

$$\begin{aligned}
P(\|T_n(\beta_0)\| \leq c^{1/2}r) &= 1 - P(\|T_n(\beta_0)\| > c^{1/2}r) \geq 1 - \frac{1}{cr^2}E\|T_n(\beta_0)\|^2 = \\
&= 1 - \frac{1}{cr^2}E[\text{tr}(M_n^{-1/2}g_n^T g_n M_n^{-1/2})] = 1 - \frac{1}{cr^2}\text{tr}[M_n^{-1/2}E(g_n^T g_n)M_n^{-1/2}] \\
&= 1 - \frac{1}{cr^2}\text{tr}(I) = 1 - \frac{p}{cr^2} = 1 - \frac{\varepsilon}{2}, \quad \forall n \geq 1,
\end{aligned} \tag{86}$$

by choosing  $r = \sqrt{(2p)/(c\varepsilon)}$ .

For the second term, we know that by  $(\mathbf{L}_w)$  that there exists  $n_1 = n_1(r(\varepsilon), \varepsilon)$  such that

$$P(\Omega_n(r)) \geq 1 - \frac{\varepsilon}{2}, \quad \forall n \geq n_1. \tag{87}$$

From (85), (86) and (87) we get

$$P(E_n(r) \cap \Omega_n(r)) \geq (1 - \frac{\varepsilon}{2}) + (1 - \frac{\varepsilon}{2}) - 1 = 1 - \varepsilon, \quad \forall n \geq n_1.$$

This concludes the proof of (83).

**(b)** Let  $\eta, \varepsilon > 0$  be arbitrary. We need to prove that there exists a  $N = N_{\eta, \varepsilon}$  such that

$$P(\|\hat{\beta}_n - \beta_0\| \leq \eta) \geq 1 - \varepsilon, \quad \forall n \geq N. \tag{88}$$

In part **(a)** we proved the existence of  $\hat{\beta}_n$  on the event  $E_n(r) \cap \Omega_n(r)$ , where  $r = \sqrt{(2p)/(c\varepsilon)}$ . Moreover  $\hat{\beta}_n \in N_n(r)$ .

By  $(\mathbf{I}_w)$ , there exists  $N = N_\eta > n_1$  such that  $\|F_n\| > (r/\eta)^2, \forall n \geq N$ .

Since  $\hat{\beta}_n \in N_n(r)$ , on the event  $E_n(r) \cap \Omega_n(r)$  we have:

$$\begin{aligned}
\|\hat{\beta}_n - \beta_0\| &\leq \|(M_n^{-1/2}H_n)^{-1}\| \|M_n^{-1/2}H_n(\hat{\beta}_n - \beta_0)\| \leq \\
&\leq \frac{r}{\|M_n^{-1/2}H_n\|} = \frac{r}{\|H_n M_n^{-1} H_n\|^{1/2}} = \frac{r}{\|F_n\|^{1/2}} \leq \eta, \quad \forall n \geq N
\end{aligned}$$

and therefore, by (83)

$$1 - \varepsilon \leq P(E_n(r) \cap \Omega_n(r)) \leq P(\|\hat{\beta}_n - \beta_0\| \leq \eta), \quad \forall n \geq N.$$

This concludes the proof of (88).

□

Our next goal is to eliminate from the set of conditions the covariance matrix  $M_n$ , which depends on the unknown true correlation matrices  $\bar{R}_i$ . This goal will be achieved in two steps. First we need the following lemma.

**Lemma 4.2.2** *We have  $\tau_n H_n(\beta) \geq M_n(\beta)$ .*

**Proof:** From Lemma 4.1.4 and Lemma 4.1.5 we have

$$\begin{aligned} & \tau_n(\beta)H_n(\beta) - M_n(\beta) = \\ & = \sum_{i=1}^n X_i^T A_i(\beta)^{1/2} R_i(\alpha)^{-1/2} (\tau_n I - R_i(\alpha)^{-1/2} \bar{R}_i R_i(\alpha)^{-1/2}) R_i(\alpha)^{-1/2} A_i(\beta)^{1/2} X_i. \end{aligned}$$

Since  $R_i^{-1/2}(\alpha) \bar{R}_i R_i^{-1/2}(\alpha) \leq \lambda_{\max}(R_i^{-1/2}(\alpha) \bar{R}_i R_i^{-1/2}(\alpha)) I \leq \tau_n I$ , for all  $i \leq n$  the conclusion follows.

□

Here is the new set of conditions, which are still dependent on  $\bar{R}_i$ , since they rely on  $\tau_n$ .

$$(\mathbf{I}_w^*) \quad \tau_n^{-1} \lambda_{\min}(H_n) \longrightarrow \infty.$$

( $\mathbf{L}_w^*$ ) There exists a constant  $c > 0$  such that for any  $r > 0$  and for any  $\varepsilon > 0$ , there exists  $n_1 = n_1(r, \varepsilon)$  with

$$P(\mathcal{D}_n(\beta) \geq cH_n, \forall \beta \in N_n^*(r)) \geq 1 - \varepsilon, \quad \forall n \geq n_1,$$

where  $N_n^*(r) = \{\beta : \|H_n^{1/2}(\beta - \beta_0)\| \leq (\tau_n)^{1/2} r\}$ .

**Theorem 4.2.3 (Theorem 2, p. 317, [20])** *Under ( $\mathbf{I}_w^*$ ) and ( $\mathbf{L}_w^*$ ), there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that*

$$(a) \quad P(g_n(\hat{\beta}_n) = 0) \longrightarrow 1.$$

(b)  $\hat{\beta}_n \longrightarrow \beta_0$  in probability.

**Proof:** (p. 317, [20]) The proof is similar to that of Theorem 4.2.1.

(a) Let  $c > 0$  be the constant given by condition ( $\mathbf{L}_w^*$ ). For any  $r > 0$ ,  $n \geq 1$ , we consider the following events:

$$\Omega_n^*(r) = \{\mathcal{D}_n(\beta) \geq cH_n, \forall \beta \in N_n^*(r)\},$$

$$E_n^*(r) = \{\|T_n^*(\beta_0)\| \leq \inf_{\beta \in \partial N_n^*(r)} \|T_n^*(\beta) - T_n^*(\beta_0)\|\}.$$

Note that on the event  $\Omega_n^*(r)$ , the function  $T_n^*(\beta) = H_n^{-1/2}g_n(\beta)$ ,  $\beta \in N_n^*(r)$  is a one to one function :

$$\dot{T}_n^*(\beta) = -H_n^{-1/2}\mathcal{D}_n(\beta) \leq -cH_n^{-1/2}H_n = -cH_n^{1/2} < 0, \quad \forall \beta \in N_n^*(r).$$

By Corollary A.3.3 (Appendix A.3), we know that on the event  $E_n^*(r) \cap \Omega_n^*(r)$  there exists  $\hat{\beta}_n \in N_n^*(r)$  such that  $T_n^*(\hat{\beta}_n) = 0$  (i.e.  $g_n(\hat{\beta}_n) = 0$ ), that is

$$E_n^*(r) \cap \Omega_n^*(r) \subseteq \{\text{there exists } \hat{\beta}_n \in N_n^*(r) \text{ such that } g_n(\hat{\beta}_n) = 0\}.$$

To prove part **(a)** of the theorem, we will show that for any  $\varepsilon > 0$  there exists  $r = r(\varepsilon)$  such that

$$P(E_n^*(r) \cap \Omega_n^*(r)) \geq 1 - \varepsilon, \quad \forall n \geq n_1, \quad (89)$$

where  $n_1 = n_1(r, \varepsilon)$  is given by condition  $(\mathbf{L}_w^*)$ .

To prove (89), let  $\varepsilon > 0$  be arbitrary and  $r = r(\varepsilon)$  be a constant to be specified later. Let  $\beta \in \partial N_n^*(r)$  be arbitrary. Using the Taylor's expansion of  $T_n^*(\cdot)$  around  $\beta_0$ , we get

$$T_n^*(\beta) - T_n^*(\beta_0) = -H_n^{-1/2}\mathcal{D}_n(\tilde{\beta}_n)(\beta - \beta_0) = -H_n^{-1/2}\mathcal{D}_n(\tilde{\beta}_n)H_n^{-1/2}H_n^{1/2}(\beta - \beta_0),$$

where  $\tilde{\beta}_n$  lies between  $\beta$  and  $\beta_0$ . Hence, using Lemma A.1.22 (Appendix A.1) with  $A = H_n^{-1/2}\mathcal{D}_n(\tilde{\beta})H_n^{-1/2}$ ,  $v = H_n^{1/2}(\beta - \beta_0)$  and  $x$  an eigenvector of norm 1 of the matrix  $A^T A$  corresponding to  $\lambda_{\min}(A^T A)$ , we get that on the event  $\Omega_n^*(r)$ , for any  $\beta \in \partial N_n^*(r)$  we have:

$$\begin{aligned} \|T_n^*(\beta) - T_n^*(\beta_0)\|^2 &\geq \left(x^T H_n^{-1/2}\mathcal{D}_n(\tilde{\beta})H_n^{-1/2}x\right)^2 \|H_n^{1/2}(\beta - \beta_0)\|^2 \geq \\ &\geq (cx^T H_n^{-1/2}H_n H_n^{-1/2}x)^2 r^2 \tau_n = c^2 r^2 \tau_n. \end{aligned}$$

Hence

$$\Omega_n^*(r) \subseteq \left\{ \inf_{\beta \in \partial N_n^*(r)} \|T_n^*(\beta) - T_n^*(\beta_0)\| \geq cr\tau_n^{1/2} \right\}. \quad (90)$$

Using Bonferroni's inequality

$$P(E_n^*(r) \cap \Omega_n^*(r)) \geq P\left(\left\{ \inf_{\beta \in \partial N_n^*(r)} \|T_n^*(\beta) - T_n^*(\beta_0)\| \geq cr\tau_n^{1/2} \geq \|T_n^*(\beta_0)\| \right\} \cap \Omega_n^*(r)\right) \geq$$

$$\begin{aligned}
&\geq P(\|T_n^*(\beta_0)\| \leq cr\tau_n^{1/2}) + P\left(\left\{\inf_{\beta \in \partial N_n^*(r)} \|T_n^*(\beta) - T_n^*(\beta_0)\| \geq cr\tau_n^{1/2}\right\} \cap \Omega_n^*(r)\right) - 1 = \\
&= P(\|T_n^*(\beta_0)\| \leq cr\tau_n^{1/2}) + P(\Omega_n^*(r)) - 1, \quad \forall n \geq 1,
\end{aligned} \tag{91}$$

where we used (90) for the last equality. For the first term, by the Chebyshev's inequality we have

$$\begin{aligned}
P(\|T_n^*(\beta_0)\| \leq cr\tau_n^{1/2}) &= 1 - P(\|T_n^*(\beta_0)\| > cr\tau_n^{1/2}) \geq 1 - \frac{E\|T_n^*(\beta_0)\|^2}{c^2r^2\tau_n} \\
&= 1 - \frac{E[\text{tr}(H_n^{-1/2}g_n g_n^T H_n^{-1/2})]}{c^2r^2\tau_n} = 1 - \frac{\text{tr}[H_n^{-1/2}E(g_n g_n^T)H_n^{-1/2}]}{c^2r^2\tau_n} \\
&= 1 - \frac{\text{tr}(H_n^{-1/2}M_n H_n^{-1/2})}{c^2r^2\tau_n} \geq 1 - \frac{\tau_n \text{tr}(H_n^{-1/2}H_n H_n^{-1/2})}{\tau_n c^2r^2} = 1 - \frac{p}{c^2r^2} \\
&= 1 - \frac{\varepsilon}{2}, \quad \forall n \geq 1
\end{aligned} \tag{92}$$

by choosing  $r = \sqrt{(2p)/(c^2\varepsilon)}$ . (Note that we used Lemma 4.2.2 for the last inequality.) For the second term, we know that by  $(\mathbf{L}_w^*)$  that there exists  $n_1 = n_1(r(\varepsilon), \varepsilon)$  such that

$$P(\Omega_n^*(r)) \geq 1 - \frac{\varepsilon}{2}, \quad \forall n \geq n_1, \tag{93}$$

From (91), (92) and (93) we get

$$P(E_n^*(r) \cap \Omega_n^*(r)) \geq (1 - \frac{\varepsilon}{2}) + (1 - \frac{\varepsilon}{2}) - 1 = 1 - \varepsilon, \quad \forall n \geq n_1.$$

This concludes the proof of (93).

**(b)** Let  $\eta, \varepsilon > 0$  be arbitrary. We need to prove that there exists  $N = N_{\varepsilon, \eta}$  such that (88) holds. In part **(a)** we proved the existence of  $\hat{\beta}_n$  on the event  $\Omega_n^*(r) \cap E_n^*(r)$ , where  $r = \sqrt{(2p)/(c^2\varepsilon)}$ . Moreover,  $\hat{\beta}_n \in N_n^*(r)$ .

By  $(\mathbf{I}_w^*)$ , there exists  $N = N_\eta > n_1$  such that  $\tau_n^{-1}\lambda_{\min}(H_n) > (r/\eta)^2, \forall n \geq N$ . Since  $\hat{\beta}_n \in N_n^*(r)$ , on the event  $E_n^*(r) \cap \Omega_n^*(r)$  we have

$$\|\hat{\beta}_n - \beta_0\| \leq \|H_n^{-1/2}\| \|H_n^{1/2}(\hat{\beta}_n - \beta_0)\| \leq \frac{\tau_n^{1/2}r}{\|H_n\|^{1/2}} \leq \eta, \quad \forall n \geq N$$

and therefore, by (89),

$$1 - \varepsilon \leq P(E_n^*(r) \cap \Omega_n^*(r)) \leq P(\|\hat{\beta}_n - \beta_0\| \leq \eta), \quad \forall n \geq N.$$

This concludes the proof of (88).

□

In order to eliminate  $\tau_n$  from the set of conditions, we consider

$$\tilde{\lambda}_n = \max_{1 \leq i \leq n} \lambda_{\max}(R_i^{-1}(\alpha)).$$

Note that by Lemma A.1.21 (Appendix A.1), we have

$$\tau_n \leq m\tilde{\lambda}_n,$$

and hence  $m\tilde{\lambda}_n H_n(\beta) \geq M_n(\beta)$  by Lemma 4.2.2.

The final set of conditions (which do not depend on  $\bar{R}_i$ ) is obtained by replacing  $\tau_n$  with  $m\tilde{\lambda}_n$ :

$$(\mathbf{I}_w^+) (m\tilde{\lambda}_n)^{-1} \lambda_{\min}(H_n) \longrightarrow \infty.$$

( $\mathbf{L}_w^+$ ) There exists  $c > 0$  such that for any  $r > 0$  and for any  $\varepsilon > 0$ , there exists  $n_1 = n_1(r, \varepsilon)$  with

$$P(\mathcal{D}_n(\beta) \geq cH_n, \forall \beta \in N_n^+(r)) \geq 1 - \varepsilon, \forall n \geq n_1,$$

where  $N_n^+(r) = \{\beta : \|H_n^{1/2}(\beta - \beta_0)\| \leq (m\tilde{\lambda}_n)^{1/2}r\}$ .

**Theorem 4.2.4 (Remark 5, p.318 [20])** Under ( $\mathbf{I}_w^+$ ) and ( $\mathbf{L}_w^+$ ), there exists a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that

$$(a) P(g_n(\hat{\beta}_n) = 0) \longrightarrow 1.$$

$$(b) \hat{\beta}_n \longrightarrow \beta_0 \text{ in probability.}$$

The proof is very similar to that of Theorem 4.2.3 and therefore it is omitted.

## 4.2.2 Asymptotic Normality

We consider the condition :

(CC) For any  $r > 0$ , we have as  $n \longrightarrow \infty$

$$\sup_{\beta \in N_n^*(r)} \|H_n^{-1/2} \mathcal{D}_n(\beta) H_n^{-T/2} - I\| \longrightarrow 0 \text{ in probability.}$$

**Lemma 4.2.5 (p. 319, [20])** (CC) implies ( $\mathbf{L}_w^*$ ).

**Proof:** (CC) is equivalent to the following assertion for any  $r > 0$ ,  $\varepsilon > 0$  and  $\eta > 0$ , there exists  $n_1 = n_1(r, \varepsilon, \eta)$ , such that

$$P\left(\|H_n^{-1/2}(\mathcal{D}_n(\beta) - H_n)H_n^{-T/2}\| \leq \eta, \forall \beta \in N_n^*(r)\right) \geq 1 - \varepsilon, \forall n \geq N_1.$$

By Corollary A.1.17 (Appendix A.1), this becomes

$$(CC') P\left(\left|\lambda^T H_n^{-1/2}(\mathcal{D}_n(\beta) - H_n)H_n^{-T/2}\lambda\right| \leq \eta(\lambda^T \lambda), \forall \lambda \in R^p, \forall \beta \in N_n^*(r)\right) \geq 1 - \varepsilon, \forall n \geq n_1.$$

By taking  $x = H_n^{-T/2}\lambda$ , the previous relation can be written as:

$$(CC'') P\left(|x^T \mathcal{D}_n(\beta) - x^T H_n x| \leq \eta(x^T H_n x), \forall x \in R^p, \forall \beta \in N_n^*(r)\right) \geq 1 - \varepsilon, \forall n \geq n_1.$$

In particular, this implies:

$$P\left(x^T \mathcal{D}_n(\beta)x \geq (1 - \eta)x^T H_n x, \forall x \in R^p, \forall \beta \in N_n^*(r)\right) \geq 1 - \varepsilon, \forall n \geq n_1,$$

i.e.

$$P\left(\mathcal{D}_n(\beta) \geq (1 - \eta)H_n, \forall \beta \in N_n^*(r)\right) \geq 1 - \varepsilon, \forall n \geq n_1.$$

By fixing  $\eta \in (0, 1)$  and letting  $c = 1 - \eta$ , we obtain exactly ( $\mathbf{L}_w^*$ ).

□

We consider the following condition:

( $\mathbf{N}_\delta$ ) There exist  $\delta > 0$  and  $K > 0$  such that

$$(i) \sup_{i \geq 1} \max_{j \leq m} E(y_{ij}^*)^{2+2/\delta} \leq K, \text{ where } y_i^* = A_i^{-1/2} \varepsilon_i.$$

$$(ii) (c_n \tilde{\lambda}_n)^{1+\delta} \gamma_n^{(D)} \longrightarrow 0, \text{ where } \gamma_n^{(D)} = \max_{1 \leq i \leq n} \lambda_{\max}(H_n^{-1/2} D_i^T V_i(\alpha)^{-1} D_i H_n^{-1/2}) \text{ and } c_n = \lambda_{\max}(M_n^{-1} H_n).$$

In order to prove the asymptotic normality of  $g_n$ , we need the following elementary result.

**Lemma 4.2.6** *Let  $\varphi(\cdot)$  be a nondecreasing function. For any positive random variable  $Y$  and  $c > 0$ , we have*

$$E(Y I_{\{Y \geq c\}}) \leq \frac{1}{\varphi(c/m)} E[Y \varphi(Y/m)].$$

**Proof:** Since  $\varphi(\cdot)$  is nondecreasing, on the event  $\{Y \geq c\}$ , we have  $1 \leq [\varphi(Y/m)]/[\varphi(c/m)]$ . Hence

$$E(YI_{\{Y \geq c\}}) \leq \frac{1}{\varphi(c/m)} E[Y\varphi(Y/m)].$$

□

**Lemma 4.2.7 (Lemma 2, p. 321, [20])** *Under  $(\mathbf{N}_\delta)$  we have*

$$M_n^{-1/2} g_n \longrightarrow N(0, I), \text{ in distribution.}$$

**Proof:** By the Cràmer-Wold theorem (Theorem A.2.4, Appendix A.2), it suffices to prove that

$$\lambda^T M_n^{-1/2} g_n \longrightarrow N(0, 1) \text{ in distribution,} \quad (94)$$

for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ . Recall that  $g_n = \sum_{i=1}^n D_i^T V_i(\alpha)^{-1} \varepsilon_i$ . Hence

$$\lambda^T M_n^{-1/2} g_n = \sum_{i=1}^n \lambda^T M_n^{-1/2} D_i^T V_i(\alpha)^{-1} \varepsilon_i := \sum_{i=1}^n Z_{ni}.$$

Note that  $\{Z_{ni}\}_{i=1, \dots, n}$  are zero-mean independent random variables with

$$E[(\lambda^T M_n^{-1/2} g_n)^2] = E(\lambda^T M_n^{-1/2} g_n g_n^T M_n^{-T/2} \lambda) = \lambda^T M_n^{-1/2} E(g_n g_n^T) M_n^{-T/2} \lambda = 1.$$

Relation (94) will follow by the Central Limit Theorem (Theorem A.2.8, Appendix A.2) providing that the Lindeberg condition holds, i.e.

$$\sum_{i=1}^n E(Z_{ni}^2 I_{\{|Z_{ni}| \geq \varepsilon\}}) \longrightarrow 0. \quad (95)$$

In order to prove (95), note that

$$\begin{aligned} Z_{ni}^2 &= (\lambda^T M_n^{-1/2} D_i^T V_i(\alpha)^{-1/2} V_i(\alpha)^{-1/2} \varepsilon_i)^2 \leq \|\lambda^T M_n^{-1/2} D_i^T V_i(\alpha)^{-1/2}\|^2 \|V_i(\alpha)^{-1/2} \varepsilon_i\|^2 = \\ &= \|\lambda^T M_n^{-1/2} H_n^{1/2} H_n^{-1/2} D_i^T V_i(\alpha)^{-1/2}\|^2 \|V_i(\alpha)^{-1/2} \varepsilon_i\|^2 \leq \\ &\leq \|\lambda\|^2 \|M_n^{-1/2} H_n^{1/2}\|^2 \|H_n^{-1/2} D_i^T V_i(\alpha)^{-1/2}\|^2 \|V_i(\alpha)^{-1/2} \varepsilon_i\|^2 \leq \\ &\leq \lambda_{\max}(M_n^{-1} H_n) (x^T H_n^{-1/2} D_i^T V_i(\alpha)^{-1} D_i H_n^{-1/2} x) (\varepsilon_i^T V_i(\alpha)^{-1} \varepsilon_i) \leq \end{aligned}$$

$$\leq \lambda_{\max}(M_n^{-1}H_n)\gamma_{n,i}^{(D)}(y_i^*)^T R_i(\alpha)^{-1}y_i^* \leq c_n\gamma_{n,i}^{(D)}\tilde{\lambda}_n(y_i^*)^T y_i^*,$$

where  $\gamma_{n,i}^{(D)} = x^T H_n^{-1/2} D_i^T V_i(\alpha)^{-1} D_i H_n^{-1/2} x$  and  $\|x\| = 1$ .

Note that  $\gamma_{n,i}^{(D)} \leq \gamma_n^{(D)}$  for all  $i \leq n$ . Hence

$$I_{\{|Z_{ni}| \geq \varepsilon\}} = I_{\{Z_{ni}^2 \geq \varepsilon^2\}} \leq I_{\{(y_i^*)^T y_i^* \geq \varepsilon^2 / (c_n \gamma_{n,i}^{(D)} \tilde{\lambda}_n)\}}. \quad (96)$$

Let  $\varphi(t) = t^{1/\delta}$ . By applying Lemma 4.2.6 with the choice  $Y = (y_i^*)^T y_i^*$  and  $c = \varepsilon^2 / (c_n \tilde{\lambda}_n \gamma_{n,i}^{(D)})$ , we obtain

$$E \left[ (y_i^*)^T y_i^* I_{\{(y_i^*)^T y_i^* \geq \varepsilon^2 / (c_n \tilde{\lambda}_n \gamma_{n,i}^{(D)})\}} \right] \leq \left[ \varphi \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_{n,i}^{(D)}} \right) \right]^{-1} E \left[ (y_i^*)^T y_i^* \varphi \left( \frac{(y_i^*)^T y_i^*}{m} \right) \right]. \quad (97)$$

By (96) and (97), we get

$$\begin{aligned} \sum_{i=1}^n E(Z_{ni}^2 I_{\{|Z_{ni}| \geq \varepsilon\}}) &\leq c_n \tilde{\lambda}_n \sum_{i=1}^n \gamma_{n,i}^{(D)} E \left[ (y_i^*)^T y_i^* I_{\{(y_i^*)^T y_i^* \geq \varepsilon^2 / (c_n \tilde{\lambda}_n \gamma_{n,i}^{(D)})\}} \right] \leq \\ &\leq c_n \tilde{\lambda}_n \left[ \varphi \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_{n,i}^{(D)}} \right) \right]^{-1} \sum_{i=1}^n \gamma_{n,i}^{(D)} E \left[ (y_i^*)^T y_i^* \varphi \left( \frac{(y_i^*)^T y_i^*}{m} \right) \right] \leq \\ &\leq m c_n \tilde{\lambda}_n \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_n^{(D)}} \right)^{-1/\delta} \sum_{i=1}^n \gamma_{n,i}^{(D)} E \left[ \frac{(y_i^*)^T y_i^*}{m} \varphi \left( \frac{(y_i^*)^T y_i^*}{m} \right) \right], \end{aligned} \quad (98)$$

where for the last inequality, we use the fact that  $\varphi(\cdot)$  is nondecreasing and  $\gamma_{n,i}^{(D)} \leq \gamma_n^{(D)}$ , for all  $i \leq n$ .

In what follows, we let  $\phi(t) = t\varphi(t) = t^{1+1/\delta}$ . Since  $\phi(\cdot)$  is a convex function, we have

$$\phi \left( \frac{1}{m} \sum_{j=1}^m (y_{ij}^*)^2 \right) \leq \frac{1}{m} \sum_{j=1}^m \phi((y_{ij}^*)^2),$$

that is

$$\frac{\sum_{j=1}^m (y_{ij}^*)^2}{m} \varphi \left( \frac{1}{m} \sum_{j=1}^m (y_{ij}^*)^2 \right) \leq \frac{1}{m} \sum_{j=1}^m (y_{ij}^*)^2 \varphi((y_{ij}^*)^2).$$

Taking expectation on both sides

$$E \left[ \frac{\sum_{j=1}^m (y_{ij}^*)^2}{m} \varphi \left( \frac{\sum_{j=1}^m (y_{ij}^*)^2}{m} \right) \right] \leq \frac{1}{m} E \left[ \sum_{j=1}^m (y_{ij}^*)^2 \varphi \left( (y_{ij}^*)^2 \right) \right] = \frac{1}{m} \sum_{j=1}^m E (y_{ij}^*)^{2+2/\delta}. \quad (99)$$

From (98) and (99) we obtain

$$\begin{aligned} \sum_{i=1}^n E (Z_{ni}^2 I_{\{|z_{ni}| \geq \varepsilon\}}) &\leq m c_n \tilde{\lambda}_n \left[ \varphi \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_n^{(D)}} \right) \right]^{-1} \sum_{i=1}^n \gamma_{n,i}^{(D)} \frac{1}{m} \sum_{j=1}^m E (y_{ij}^*)^{2+2/\delta} \\ &\leq m c_n \tilde{\lambda}_n \left[ \varphi \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_n^{(D)}} \right) \right]^{-1} \sum_{i=1}^n \gamma_{n,i}^{(D)} \frac{1}{m} (mK) \\ &= mK c_n \tilde{\lambda}_n \left[ \varphi \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_n^{(D)}} \right) \right]^{-1} \sum_{i=1}^n \gamma_{n,i}^{(D)}, \end{aligned} \quad (100)$$

where we used  $(\mathbf{N}_\delta)$  (i) in the second inequality above.

Note that

$$\begin{aligned} \sum_{i=1}^n \gamma_{n,i}^{(D)} &= \sum_{i=1}^n x^T H_n^{-1/2} D_i^T V_i(\alpha)^{-1} D_i H_n^{-1/2} x = \\ &= x^T H_n^{-1/2} \left( \sum_{i=1}^n D_i^T V_i(\alpha)^{-1} D_i \right) H_n^{-1/2} x = x^T H_n^{-1/2} H_n H_n^{-1/2} x = 1. \end{aligned} \quad (101)$$

and

$$c_n \tilde{\lambda}_n \left[ \varphi \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_n^{(D)}} \right) \right]^{-1} = c_n \tilde{\lambda}_n \left( \frac{\varepsilon^2}{m c_n \tilde{\lambda}_n \gamma_n^{(D)}} \right)^{-1/\delta} = \left[ \frac{1}{\varepsilon^2} m (c_n \tilde{\lambda}_n)^{1+\delta} \gamma_n^{(D)} \right]^{1/\delta}. \quad (102)$$

From (100), (101) and (102) we conclude that

$$\sum_{i=1}^n E (Z_{n,i}^2 I_{\{|Z_{n,i}| \geq \varepsilon\}}) \leq mK \left[ \frac{1}{\varepsilon^2} m (c_n \tilde{\lambda}_n)^{1+\delta} \gamma_n^{(D)} \right]^{1/\delta}.$$

Relation (95) follows by  $(\mathbf{N}_\delta)$  (ii).

□

**Theorem 4.2.8 (Theorem 4, p. 322, [20])** *Suppose that conditions  $(\mathbf{I}_w)$ ,  $(\mathbf{L}_w)$  and  $(\mathbf{CC})$  hold, or conditions  $(\mathbf{I}_w^*)$  and  $(\mathbf{CC})$  hold. If in addition  $(\mathbf{N}_\delta)$  holds, then*

$$M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \longrightarrow N(0, I) \text{ in distribution.}$$

**Proof:** We focus on the event  $\{g_n(\hat{\beta}_n) = 0\}$  whose probability converges to 1, by Theorem 4.2.1, respectively Theorem 4.2.2. Using the Taylor's expansion for the function  $H_n^{-1/2} g_n(\cdot)$  around  $\beta_0$ , we get

$$\begin{aligned} H_n^{-1/2} g_n &= -H_n^{-1/2} \left( g_n(\hat{\beta}_n) - g_n \right) = H_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n)(\hat{\beta}_n - \beta_0) \\ &= \left[ H_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-T/2} \right] H_n^{T/2}(\hat{\beta}_n - \beta_0), \end{aligned}$$

where  $\tilde{\beta}_n$  is a random vector which lies between  $\hat{\beta}_n$  and  $\beta_0$ .

Multiplying by  $M_n^{-1/2} H_n^{1/2}$ , we get

$$M_n^{-1/2} g_n = M_n^{-1/2} H_n^{1/2} \left[ H_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-T/2} \right] \cdot H_n^{T/2}(\hat{\beta}_n - \beta_0).$$

Exactly as for the proof of Theorem 3.2.4, one can show that under  $(\mathbf{CC})$ ,

$$\left\| H_n^{-1/2} \mathcal{D}_n(\tilde{\beta}_n) H_n^{-T/2} - I \right\| \longrightarrow 0 \text{ in probability.}$$

The conclusion follows by Lemma 4.2.7.

□

### 4.2.3 Asymptotic Existence and Strong Consistency

We begin now to investigate the existence of a sequence of strongly consistent estimators. Consider the following conditions:

$(\mathbf{I}_s)^{\text{SP}}$   $\lambda_{\min}(M_n) \longrightarrow \infty$ , when  $n \longrightarrow \infty$ .

$(\mathbf{L}_s)^{\text{SP}}$  There exist some constants  $c > 0$ ,  $\delta > 0$  and a neighborhood  $N$  of  $\beta_0$  such that with probability 1, there exists a random number  $n_1$  such that

$$|x^T \mathcal{D}_n(\beta) x| \geq c[\lambda_{\max}(M_n)]^{1/2+\delta}, \quad \forall x \in \mathbb{R}^p, \quad \forall \beta \in N, \quad \forall n \geq n_1.$$

**Remarks:** (a) In general, the matrix  $\mathcal{D}_n(\beta)$  is not symmetric. Hence the eigenvalues of  $\mathcal{D}_n(\beta)$  are not necessarily real numbers, and we can not define  $\lambda_{\min}(\mathcal{D}_n(\beta))$ .

(b) Under conditions  $(\mathbf{I}_s)^{\text{SP}}$  and  $(\mathbf{L}_s)^{\text{SP}}$ , with probability 1, we have

$$|x^T \mathcal{D}_n(\beta)x| > 0, \forall x \in R^p, \forall \beta \in N, \forall n \geq n_1$$

and hence, with probability 1

$$\mathcal{D}_n(\beta) \text{ is nonsingular } \forall \beta \in N, \forall n \geq n_1.$$

**Lemma 4.2.9** Under  $(\mathbf{I}_s)^{\text{SP}}$ , for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ , we have

$$\frac{\lambda^T g_n}{[\lambda_{\max}(M_n)]^{1/2+\delta}} \longrightarrow 0 \text{ a.s.}$$

**Proof:** Note that

$$\lambda^T g_n = \lambda^T \sum_{i=1}^n u_i = \sum_{i=1}^n \lambda^T u_i,$$

where  $u_i = X_i^T A_i^{1/2} R_i(\alpha)^{-1} A_i^{-1/2} \varepsilon_i$ ,  $i \geq 1$  are zero-mean independent random variables, and

$$E[(\lambda^T g_n)^2] = \lambda^T E(g_n g_n^T) \lambda = \lambda^T M_n \lambda \leq \lambda_{\max}(M_n).$$

The conclusion follows by Theorem A.2.1 (Appendix A.2).

□

**Theorem 4.2.10 (Theorem 7, p. 327 [20])** Under  $(\mathbf{I}_s)^{\text{SP}}$  and  $(\mathbf{L}_s)^{\text{SP}}$ , there exist a sequence  $\{\hat{\beta}_n\}_{n \geq 1} \subseteq N$  of random variables and there exists a random number  $n_2$  such that

(a)  $P(g(\hat{\beta}_n) = 0, \forall n \geq n_2) = 1.$

(b)  $\hat{\beta}_n \longrightarrow \beta_0 \text{ a.s.}$

**Proof:** Choose  $\varepsilon_0 > 0$  such that the neighborhood  $B_\varepsilon(\beta_0) = \{\beta : \|\beta - \beta_0\| < \varepsilon_0\} \subseteq N$ . By  $(\mathbf{L}_s)^{\text{SP}}$  we conclude that with probability 1, the function  $T_n(\beta) = [\lambda_{\max}(M_n)]^{-1/2-\delta} g_n(\beta)$  is a one to one function on  $N$ , because

$$\dot{T}_n(\beta) = -[\lambda_{\max}(M_n)]^{-1/2-\delta} \mathcal{D}_n(\beta)$$

is nonsingular for all  $\beta \in N$  and  $n \geq n_1$ .

We will show that with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$  there exists a random number  $n_2 = n_2(\varepsilon)$  such that

$$\|T_n(\beta_0)\| \leq \inf_{\beta \in \partial B_\varepsilon(\beta_0)} \|T_n(\beta) - T_n(\beta_0)\|, \quad \forall n \geq n_2. \quad (103)$$

By Corollary A.3.3 (Appendix A.3), this will imply the conclusion of our theorem in the following format: with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$ , there exists  $n_2 = n_2(\varepsilon)$  (random) such that for all  $n \geq n_2$

$$\text{there exists } \hat{\beta}_n \in B_\varepsilon(\beta_0) \text{ with } g_n(\hat{\beta}_n) = 0.$$

(The fact that  $\hat{\beta}_n$  does not depend on  $\varepsilon$  follows from the uniqueness of the zeros' of the function  $g_n(\cdot)$ . It is also that  $\hat{\beta}_n \rightarrow \beta_0$  a.s.)

To prove (103), we use the Taylor's expansion of  $T_n(\cdot)$  around  $\beta_0$ : for  $\beta \in \partial B_\varepsilon(\beta_0)$ ,  $\varepsilon \in (0, \varepsilon_0)$  we have

$$T_n(\beta) - T_n(\beta_0) = [\lambda_{\max}(M_n)]^{-(1/2+\delta)} \mathcal{D}_n(\tilde{\beta}_n)(\beta - \beta_0),$$

where  $\tilde{\beta}_n$  lies between  $\beta$  and  $\beta_0$ . Hence using Lemma A.1.22 (Appendix A.1) with  $A = \mathcal{D}_n(\tilde{\beta})$ ,  $v = \beta - \beta_0$ ,  $x$  is an eigenvector corresponding to  $\lambda_{\min}(\mathcal{D}_n(\tilde{\beta}_n)^T \mathcal{D}_n(\tilde{\beta}_n))$  ( $\|x\| = 1$ ) and  $(\mathbf{L}_s)^{\text{SP}}$  we get that with probability 1, there exists a random number  $n_1$  such that for any  $\varepsilon \in (0, \varepsilon_0)$ ,  $\beta \in \partial B_\varepsilon(\beta_0)$ , we have:

$$\begin{aligned} \|T_n(\beta) - T_n(\beta_0)\|^2 &= [\lambda_{\max}(M_n)]^{-2(1/2+\delta)} \|\mathcal{D}_n(\tilde{\beta}_n)(\beta - \beta_0)\|^2 \geq \\ &\geq [\lambda_{\max}(M_n)]^{-2(1/2+\delta)} (x^T \mathcal{D}_n(\tilde{\beta}_n) x)^2 \varepsilon^2 \geq \\ &\geq [\lambda_{\max}(M_n)]^{-2(1/2+\delta)} [c \lambda_{\max}(M_n)]^{2(1/2+\delta)} \varepsilon^2 = c^2 \varepsilon^2, \quad \forall n \geq n_1, \end{aligned}$$

Hence with probability 1,  $\forall \varepsilon \in (0, \varepsilon_0)$

$$\inf_{\beta \in \partial B_\varepsilon(\beta_0)} \|T_n(\beta) - T_n(\beta_0)\| \geq c\varepsilon, \quad \forall n \geq n_1. \quad (104)$$

From Lemma 4.2.9, it follows that with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$ , there exists a random number  $n_2 = n_2(\varepsilon) > n_1$  such that

$$\|T_n(\beta_0)\| = \|g_n\| / [\lambda_{\max}(M_n)]^{1/2+\delta} \leq c\varepsilon, \quad \forall n \geq n_2. \quad (105)$$

From (104) and (105) we obtain (103).

□

**Conclusion** The following table summarizes the sets of conditions needed for the asymptotic results presented in this section.

Results	Conditions
Weak Consistency	( $\mathbf{I}_w$ ) and ( $\mathbf{L}_w$ ) ( $\mathbf{I}_w^*$ ) and ( $\mathbf{L}_w^*$ ) ( $\mathbf{I}_w^+$ ) and ( $\mathbf{L}_w^+$ )
Asymptotic Normality	( $\mathbf{I}_w$ ), ( $\mathbf{CC}$ ) and ( $\mathbf{N}_\delta$ ) ( $\mathbf{I}_w^*$ ), ( $\mathbf{CC}$ ) and ( $\mathbf{N}_\delta$ )
Strong Consistency	( $\mathbf{I}_s$ ) <sup>SP</sup> and ( $\mathbf{L}_s$ ) <sup>SP</sup>

In Section 4.4 the conditions in the table will be examined for some particular cases and examples.

### 4.3 Verification of Condition (CC)

In order to check the above mentioned conditions in practice, we need to impose the following assumption: for any  $r > 0$

$$(\mathbf{AH}) \quad k_n^{[1]} = \sup_{\beta \in N_n^*(r)} \max_{i \leq n} \max_{j \leq m} \left| \frac{\mu''(x_{ij}^T \beta)}{\mu'(x_{ij}^T \beta)} \right| \leq c, \quad k_n^{[2]} = \sup_{\beta \in N_n^*(r)} \max_{i \leq n} \max_{j \leq m} \left| \frac{\mu'''(x_{ij}^T \beta)}{\mu'(x_{ij}^T \beta)} \right| \leq c,$$

where  $c$  is a constant. We define new sequences of constants:

$$\gamma_n^{(0)} = \max_{i \leq n} \max_{j \leq m} (x_{ij}^T H_n^{-1} x_{ij}), \quad \gamma_n^* := \tau_n \gamma_n^{(0)}, \quad \pi_n = \frac{\max_{i \leq n} \{\lambda_{\max}(R_i^{-1}(\alpha))\}}{\min_{i \leq n} \{\lambda_{\min}(R_i^{-1}(\alpha))\}} := \frac{\tilde{\lambda}_n}{\lambda_n},$$

$$\nu_n^* = \min\left\{\sqrt{mn}, \frac{m\pi_n}{b_n}\right\}, \quad \text{where } b_n = \min_{i \leq n} \min_{j \leq m} \sigma_{ij}^2.$$

We introduce the following conditions:

$$(\mathbf{CC}_1) \quad \pi_n^2 \gamma_n^* \rightarrow 0, \quad (\mathbf{CC}_2) \quad \nu_n^* \pi_n \gamma_n^* \rightarrow 0, \quad (\mathbf{CC}_3) \quad \pi_n \gamma_n^* \rightarrow 0.$$

Note that ( $\mathbf{CC}_1$ ) implies ( $\mathbf{CC}_3$ ), since  $\pi_n \geq 1$ . The proof of the following lemma is quite technical and is omitted (see Appendix C, p. 338 of [20] )

**Lemma 4.3.1 (Lemma A.1-Lemma A.3, p. 338, [20])** *Suppose that (AH) holds.*

(a) *If (CC<sub>3</sub>) holds, then  $\sup_{\beta \in N_n^*(r)} |x^T H_n^{-1/2} B_n(\beta) H_n^{-T/2} x - 1| \rightarrow 0$ , for any  $x$  with  $\|x\|=1$ .*

(b) *If (CC<sub>1</sub>) holds, then  $\sup_{\beta \in N_n^*(r)} (x^T H_n^{-1/2} H_n(\beta) H_n^{-T/2} x) \rightarrow 0$ , for any  $x$  with  $\|x\|=1$ .*

(c) *If (CC<sub>2</sub>) holds, then  $\sup_{\beta \in N_n^*(r)} (x^T H_n^{-1/2} \mathcal{E}(\beta) H_n^{-T/2} x) \rightarrow 0$  in probability, for any  $x$  with  $\|x\|=1$ .*

**Theorem 4.3.2 (Theorem A.2, p. 339 [20])** *Assume that (AH), (CC<sub>1</sub>) and (CC<sub>2</sub>) hold. Then (CC) holds, and in particular (L<sub>w</sub><sup>\*</sup>) holds.*

**Proof:** This follows from Lemma 4.1.6. and Lemma 4.3.1.

□

## 4.4 Particular Cases and Examples

In this section, we present some particular cases and examples.

### 4.4.1 Bounded Regressors

In this subsection, we suppose that the covariates  $\{x_{ij}\}_{i,j}$  satisfy

$$\sup_{i \geq 1} \sup_{j \leq m} \|x_{ij}\| \leq c < \infty. \quad (106)$$

In this case, we can show that assumption (AH) holds and condition (CC<sub>2</sub>) follows from (CC<sub>1</sub>), whereas condition (N<sub>δ</sub>) is implied by

$$(\mathbf{B}_\delta) \quad c_n^{1+\delta} \tilde{\lambda}_n^{2+\delta} \gamma_n^{(0)} \rightarrow 0.$$

More precisely, as a consequence of Theorem 4.2.8 and Theorem 4.3.2 we have the following result.

**Corollary 4.4.1 (Corollary 1, p. 331, [20])** *Suppose that the covariates  $\{x_{ij}\}_{i,j}$  satisfy (106), and (I<sub>w</sub><sup>\*</sup>), (CC<sub>1</sub>) hold. Then there exist a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that  $\hat{\beta}_n \rightarrow \beta_0$  in probability. If in addition, we suppose (B<sub>δ</sub>) holds, then  $M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \rightarrow N(0, I)$  in distribution.*

**Proof:** (p. 332, [20]) Note that  $|x_{ij}^T \beta_0| \leq \|x_{ij}\| \|\beta_0\| \leq c \|\beta_0\|$ , for all  $i \geq 1$ ,  $j \leq m$ . Since a continuous function maps a compact set into a compact set, we get

$$0 < c_1 \leq \mu'(x_{ij}^T \beta) < c_2, \quad 0 < c'_1 \leq \mu''(x_{ij}^T \beta) < c'_2, \quad 0 < c_1^* \leq \mu'''(x_{ij}^T \beta) < c_2^*. \quad (107)$$

Hence assumption **(AH)** is satisfied.

Note that in this case **(CC<sub>1</sub>)** implies **(CC<sub>2</sub>)** :

$$\nu_n^* \pi_n \gamma_n^* \leq \frac{m \pi_n}{b_n} (\pi_n \gamma_n^*) \leq \left(\frac{1}{c_1} m \pi_n\right) \pi_n \gamma_n^* \longrightarrow 0,$$

since  $b_n \geq c_1 > 0$ .

From Theorem 4.3.2, condition **(CC)** holds. From Theorem 4.2.3, we get  $\hat{\beta}_n \longrightarrow \beta_0$  in probability. In order to prove the asymptotic normality of  $\hat{\beta}_n$ , we will use Theorem 4.2.8.

To see that **(N<sub>δ</sub>) (i)** holds, note that since  $E((y_{ij}^*)^{2(1+1/\delta)})$  is continuous function of  $x_{ij}$  and  $x_{ij}$  lies in a compact set, there exists a constant  $0 < K_1 < \infty$  such that  $E((y_{ij}^*)^{2(1+1/\delta)}) \leq K_1$ .

We claim that

$$\mathbf{(B}_\delta) \text{ implies } \mathbf{(N}_\delta) \text{ (ii)}. \quad (108)$$

To see this, recall that

$$\begin{aligned} \gamma_n^{(D)} &= \max_{1 \leq i \leq n} \lambda_{\max}(H_n^{-1/2} D_i^T V_i(\alpha)^{-1} D_i H_n^{-1/2}) = \\ &= \max_{1 \leq i \leq n} \lambda_{\max}(H_n^{-1/2} X_i^T A_i^{1/2}(\beta) R_i(\alpha)^{-1} A_i^{1/2}(\beta) X_i H_n^{-1/2}). \end{aligned}$$

Note that if  $x$  is a  $p$ -dimensional vector of norm 1, then for each  $i \leq n$ ,

$$\begin{aligned} x^T H_n^{-1/2} X_i^T A_i^{1/2} R_i^{-1} A_i^{1/2} X_i H_n^{-1/2} x &\leq \lambda_{\max}(R_i^{-1}) x^T H_n^{-1/2} X_i^T A_i X_i H_n^{-1/2} x \leq \\ &\leq \lambda_{\max}(R_i^{-1}) \lambda_{\max}(A_i) x^T H_n^{-1/2} X_i^T X_i H_n^{-1/2} x = \lambda_{\max}(R_i^{-1}) \lambda_{\max}(A_i) \|X_i H_n^{-1/2} x\|^2 \leq \\ &\leq \lambda_{\max}(R_i^{-1}) \lambda_{\max}(A_i) \|X_i H_n^{-1/2}\|_E^2 = \lambda_{\max}(R_i^{-1}) \lambda_{\max}(A_i) \text{tr}(X_i H_n^{-1} X_i^T) \leq \\ &\leq \tilde{\lambda}_n c_2 m \max_{j \leq m} (x_{ij}^T H_n^{-1} x_{ij}), \end{aligned}$$

where  $\lambda_{\max}(A_i) = \max_{j \leq m} \mu'(x_{ij}^T \beta_0) \leq c_2$ . Taking maximum over  $i \leq n$ , we get  $\gamma_n^{(D)} \leq c_2(m\tilde{\lambda}_n)\gamma_n^{(0)}$ . Hence

$$(c_n \tilde{\lambda}_n m)^{1+\delta} \gamma_n^{(D)} \leq (c_n^{1+\delta})(\tilde{\lambda}_n m)^{2+\delta} \gamma_n^{(0)}.$$

This concludes the proof of (108).

□

Our next goal is to show that the conditions of the previous corollary can be checked directly using only the working correlated matrices  $R_i(\alpha)$  and the design matrices  $X_i$ . For this purpose, we consider the following two conditions:

$$(\mathbf{B}) \frac{\pi_n^3}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} \longrightarrow 0, \quad (\mathbf{B}_\delta^*) c_n^{1+\delta} \tilde{\lambda}_n^{1+\delta} \frac{\pi_n}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} \longrightarrow 0.$$

**Proposition 4.4.2 (Remark 13, p. 332, [20])** *Suppose that the covariates  $(x_{ij})_{i,j}$  satisfy (106). Then  $(\mathbf{B})$  implies  $(\mathbf{I}_w^*)$  and  $(\mathbf{CC}_1)$ , and  $(\mathbf{B}_\delta^*)$  implies  $(\mathbf{B}_\delta)$ .*

**Proof:** From (107), we get that  $A_i \geq \lambda_{\min}(A_i)I = \min_{j \leq m} \mu'(x_{ij}^T \beta_0)I \geq c_1 I$ . Hence

$$R_i(\alpha) \leq \lambda_{\max}(R_i(\alpha))I \leq \frac{1}{c_1} \lambda_{\max}(R_i(\alpha))A_i$$

and

$$R_i^{-1}(\alpha) \geq c_1 \lambda_{\min}(R_i^{-1}(\alpha))A_i^{-1} \geq c_1 \bar{\lambda}_n A_i^{-1}, \forall i \leq n.$$

It follows that

$$H_n - (c_1 \bar{\lambda}_n) \sum_{i=1}^n X_i^T X_i = \sum_{i=1}^n X_i^T A_i^{1/2} (R_i^{-1}(\alpha) - c_1 \bar{\lambda}_n A_i^{-1}) A_i^{1/2} X_i \geq 0$$

and

$$\lambda_{\min}(H_n) \geq c_1 \bar{\lambda}_n \lambda_{\min}\left(\sum_{i=1}^n X_i^T X_i\right). \quad (109)$$

By condition (106), there exists a constant  $K > 0$  such that  $\|X_i\|^2 \leq K$  for all  $i$ . We have

$$\max_{j \leq m} (x_{ij}^T H_n^{-1} x_{ij}) \leq \lambda_{\max}(X_i H_n^{-1} X_i^T) = \lambda_{\max}(X_i H_n^{-1/2} H_n^{-1/2} X_i^T) =$$

$$= \|H_n^{-1/2} X_i^T\|^2 \leq \|H_n^{-1/2}\|^2 \|X_i\|^2 \leq K \|H_n^{-1}\| = \frac{K}{\lambda_{\min}(H_n)}.$$

Taking maximum over  $i \leq n$  and using (109), we obtain

$$\gamma_n^{(0)} \leq \frac{K}{\lambda_{\min}(H_n)} \leq \frac{K}{c_1 \bar{\lambda}_n \lambda_{\min}(\sum_{i=1}^n X_i^T X_i)}. \quad (110)$$

To prove that **(B)** implies **(I<sub>w</sub><sup>\*</sup>)**, recall that  $\tau_n = \max_{i \leq n} \lambda_{\max}(R_i(\alpha)^{-1} \bar{R}_i) \leq m \tilde{\lambda}_n$ . Using (109), we obtain

$$\frac{\tau_n}{\lambda_{\min}(H_n)} \leq \frac{m \tilde{\lambda}_n}{c_1 \bar{\lambda}_n \lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} = \frac{1}{c_1} \frac{m \pi_n}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)}.$$

To prove that **(B)** implies **(CC<sub>1</sub>)**, we use (110) and the fact that  $\tau_n \leq m \tilde{\lambda}_n$ :

$$\begin{aligned} \pi_n^2 \gamma_n^* &= \pi_n^2 \tau_n \gamma_n^{(0)} \leq \frac{K}{c_1} \frac{\pi_n^2}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} \frac{m \tilde{\lambda}_n}{\bar{\lambda}_n} \frac{\tau_n}{m \tilde{\lambda}_n} = \\ &= \frac{K}{c_1} \frac{\pi_n^2}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} m \pi_n \frac{\tau_n}{m \tilde{\lambda}_n} \leq \frac{K}{c_1} \frac{m \pi_n^3}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)}. \end{aligned}$$

To prove **(B<sub>δ</sub><sup>\*</sup>)** implies **(B<sub>δ</sub>)**, we use (110):

$$c_n^{1+\delta} (\tilde{\lambda}_n)^{2+\delta} \gamma_n^{(0)} \leq \frac{K}{c_1} \frac{c_n^{1+\delta} (\tilde{\lambda}_n)^{1+\delta}}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} \frac{\tilde{\lambda}_n}{\bar{\lambda}_n} = \frac{K}{c_1} \frac{c_n^{1+\delta} (\tilde{\lambda}_n)^{1+\delta}}{\lambda_{\min}(\sum_{i=1}^n X_i^T X_i)} \pi_n.$$

□

By combining Proposition 4.4.2 and Corollary 4.4.1, we obtain the following result.

**Corollary 4.4.3** *Suppose that the covariates  $(x_{ij})_{i,j}$  satisfy (106), and conditions **(B)** and **(B<sub>δ</sub><sup>\*</sup>)** hold. Then the same conclusion as in Corollary 4.4.1 holds.*

### 4.4.2 Bounded responses

We suppose that the responses  $(y_{ij})_{ij}$  satisfy

$$\sup_{i \geq 1} \sup_{j \leq m} |y_{ij}| \leq c < \infty. \quad (111)$$

Recall that we have assumed that  $y_{ij}$  has a density function given by formula (77) (with  $\phi = 1$ ). This implies that for any  $\beta \in B$ , and for any  $i \geq 1$ ,  $j = 1, \dots, m$ , we have:

$$|\mu_{ij}(\beta)| = |E_{\beta}(y_{ij})| \leq E_{\beta}|y_{ij}| \leq c, \quad \mu'(x_{ij}^T \beta) = E_{\beta}(y_{ij} - \mu_{ij})^2 \leq (2c)^2,$$

$$|\mu''(x_{ij}^T \beta)| = |E(y_{ij} - \mu_{ij})^3| \leq (2c)^3,$$

$$|\mu'''(x_{ij}^T \beta)| = |E(y_{ij} - \mu_{ij})^4 - 3[\mu'(x_{ij}^T \beta)]^2| \leq (2c)^4 + 3(2c)^4.$$

Therefore under (111), **(AH')** implies **(AH)**, where

$$\mathbf{(AH')} \quad \inf_{\beta \in N_n^*(r)} \inf_{i \leq n} \inf_{j \leq m} \mu'(x_{ij}^T \beta) > 0.$$

Consider conditions

$$\mathbf{(CC'_2)} \quad \frac{\pi_n^2 \gamma_n^*}{b_n} \rightarrow 0, \quad \mathbf{(N_0)} \quad \frac{c_n \tilde{\lambda}_n^2 \gamma_n^{(0)}}{b_n} \rightarrow 0.$$

**Corollary 4.4.4 (Corollary 3, p. 334, [20])** *Suppose that the responses  $y_{ij}$ s satisfy (111), **(AH')**, **(I\_w^\*)** and **(CC'\_2)** hold. Then there exist a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that  $\hat{\beta}_n \rightarrow \beta_0$  in probability. If in addition, we suppose **(N\_0)** holds, then  $M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \rightarrow N(0, I)$  in distribution, when  $n \rightarrow \infty$ .*

**Proof:** (p. 334, [20]) Note that  $b_n = \min_{i \leq n} \min_{j \leq m} \mu'(x_{ij}^T \beta) \leq (2c)^2$ . Therefore **(CC'\_2)** implies **(CC\_1)**, since

$$\pi_n^2 \gamma_n^* = \frac{\pi_n^2 \gamma_n^*}{b_n} b_n \leq (2c)^2 \frac{\pi_n^2 \gamma_n^*}{b_n}.$$

Also **(CC'\_2)** implies **(CC\_2)**, since

$$\nu_n^* \pi_n \gamma_n^* \leq \frac{m \pi_n}{b_n} \pi_n \gamma_n^* = m \frac{\pi_n^2 \gamma_n^*}{b_n}.$$

From Theorem 4.3.2, condition **(CC)** holds. From Theorem 4.2.3, we get  $\hat{\beta}_n \rightarrow \beta_0$  in probability. As in the proof of Theorem 4.2.8, in order to show that  $M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \rightarrow N(0, I)$  it suffices to prove that

$$M_n^{-1/2} g_n \rightarrow N(0, I) \text{ in distribution.}$$

By the Cràmer-Wold theorem (Theorem A.2.4, Appendix A.2), this is equivalent to saying that

$$x^T M_n^{-1/2} g_n \rightarrow N(0, 1) \text{ in distribution,}$$

for any  $p$ -dimensional vector  $x$  with  $\|x\| = 1$ . Recall that  $g_n = \sum_{i=1}^n D_i^T V_i^{-1}(\alpha) \varepsilon_i$ . Hence

$$x^T M_n^{-1/2} g_n = \sum_{i=1}^n x^T M_n^{-1/2} D_i^T V_i^{-1}(\alpha) \varepsilon_i := \sum_{i=1}^n Z_{ni}.$$

We need to verify the Lindeberg condition

$$\sum_{i=1}^n E(z_{ni}^2 I_{\{|z_{ni}| \geq \varepsilon\}}) \rightarrow 0, \quad \forall \varepsilon > 0.$$

We know that  $Z_{ni}^2 \leq c_n \tilde{\lambda}_n \gamma_n^{(D)} (y_i^*)^T y_i^*$ , where  $y_i^* = A_i^{-1/2} (y_i - \mu_i)$ . (See the proof of Lemma 4.2.7)

We claim that there exists a constant  $K > 0$  such that  $\gamma_{ni}^{(D)} \leq K m \tilde{\lambda}_n \gamma_n^{(0)}$ . To see this, note that

$$\begin{aligned} \gamma_{ni}^{(D)} &= x^T H_n^{-1/2} X_i^T A_i^{1/2} R_i(\alpha)^{-1} A_i^{1/2} X_i H_n^{-1/2} x = \|R_i^{-1/2} A_i^{1/2} X_i H_n^{-1/2} x\|^2 \leq \\ &\leq \|R_i^{-1}\| \|A_i^{1/2}\|^2 \|X_i H_n^{-1/2}\|^2 \|x\|^2 \leq (2c)^2 (m \gamma_n^{(0)}) \tilde{\lambda}_n, \end{aligned}$$

since  $\|X_i H_n^{-1/2}\|^2 \leq m \gamma_n^{(0)}$ ,  $\|x\| = 1$  and  $\|A_i\| = \lambda_{\max}(A_i) = \max_{j \leq m} \mu'(x_{ij}^T \beta_0) \leq (2c)^2$ . Hence

$$Z_{ni}^2 \leq K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)} (y_i^*)^T y_i^*, \quad \forall i \leq n \quad (112)$$

and

$$I_{\{|Z_{ni}| \geq \varepsilon\}} = I_{\{Z_{ni}^2 \geq \varepsilon^2\}} \leq I_{\{(y_i^*)^T y_i^* \geq \varepsilon^2 / (K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)})\}}.$$

We claim that there exists  $n_0 = n_0(\varepsilon)$  such that for any  $n \geq n_0$  and  $i = 1, \dots, n$ ,

$$\{(y_i^*)^T y_i^* \geq \frac{\varepsilon^2}{K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)}}\} = \emptyset, \quad \forall n \geq n_0. \quad (113)$$

To see this, note that  $\lambda_{\min}(A_i) = \min_{j \leq m} \mu'(x_{ij}^T \beta_0) I \geq b_n I$  and under (111) there exists a constant  $K_1$  such that  $\|y_i - \mu_i\|^2 \leq K_1^2$ . By  $(\mathbf{N}_0)$ , there exists  $n_0$  such that

$$\frac{c_n \tilde{\lambda}_n^2 \gamma_n^{(0)}}{b_n} \leq \frac{\varepsilon^2}{m K K_1^2}, \quad \forall n \geq n_0.$$

Hence

$$(y_i^*)^T y_i^* = (y_i - \mu_i)^T A_i^{-1} (y_i - \mu_i) \leq \frac{1}{b_n} \|y_i - \mu_i\|^2 \leq K_1^2 \frac{1}{b_n} \leq \frac{\varepsilon^2}{K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)}},$$

which concludes the proof of (113). From (112) and (113), we conclude that

$$\begin{aligned} \sum_{i=1}^n E [z_{ni}^2 I_{\{|z_{ni}| \geq \varepsilon\}}] &\leq K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)} \sum_{i=1}^n E \left[ (y_i^*)^2 I_{\{(y_i^*)^T y_i^* \geq \varepsilon^2 / (K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)})\}} \right] \\ &= K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)} \sum_{i=1}^{n_0} E \left[ (y_i^*)^2 I_{\{(y_i^*)^T y_i^* \geq \varepsilon^2 / (K c_n \tilde{\lambda}_n^2 m \gamma_n^{(0)})\}} \right]. \end{aligned}$$

The conclusion follows since  $(\mathbf{N}_0)$  and the fact that  $b_n \leq (2c)^2$  imply that  $c_n \tilde{\lambda}_n^2 \gamma_n^{(0)} \rightarrow 0$ .

□

### 4.4.3 Logistic Regression for Binary Data

In this subsection we assume that the responses  $y_{ij}$  assume only the values 0 and 1, say

$$P(y_{ij} = 1) = p_{ij} \text{ and } P(y_{ij} = 0) = 1 - p_{ij}.$$

Then density function of  $y_{ij}$  is

$$\begin{aligned} f(y_{ij} | p_{ij}) &= p_{ij}^{y_{ij}} (1 - p_{ij})^{1 - y_{ij}} = \exp \{ y_{ij} \ln p_{ij} + [(1 - y_{ij}) \ln(1 - p_{ij})] \} = \\ &= \exp \left[ y_{ij} \ln \frac{p_{ij}}{1 - p_{ij}} + \ln(1 - p_{ij}) \right] = \exp [y_{ij} \text{logit}(p_{ij}) + \ln(1 - p_{ij})]. \end{aligned}$$

From here we see that the natural link function is:

$$\text{logit}(p_{ij}) = \theta_{ij} = x_{ij}^T \beta.$$

Therefore, we obtain that  $f(y_{ij}|\theta_{ij}) = c(y_{ij}) \exp[\theta_{ij}y_{ij} - a(\theta_{ij})]$ . Note that

$$c(y_{ij}) = 1, \quad a(\theta_{ij}) = \ln(1 + e^{\theta_{ij}}), \quad \mu(\theta_{ij}) = E(y_{ij}) = p_{ij} = \frac{e^{\theta_{ij}}}{1 + e^{\theta_{ij}}}.$$

In what follows, we will check the assumption **(AH)**. Note that

$$\mu'(\theta) = \frac{e^\theta}{(1 + e^\theta)^2}, \quad \mu''(\theta) = -\frac{e^\theta(1 - e^\theta)}{(1 + e^\theta)^3}, \quad \mu'''(\theta) = -\frac{e^{3\theta} - 4e^{2\theta} + e^\theta}{(1 + e^\theta)^4},$$

and hence

$$\left| \frac{\mu''(\theta)}{\mu'(\theta)} \right| = \left| \frac{1 - e^\theta}{1 + e^\theta} \right| = \left| 1 - \frac{2e^\theta}{1 + e^\theta} \right| \leq 1 + 2 \left| \frac{e^\theta}{1 + e^\theta} \right| \leq 3,$$

$$\left| \frac{\mu'''(\theta)}{\mu'(\theta)} \right| = \left| \frac{e^{2\theta} - 4e^\theta + 1}{(1 + e^\theta)^2} \right| = \left| 1 - \frac{6e^\theta}{1 + e^\theta} \cdot \frac{1}{1 + e^\theta} \right| \leq 1 + \left| \frac{6e^\theta}{1 + e^\theta} \right| \cdot \left| \frac{1}{1 + e^\theta} \right| \leq 7.$$

From here we conclude that

$$k_n^{[1]} = \sup_{\beta \in N_n^*(r)} \max_{i \leq n} \max_{j \leq m} \left| \frac{\mu''(x_{ij}^T \beta)}{\mu'(x_{ij}^T \beta)} \right| \leq 3$$

and

$$k_n^{[2]} = \sup_{\beta \in N_n^*(\delta)} \max_{i \leq n} \max_{j \leq m} \left| \frac{\mu'''(x_{ij}^T \beta)}{\mu'(x_{ij}^T \beta)} \right| \leq 7.$$

Since  $y_{ij} \in \{0, 1\}$ , this situation belongs to the Bounded Responses Case. From Corollary 4.4.4, we obtain that if **(I $_{\mathbf{w}}^*$ )** and **(CC $_2'$ )** hold, then there exist a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that  $\hat{\beta}_n \rightarrow \beta_0$  in probability. If in addition, we suppose **(N $_0$ )** holds, then  $M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \rightarrow N(0, I)$  in distribution.

In practice, for a given longitudinal data set, we can find the GEE estimator and build a logistic regression model for binary data using the SAS system. (see Appendix B.1)

#### 4.4.4 Linear Regression

In this subsection, we suppose that for all  $i \geq 1, j = 1, \dots, m,$

$$E_\beta(y_{ij}) = x_{ij}^T \beta \quad \text{and} \quad \text{Var}_\beta(y_{ij}) = 1. \quad (114)$$

This situation may be encountered for instance if  $y_{ij}$  has a normal distribution with mean  $\theta_{ij} = x_{ij}^T \beta$  and variance 1, i.e. the density of  $y_{ij}$  is

$$f(y_{ij}|\theta_{ij}) = c(y_{ij}) \exp\left(y_{ij}\theta_{ij} - \frac{\theta_{ij}^2}{2}\right).$$

In this case, the theory simplifies considerably, since

$$A_i(\beta) = I, \quad D_i(\beta) = A_i(\beta)X_i = X_i, \quad V_i(\beta, \alpha) = A_i(\beta)^{1/2}R_i(\alpha)A_i(\beta)^{1/2} = R_i(\alpha),$$

$$g_n(\beta) = \sum_{i=1}^n X_i^T R_i(\alpha)^{-1}(y_i - \mu_i(\beta)) = \sum_{i=1}^n X_i^T R_i(\alpha)^{-1}(y_i - X_i\beta),$$

$$\mathcal{D}_n(\beta) = -\frac{\partial g_n(\beta)}{\partial \beta} = \sum_{i=1}^n X_i^T R_i(\alpha)^{-1}X_i = H_n(\beta),$$

$$M_n = \sum_{i=1}^n X_i^T R_i(\alpha)^{-1} \bar{R}_i R_i(\alpha)^{-1} X_i.$$

Since  $\mathcal{D}_n(\beta) = H_n(\beta)$ , conditions  $(\mathbf{L}_w)$ ,  $(\mathbf{L}_w^*)$  and  $(\mathbf{CC})$  hold, whereas condition  $(\mathbf{N}_\delta)$  is implied by

$$(\mathbf{L}_\delta) \quad (c_n)^{1+\delta} \tilde{\lambda}_n^{2+\delta} \max_{1 \leq i \leq n} \lambda_{\max}(X_i H_n^{-1} X_i^T) \rightarrow 0.$$

More precisely, we have the following result as a consequence of Theorem 4.2.8 and Theorem 4.2.3.

**Corollary 4.4.5 (Example 5.1, p. 329, [20])** *Suppose that the responses  $y_{ij}$  satisfy (114) and  $(\mathbf{I}_w^*)$  holds. Then there exist a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that  $\hat{\beta}_n \rightarrow \beta_0$  in probability. If in addition, we suppose  $(\mathbf{L}_\delta)$  holds, then  $M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \rightarrow N(0, I)$  in distribution, when  $n \rightarrow \infty$ .*

**Proof:** From Theorem 4.2.3, we get  $\hat{\beta}_n \rightarrow \beta_0$  in probability.

In order to prove the asymptotic normality of  $\hat{\beta}_n$ , we will use Theorem 4.2.8. We claim that

$$(\mathbf{L}_\delta) \text{ implies } (\mathbf{N}_\delta). \quad (115)$$

To see this, recall that

$$\begin{aligned} \gamma_n^{(D)} &= \max_{1 \leq i \leq n} \lambda_{\max}(H_n^{-1/2} D_i^T(\beta) V_i(\beta, \alpha)^{-1} D_i(\beta) H_n^{-T/2}) \\ &= \max_{1 \leq i \leq n} \lambda_{\max}(H_n^{-1/2} X_i^T R_i(\alpha)^{-1} X_i H_n^{-T/2}). \end{aligned}$$

Note that for each  $i \leq n$ ,

$$\begin{aligned} x^T H_n^{-1/2} X_i^T R_i^{-1}(\alpha) X_i H_n^{-T/2} x &\leq \lambda_{\max}[R_i^{-1}(\alpha)] x^T H_n^{-1/2} X_i^T X_i H_n^{-T/2} x \leq \\ &\leq \lambda_{\max}[R_i^{-1}(\alpha)] x^T H_n^{-1/2} X_i^T X_i H_n^{-T/2} x = \lambda_{\max}[R_i^{-1}(\alpha)] \|X_i H_n^{-1/2} x\|^2 = \\ &= \lambda_{\max}[R_i^{-1}(\alpha)] \lambda_{\max}(X_i H_n^{-1} X_i^T) \leq \tilde{\lambda}_n \lambda_{\max}(X_i H_n^{-1} X_i^T). \end{aligned}$$

Taking the maximum over  $i \leq n$ , we get  $\gamma_n^{(D)} \leq \tilde{\lambda}_n \cdot \max_{i \leq n} \lambda_{\max}(X_i H_n^{-1} X_i^T)$ . Hence

$$(c_n \tilde{\lambda}_n)^{1+\delta} \gamma_n^{(D)} \leq (c_n)^{1+\delta} \tilde{\lambda}_n^{2+\delta} \max_{1 \leq i \leq n} \lambda_{\max}(X_i H_n^{-1} X_i),$$

which concludes the proof of (115).

□

#### 4.4.5 Poisson Log-linear Regression

In this subsection, we suppose that  $y_{ij}$  is a count measurement with the following Poisson density:

$$f(y_{ij} | \lambda_{ij}) = e^{-\lambda_{ij}} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} = (1/y_{ij}!) \exp(y_{ij} \ln \lambda_{ij} - \lambda_{ij})$$

From here we see that the natural link function for this model is

$$\ln \lambda_{ij} = \theta_{ij} = x_{ij}^T \beta.$$

Hence

$$\mu_{ij}(\beta) = \sigma_{ij}^2(\beta) = e^{x_{ij}^T \beta}.$$

In this case  $\mu(\theta) = e^\theta$  and assumption **(AH)** is clearly satisfied.

We introduce the following assumption:

**(P)** There exist constants  $c_1 > 0$  and  $c_2 > 0$  such that

$$\min_{i \leq n} \min_{j \leq m} e^{x_{ij}^T \beta_0} \geq c_1, \quad \forall n, \quad \max_{i \leq n} \max_{j \leq m} e^{x_{ij}^T \beta_0} \leq c_2, \quad \forall n$$

**Corollary 4.4.6 (Corollary 2, p. 332, [20])** *Suppose that **(P)**, **(I<sub>w</sub><sup>\*</sup>)** and **(CC<sub>1</sub>)** hold. Then there exist a sequence  $\{\hat{\beta}_n\}_{n \geq 1}$  of random variables such that  $\hat{\beta}_n \rightarrow \beta_0$  in probability. If in addition, we suppose that **(B<sub>δ</sub>)** holds, then  $M_n^{-1/2} H_n(\hat{\beta}_n - \beta_0) \rightarrow N(0, I)$  in distribution, when  $n \rightarrow \infty$ .*

**Proof:** (p. 333, [20]) Under **(P)**, we have  $b_n \geq c_1$ , and hence

$$\nu_n^* \leq \frac{m\pi_n}{b_n} \leq \frac{m}{c_1}\pi_n, \quad \nu_n^*\pi_n\gamma_n^* \leq \frac{m}{c_1}\pi_n^2\gamma_n^*.$$

It follows that **(CC<sub>1</sub>)** implies **(CC<sub>2</sub>)**. From Theorem 4.3.2, condition **(CC)** holds. From Theorem 4.2.3, we get  $\hat{\beta}_n \rightarrow \beta_0$  in probability.

In order to prove the asymptotic normality of  $\hat{\beta}_n$ , we will use Theorem 4.2.8. We claim that **(B<sub>δ</sub>)** implies **(N<sub>δ</sub>)**. First, using a property of the Poisson distribution and condition **(P)**, we conclude that there exist some constants  $c'_1 > 0$  and  $c'_2 > 0$  (depending on  $\delta$ ) such that

$$E|y_{ij}^*|^{2+2/\delta} = E \left| \frac{y_{ij} - \mu_{ij}(\beta_0)}{\sqrt{\mu_{ij}(\beta_0)}} \right|^{2+\delta/2} \leq c'_1 + c'_2 \{\mu_{ij}(\beta_0)\}^{3+2/\delta} \leq c'_1 + c'_2 c_2.$$

Note that under **(P)**, we have  $\gamma_n^{(D)} \leq c_2 m \tilde{\lambda}_n \gamma_n^{(0)}$ . (To see this, we use the same argument as in the proof of Corollary 4.4.1, since  $\lambda_{\max}(A_i) \leq c_2$ , for all  $i$ , due to condition **(P)**). Therefore

$$(c_n \tilde{\lambda}_n m)^{1+\delta} \gamma_n^{(D)} \leq (c_n^{1+\delta}) (\tilde{\lambda}_n m)^{2+\delta} \gamma_n^{(0)},$$

and **(N<sub>δ</sub>)** follows from **(B<sub>δ</sub>)**.

□

In practice, for a given longitudinal data set, we can find the GEE estimator and build a Poisson log-linear model for binary data using the SAS system. (see Appendix B.2)

# Chapter 5

## The Quasi-likelihood Approach

In this chapter, we present the asymptotic theory for a GLM in which the density function of the response variables  $y_i$  is not specified. As in Chapter 3, we suppose that  $(y_i)_{i \geq 1}$  are random variables defined on probability space  $(\Omega, \mathcal{F}, P_\beta)$ , which are recorded together with the  $p$ -dimensional regressors  $(x_i)_{i \geq 1}$  such that

$$E_\beta(y_i) = \mu(x_i^T \beta) \text{ and } Var_\beta(y_i) = \phi \mu'(x_i^T \beta), \quad \forall i \geq 1. \quad (116)$$

The solution  $\hat{\beta}_n$  of the equation

$$s_n(\beta) := \sum_{i=1}^n x_i (y_i - \mu(x_i^T \beta)) = 0, \quad (117)$$

is called the **maximum quasi-likelihood estimator** (MQLE) of  $\beta$ . The question of its existence and strong consistency has been examined by Chen, Hu and Ying in [4]. However, we discovered that the proof of the main result in the above-mentioned article relies on some incorrect arguments, and therefore we believe that this result does not hold true.

This chapter is organized as follows. In Section 5.1, we present the general argument of Chen, Hu, and Ying (used for the proof of Theorem 1 of [4]), and the reasons we believe it is incorrect. In Section 5.2, we apply the theory developed in Chapter 4, to provide the asymptotic results for a GLM specified by (116). In particular, our Theorem 5.2, yields a correction to Theorem 1 of [4].

## 5.1 The Argument of Chen, Hu and Ying

The main result of [4] states that if the regressors are bounded, the existence and strong consistency of the MQLE  $\hat{\beta}_n$  can be obtained under virtually the same conditions as in the linear case, and moreover, the rate of convergence of  $\hat{\beta}_n$  to  $\beta_0$  is exactly the same as the rate of convergence of the LSE  $\hat{\beta}_n^{(LS)}$  to  $\beta_0$ .

We begin to investigate this result.

Using the notation of Chapter 2 and Chapter 3, we let  $H_n = \sum_{i=1}^n x_i x_i^T$  be the design matrix,  $V_n = H_n^{-1} = (v_{ij}^{(n)})_{i,j=1,\dots,p}$  and  $F_n(\beta) = \sum_{i=1}^n \mu'(x_i^T \beta) x_i x_i^T$ . We suppose that the matrix  $H_n$  satisfies the divergence condition  $(\mathbf{D}^*)$  and  $\sup_{i \geq 1} \|x_i\| < \infty$ .

We define the **errors**

$$\varepsilon_i(\beta) = y_i - \mu(x_i^T \beta), \quad i \geq 1$$

and we suppose that the residuals  $(\varepsilon_i)_{i \geq 1}$  satisfy condition  $(\mathbf{C}_1)$  given in Chapter 2. Let us also consider the **linear-errors**

$$\varepsilon_i^{(LS)}(\beta) = y_i - x_i^T \beta, \quad i \geq 1.$$

The first error of the authors of [4] is the fact that they do **not** make a distinction between  $\varepsilon_i$  and  $\varepsilon_i^{(LS)}$ , as it can be seen from their relations (2.7) and (2.9) on page 1159.

It is clearly true that

$$\begin{aligned} a_n^{(LS)} &:= H_n^{-1} \sum_{i=1}^n x_i \varepsilon_i^{(LS)} = \hat{\beta}_n^{(LS)} - \beta_0 = o\left(\max_{j \leq p} \{v_{jj}^{(n)} |\log v_{jj}^{(n)}|^{1+\delta}\}^{1/2}\right) \\ &= o\left(\left[\frac{|\log \|H_n\||^{1+\delta}}{\lambda_{\min}(H_n)}\right]^{1/2}\right) \text{ a.s.}, \end{aligned}$$

where  $\hat{\beta}_n^{(LS)} = H_n^{-1} \sum_{i=1}^n x_i y_i$  is the LSE of  $\beta$  (see our Theorem 2.2.1 and Corollary 2.2.4). However, it is not true that one would have

$$a_n := H_n^{-1} \sum_{i=1}^n x_i \varepsilon_i = o\left(\max_{j \leq p} \{v_{jj}^{(n)} |\log v_{jj}^{(n)}|^{1+\delta}\}^{1/2}\right)$$

as it is claimed by relation (2.9) of [4].

Using the same notation as in [4], let us define

$$G_n(\beta) = \sum_{i=1}^n x_i (\mu(x_i^T \beta) - \mu(x_i^T \beta_0)) = s_n(\beta_0) - s_n(\beta)$$

and

$$\tilde{G}_n(\beta) = H_n^{-1} G_n(\beta).$$

Suppose that there exists  $\hat{\beta}_n = \tilde{G}_n^{-1}(a_n)$ , that is

$$\tilde{G}_n(\hat{\beta}_n) = a_n.$$

Using the definitions of  $\tilde{G}_n(\cdot)$  and  $a_n$ , this is equivalent to saying that

$$H_n^{-1} (s_n(\beta_0) - s_n(\hat{\beta}_n)) = H_n^{-1} s_n(\beta_0),$$

which in turn forces  $s_n(\hat{\beta}_n) = 0$ .

In order to prove the existence of  $\hat{\beta}_n$ , the authors of [4] use Lemma A.3.2 (Appendix A.3). For this purpose, they wish to prove that for any  $\eta > 0$ , there exists  $r = r_\eta$  such that

$$\inf_{\beta \in \partial B_\eta(\beta_0)} \|\tilde{G}_n(\beta) - 0\| \geq r \tag{118}$$

(their relation (2.12)). By the above mentioned lemma, this implies that  $\tilde{G}_n^{-1}(\overline{B_r}(0)) \subseteq \overline{B_\eta}(\beta_0)$ . Hence, if  $a_n \rightarrow 0$ , then there exists  $n_0 = n_0(\eta)$  such that

$$a_n \in \overline{B_r}(0), \quad \forall n \geq n_0.$$

Therefore  $\hat{\beta}_n = \tilde{G}_n^{-1}(a_n)$  would exist and lie in  $\overline{B_\eta}(\beta_0)$  for all  $n \geq n_0$ .

The second error of [4] is when proving (118). Using Taylor's expansion for the function  $s_n(\cdot)$  around  $\beta_0$ , we have: for  $\beta \in \partial B_\eta(\beta_0)$

$$\begin{aligned} \|\tilde{G}_n(\beta)\|^2 &= \|H_n^{-1} (s_n(\beta_0) - s_n(\beta))\|^2 = \|H_n^{-1} F_n(\tilde{\beta}_n)(\beta - \beta_0)\|^2 \\ &= (\beta - \beta_0)^T F_n(\tilde{\beta}_n) H_n^{-2} F_n(\tilde{\beta}_n)(\beta - \beta_0), \end{aligned}$$

where  $\tilde{\beta}_n$  lies between  $\beta$  and  $\beta_0$ . Relation (2.12) of [4] claims that

$$(\beta - \beta_0)^T F_n(\tilde{\beta}_n) H_n^{-2} F_n(\tilde{\beta}_n)(\beta - \beta_0) \geq m^2 (\beta - \beta_0)^T F_n(\tilde{\beta}_n) F_n(\tilde{\beta}_n)^{-2} F_n(\tilde{\beta}_n)(\beta - \beta_0), \tag{119}$$

where  $m = \inf_{i \geq 1} \mu'(x_i^T \beta)$ .

In what follows we will prove that (119) is false by providing a counter-example. (119) is equivalent to

$$(\beta - \beta_0)^T \frac{F_n(\tilde{\beta}_n)}{m} H_n^{-2} \frac{F_n(\tilde{\beta}_n)}{m} (\beta - \beta_0) \geq (\beta - \beta_0)^T (\beta - \beta_0). \quad (120)$$

We choose  $\tilde{\beta}_n$  such that  $\mu'(x_i^T \tilde{\beta}_n) = m$ ,  $\forall i \geq 2$  and  $\mu'(x_i^T \tilde{\beta}_n) > m$ ,  $i=1$ . We denote  $\alpha_i = \mu'(x_i^T \tilde{\beta}_n)/m - 1$ . It follows that  $\alpha_i = 0$ ,  $\forall i \geq 2$  and  $\alpha_1 > 0$ . Recall that

$$F_n(\tilde{\beta}_n) = \sum_{i=1}^n \mu'(x_i^T \tilde{\beta}_n) x_i x_i^T.$$

Hence

$$\begin{aligned} \frac{F_n(\tilde{\beta}_n)}{m} &= \sum_{i=1}^n \frac{\mu'(x_i^T \tilde{\beta}_n)}{m} x_i x_i^T = \sum_{i=1}^n x_i x_i^T + \sum_{i=1}^n \left( \frac{\mu'(x_i^T \tilde{\beta}_n)}{m} - 1 \right) x_i x_i^T \\ &= H_n + \sum_{i=1}^n \alpha_i x_i x_i^T = H_n + \alpha_1 x_1 x_1^T. \end{aligned} \quad (121)$$

By (121), (120) is equivalent to

$$(\beta - \beta_0)^T (H_n + \alpha_1 x_1 x_1^T) H_n^{-2} (H_n + \alpha_1 x_1 x_1^T) (\beta - \beta_0) \geq (\beta - \beta_0)^T (\beta - \beta_0)$$

which is equivalent to

$$2(\beta - \beta_0)^T H_n^{-1} x_1 x_1^T (\beta - \beta_0) + \alpha_1 (\beta - \beta_0)^T x_1 x_1^T H_n^{-2} x_1 x_1^T (\beta - \beta_0) \geq 0. \quad (122)$$

If we take  $\alpha_1$  sufficiently small, then (122) is equivalent to

$$(\beta - \beta_0)^T H_n^{-1} x_1 x_1^T (\beta - \beta_0) > 0. \quad (123)$$

If let  $p = 3$ ,  $x_1 = (1, 0, 0)^T$ ,  $x_2 = (0, 1, 0)^T$ ,  $x_3 = (1, 0, 1)^T$ ,  $\beta_1 - \beta_0 = 1$ ,  $\beta_3 - \beta_0 = 3$  and  $\beta_2$  be arbitrary, then relation (123) is not true. Hence (120) is not valid.

## 5.2 The Asymptotic Results

### 5.2.1 The Weak Consistency

In this subsection, we will show that Theorem 3.2.1 (whose proof relies heavily on the fact that  $y_i$  has an exponential density), continues to hold in the case of a GLM specified just by (116). Note that in this subsection we do not need to impose any dependence structure on the residuals  $(\varepsilon_i)_{i \geq 1}$ .

To see this, we use an argument similar to that given in the proof of Theorem 4.2.3, by taking  $T_n(\beta) = F_n^{-1/2} s_n(\beta)$ . Using Taylor's expansion for the function  $T_n(\cdot)$  around  $\beta_0$ , Lemma A.1.22 (Appendix A.1) and condition (C), we get: for any  $\beta \in \partial N_n(r)$

$$\begin{aligned} \|T_n(\beta) - T_n(\beta_0)\|^2 &= \left\| -F_n^{-1/2} F_n(\tilde{\beta}_n) F_n^{-T/2} F_n^{T/2} (\beta - \beta_0) \right\|^2 \geq \\ &\geq \left( x^T F_n^{-1/2} F_n(\tilde{\beta}_n) F_n^{-T/2} x \right)^2 \|F_n^{T/2} (\beta - \beta_0)\|^2 \geq c^2 r^2, \quad \forall n \geq n_1, \end{aligned}$$

where  $c$  and  $n_1 = n_1(r)$  are the constant given by condition (C). By Chebyshev's inequality

$$\begin{aligned} P(\|T_n(\beta_0)\| \leq cr) &\geq 1 - \frac{E\|T_n(\beta_0)\|^2}{c^2 r^2} = 1 - \frac{\text{tr}\left(F_n^{-1/2} E(s_n s_n^T) F_n^{-T/2}\right)}{c^2 r^2} \\ &= 1 - \frac{p}{c^2 r^2} = 1 - \varepsilon, \quad \forall n \geq 1, \end{aligned}$$

by choosing  $r = \sqrt{p/(c^2 \varepsilon)}$ . Hence

$$P\left(\|T_n(\beta_0)\| \leq \inf_{\beta \in \partial N_n(r)} \|T_n(\beta) - T_n(\beta_0)\|\right) \geq 1 - \varepsilon, \quad \forall n \geq n_1.$$

The argument is complete by invoking Corollary A.3.3 (Appendix A.3).

□

### 5.2.2 The Asymptotic Normality

Note that the proof of Lemma 3.2.2 (i.e. (N) implies (C)) does not rely on the density assumption on  $y_i$ . However, the proof of Lemma 3.2.3 (which gives the asymptotic

normality of  $F_n^{-1/2}s_n$ ) uses this density assumption. To overcome this difficulty, we introduce the following condition:

- ( $\mathbf{N}_\delta$ )(i) There exist  $\delta > 0$ ,  $K > 0$  such that  $\sup_{i \geq 1} E|y_i^*|^{2(1+\delta)} \leq K$ , where  $y_i^* = \varepsilon_i/\sigma_i$ .
- (ii)  $\gamma_n^{(D)} \rightarrow 0$ , where  $\gamma_n^{(D)} = \max_{1 \leq i \leq n} \sigma_i^2 x_i^T F_n^{-1} x_i$ .

**Lemma 5.2.1** *Suppose that the residuals  $(\varepsilon_i)_{i \geq 1}$  are independent. Under  $(\mathbf{N}_\delta)$ , we have*

$$F_n^{-1/2}s_n \rightarrow N(0, I) \text{ in distribution.}$$

**Proof:** By the Cràmer-Wold Theorem, it suffices to prove that

$$\lambda^T F_n^{-1/2}s_n \rightarrow N(0, I) \text{ in distribution,} \quad (124)$$

for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$ . Recall that  $s_n = \sum_{i=1}^n x_i \varepsilon_i$ . Hence

$$\lambda^T F_n^{-1/2}s_n = \sum_{i=1}^n \lambda^T F_n^{-1/2} x_i \varepsilon_i := \sum_{i=1}^n Z_{n,i}.$$

Note that  $(Z_{n,i})_{i=1, \dots, n}$  are zero-mean independent random variables and

$$E[(\lambda^T F_n^{-1/2}s_n)^2] = E(\lambda^T F_n^{-1/2}s_n s_n^T F_n^{-T/2}\lambda) = \lambda^T F_n^{-1/2} E(s_n s_n^T) F_n^{-T/2}\lambda = 1.$$

Relation (124) will follow by the Center Limit Theorem (Theorem A.2.8, Appendix A.2) providing we show that the Lindeberg condition holds, i.e.

$$\sum_{i=1}^n E(Z_{n,i}^2 I_{\{|Z_{n,i}| \geq \varepsilon\}}) \rightarrow 0, \quad \forall \varepsilon > 0. \quad (125)$$

For this, we note that

$$Z_{n,i}^2 = (\lambda^T F_n^{-1/2} x_i)^2 \varepsilon_i^2 = \gamma_{n,i}^{(D)} (y_i^*)^2,$$

where  $\gamma_{n,i}^{(D)} = \sigma_i^2 (\lambda^T F_n^{-1/2} x_i)^2$ . Note that

$$\gamma_{n,i}^{(D)} = \sigma_i^2 \cdot x_i^T F_n^{-1/2} \lambda \lambda^T F_n^{-T/2} x_i \leq \sigma_i^2 x_i^T F_n^{-1} x_i \leq \gamma_n^{(D)}, \quad \forall i \leq n,$$

since  $\lambda_{\max}(\lambda\lambda^T) = \|\lambda\|^2 = 1$ . Hence

$$I_{\{Z_{n,i}^2 \geq \varepsilon^2\}} = I_{\{(y_i^*)^2 \geq \varepsilon^2/\gamma_{n,i}^{(D)}\}} \leq I_{\{(y_i^*)^2 \geq \varepsilon^2/\gamma_n^{(D)}\}}$$

and

$$\begin{aligned} \sum_{i=1}^n E(Z_{n,i}^2 I_{\{|Z_{n,i}| \geq \varepsilon\}}) &= \sum_{i=1}^n \gamma_{n,i}^{(D)} E\left((y_i^*)^2 I_{\{(y_i^*)^2 \geq \varepsilon^2/\gamma_{n,i}^{(D)}\}}\right) \leq \sum_{i=1}^n \gamma_{n,i}^{(D)} E\left[(y_i^*)^2 \frac{(y_i^*)^{2\delta}}{(\varepsilon^2/\gamma_n^{(D)})^{2\delta}}\right] \\ &= \sum_{i=1}^n \gamma_{n,i}^{(D)} E|y_i^*|^{2(1+\delta)} \left(\frac{\gamma_n^{(D)}}{\varepsilon^2}\right)^{2\delta} \leq K\varepsilon^{-4\delta} \gamma_n^{(D)} \sum_{i=1}^n \gamma_{n,i}^{(D)}, \end{aligned} \quad (126)$$

where we used  $(\mathbf{N}_\delta)$  (i) for the last inequality above. Note that

$$\begin{aligned} \sum_{i=1}^n \gamma_{n,i}^{(D)} &= \sum_{i=1}^n \sigma_i^2 (\lambda^T F_n^{-1/2} x_i)^2 = \lambda^T F_n^{-1/2} \left(\sum_{i=1}^n \sigma_i^2 x_i x_i^T\right) F_n^{-T/2} \lambda = \\ &= \lambda^T F_n^{-1/2} F_n F_n^{-T/2} \lambda = 1. \end{aligned} \quad (127)$$

From (126) and (127) we conclude that

$$\sum_{i=1}^n E(Z_{n,i}^2 I_{\{|Z_{n,i}| \geq \varepsilon\}}) \leq K\varepsilon^{-4\delta} \gamma_n^{(D)}.$$

Relation (125) follows by  $(\mathbf{N}_\delta)$  (ii).

□

We conclude that under the additional assumption  $(\mathbf{N}_\delta)$ , Theorem 3.2.4 continues to hold for a GLM model specified just by (116), whose residuals  $(\varepsilon_i)_{i \geq 1}$  are independent.

### 5.2.3 The Strong Consistency

This subsection contains the correction to Theorem 1 of [4].

**Lemma 5.2.2** *Suppose that the residuals  $(\varepsilon_i)_{i \geq 1}$  form a martingale difference sequence. If  $(\mathbf{D})$  holds, then for any  $p$ -dimensional vector  $\lambda$  with  $\|\lambda\| = 1$  we have*

$$\frac{\lambda^T s_n}{[\lambda_{\max}(F_n)]^{1/2+\delta}} \longrightarrow 0 \text{ a.s.}$$

**Proof:** The proof follows from Theorem A.2.2 (i) (Appendix A.2) since  $(\lambda^T s_n)_{n \geq 1}$  is a zero-mean martingale with  $E[(\lambda^T s_n)^2] = \lambda^T F_n \lambda \leq \lambda_{\max}(F_n)$ .

□

We claim that Theorem 3.2.6 continues to hold in the case of a GLM specified by (116), if  $(\varepsilon_i)_{i \geq 1}$  form a martingale difference sequence.

To see this, we use an argument similar to that given in the proof of Theorem 4.2.10, by taking

$$T_n(\beta) = [\lambda_{\max}(F_n)]^{-1/2-\delta} s_n(\beta).$$

Using the Taylor's expansion for the function  $T_n(\cdot)$  around  $\beta_0$ , Lemma A.1.22 (Appendix A.1) and condition  $(\mathbf{S}_\delta)$ , we get: for any  $\beta \in \partial B_\varepsilon(\beta_0)$

$$\begin{aligned} \|T_n(\beta) - T_n(\beta_0)\|^2 &= [\lambda_{\max}(F_n)]^{-1-2\delta} \|F_n(\tilde{\beta}_n)(\beta - \beta_0)\|^2 \geq \\ &\geq [\lambda_{\max}(F_n)]^{-1-2\delta} \left(x^T F_n(\tilde{\beta}_n)x\right)^2 \|\beta - \beta_0\|^2 \geq c^2 \varepsilon^2, \quad \forall n \geq n_1, \end{aligned} \quad (128)$$

where  $c$  and  $n_1$  are the constant given by condition  $(\mathbf{S}_\delta)$ . By Lemma 5.2.2, with probability 1, for any  $\varepsilon \in (0, \varepsilon_0)$  there exists a random variable  $n_2(> n_1)$  such that

$$\|T_n(\beta_0)\| = \frac{\|s_n\|}{[\lambda_{\max}(F_n)]^{1/2+\delta}} \leq c\varepsilon, \quad \forall n \geq n_2. \quad (129)$$

From (128) and (129), we conclude that

$$\|T_n(\beta_0)\| \leq \inf_{\beta \in \partial B_\varepsilon(\beta_0)} \|T_n(\beta) - T_n(\beta_0)\|, \quad \forall n \geq n_2.$$

The argument is completed by invoking Corollary A.3.3 (Appendix A.3).

**Remark:** As it easily seen from the above argument, by using part (ii) of Theorem A.2.2, we can modify condition  $(\mathbf{S}_\delta)$  as follows:

$(\mathbf{S}_\delta^*)$  There exist some constant  $c > 0$ ,  $\delta > 0$  and  $n_1 \geq 1$  and a neighborhood  $N \subset B$  of  $\beta_0$  such that

$$\lambda_{\min}(F_n(\beta)) \geq c [\lambda_{\max}(F_n)]^{1/2} |\log \lambda_{\max}(F_n)|^\delta, \quad \forall \beta \in N, \quad \forall n \geq n_1.$$

The correction to Theorem 1 of [4] is the following:

**Theorem 5.2.3** *In a GLM specified by (116), if the residuals  $(\varepsilon_i)_{i \geq 1}$  form a martingale of difference sequence,  $(\mathbf{D})$  and  $(\mathbf{S}_\delta^*)$  hold, then there exists a sequence  $\{\hat{\beta}_n\}_n$  of random variable and a random number  $n_2$  such that*

**(a)**  $P\left(s_n(\hat{\beta}_n) = 0, \forall n \geq n_2\right) = 1$

**(b)**  $\hat{\beta}_n \longrightarrow \beta_0$  a.s.

# Appendix A

## Background Material

### A.1 Matrix Analysis Results

The following results can be found in [17].

**Definition A.1.1** Let  $A$  be a  $p \times p$  matrix. (a) Suppose that  $A$  is symmetric (i.e.  $A = A^T$ ). We say that  $A$  is **nonnegative definite** (and we write  $A \geq 0$ ) if  $x^T Ax \geq 0$  for all  $x \in R^p$ . We say that  $A$  is **positive definite** (and we write  $A > 0$ ) if  $x^T Ax > 0$  for all  $x \in R^p$ ,  $x \neq 0$ .

(b) We say that  $A$  is **nonsingular**, if  $\det(A) \neq 0$ .

**Note A.1.2**  $A > 0$  if and only if  $A \geq 0$  and  $A$  is nonsingular.

**Definition A.1.3** We say that  $\lambda \in R$  is an **eigenvalue** of  $A$  if there exists  $x \in R^p$  such that  $Ax = \lambda x$  (or equivalently  $\det(A - \lambda I) = 0$ ).

**Theorem A.1.4** If  $A$  is a symmetric matrix then all the eigenvalues of  $A$  are real numbers. (We usually denote with  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$  the smallest, respectively and the largest eigenvalue of  $A$ .)

**Theorem A.1.5** (a)  $A$  is nonnegative definite if and only if all the eigenvalues of  $A$  are nonnegative.

(b)  $A$  is positive definite if and only if all the eigenvalues of  $A$  are positive.

**Theorem A.1.6** Let  $\lambda$  be an eigenvalue of  $A$  and  $x$  be a corresponding eigenvector.

- (a) Then  $\lambda^n$  is an eigenvalue of  $A^n$  corresponding to the eigenvector  $x$ , for any  $n \geq 0$ .
- (b) If  $A$  is nonsingular, then  $\lambda^{-1}$  is an eigenvalue of  $A^{-1}$  corresponding to the eigenvector  $x$ .

**Theorem A.1.7** If denote with  $\lambda_1(A), \dots, \lambda_p(A)$  the eigenvalues of  $A$ , then

- (a)  $\text{tr}(A) = \sum_{i=1}^p \lambda_i(A)$ .
- (b)  $\det(A) = \prod_{i=1}^p \lambda_i(A)$ .

**Theorem A.1.8** If  $A$  is a symmetric matrix, then

$$\lambda_{\min}(A)x^T x \leq x^T A x \leq \lambda_{\max}(A)x^T x, \quad \forall x \in R^p.$$

**Corollary A.1.9** If  $A$  is a symmetric matrix, then

$$\lambda_{\min}(A) \leq a_{ii} \leq \lambda_{\max}(A), \quad \forall i = 1, \dots, p.$$

**Definition A.1.10** Let  $A$  be a  $p \times q$  matrix. We define the **Euclidean norm** of  $A$  by  $\|A\|_E = \left\{ \sum_{i=1}^p \sum_{j=1}^q a_{ij}^2 \right\}^{\frac{1}{2}}$ . We define the **spectral norm** of  $A$  by  $\|A\| = \left\{ \lambda_{\max}(A^T A) \right\}^{\frac{1}{2}}$ .

**Note A.1.11**  $\|A\| = \sup_{\|x\|=1} \|Ax\| = \sup_{x \neq 0} \|Ax\|/\|x\|$ .

**Note A.1.12** The Euclidean norm and the spectral norm are equivalent. i.e.  $\exists c_1 > 0, c_2 > 0$ , such that

$$c_1 \|A\|_E \leq \|A\| \leq c_2 \|A\|_E.$$

**Note A.1.13** If  $x \in R^p$  is a vector, then  $\|x\|_E = (x^T x)^{\frac{1}{2}} = \left\{ \text{tr}(x^T x) \right\}^{\frac{1}{2}} = \left( \sum_{i=1}^p x_i^2 \right)^{\frac{1}{2}}$ .

**Property (a)** If  $A \geq 0$ , then  $\|A\| = \lambda_{\max}(A)$ .

**(b)** If  $A > 0$ , then  $\|A^{-1}\| = \lambda_{\max}(A^{-1}) = 1/\lambda_{\min}(A)$ .

**Definition A.1.14** Let  $A$  be a positive definite matrix. The matrix  $A^{1/2}$  is called the **left square root** of  $A$ , if

$$A^{1/2}(A^{\frac{1}{2}})^T = A.$$

The matrix  $(A^{1/2})^T$  is denoted with  $A^{T/2}$  and is called **the right square root** of  $A$ .

**Note A.1.15** *The left square root of a positive definite matrix always exists and is positive definite. We denote with  $A^{-1/2}$  the matrix  $(A^{1/2})^{-1}$ , and with  $A^{-T/2}$  the matrix  $(A^{T/2})^{-1}$ .*

**Theorem A.1.16** *If  $A$  is a symmetric  $p \times p$  matrix and  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of  $A$ , then  $\|A\| \geq \rho(A) := \max_{i=1, \dots, p} |\lambda_i|$ .*

**Corollary A.1.17** *If  $A$  is a symmetric matrix, then  $|\lambda^T A \lambda| \leq \|A\| \|\lambda\|^2$ , for all  $\lambda$ .*

**Proof:** For all  $\lambda$ , we have

$$\lambda_{\min}(A) \lambda^T \lambda \leq \lambda^T A \lambda \leq \lambda_{\max}^T(A) \lambda^T \lambda.$$

$$\lambda_{\min}(A) \|\lambda\|^2 \leq \lambda^T A \lambda \leq \lambda_{\max}^T(A) \|\lambda\|^2.$$

Hence  $|\lambda^T A \lambda| \leq \max\{|\lambda_{\min}(A) \|\lambda\|^2|, |\lambda_{\max}(A) \|\lambda\|^2|\}$ , by Theorem A.1.16.

□

**Lemma A.1.18** *Let  $A$  be an  $p \times p$  matrix. Then the eigenvalues of  $BAB^{-1}$  are the same as the eigenvalues of  $A$ , if  $B$  is a nonsingular  $m \times m$  matrix.*

**Corollary A.1.19** *The matrices  $A^{-1}B$  and  $A^{-1/2}BA^{-T/2}$  have the same eigenvalues.*

**Lemma A.1.20 (Lemma 1, p. 317, [20])** *Suppose that  $A$  is a  $p \times q$  matrix. For any vector  $\lambda$  with  $\|\lambda\| = 1$ , we have  $\lambda^T A^T A \lambda \geq (\lambda^T A \lambda)^2$ .*

**Proof:** For any given  $p \times 1$  vector  $\lambda$  with  $\|\lambda\| = 1$ , construct an orthogonal matrix  $\Lambda$  such that its first column is  $\lambda$ . Denote  $b = \Lambda^T A \lambda$ . The first element of  $b$  is then  $b_1 = \lambda^T A \lambda$  and  $A \lambda = \Lambda b$ . So  $\lambda^T A^T A \lambda = b^T b \geq b_1^2 = (\lambda^T A \lambda)^2$ .

□

**Lemma A.1.21** *Let  $A$  and  $B$  be  $p \times p$  matrices. Suppose that  $A = (a_{ij})_{i,j=1, \dots, p}$  and  $|a_{ij}| < \varepsilon$ ,  $\forall i, j$ . Then  $|\lambda_{\max}(AB)| \leq p\varepsilon \|B\|$ .*

**Proof:** We suppose that  $x$  is the eigenvector corresponding to  $\lambda = \lambda_{\max}(AB)$  with  $\|x\| = 1$ . Then  $(BA)x = \lambda x$ . Hence we get

$$|\lambda| = \|\lambda x\| = \|BAx\| \leq \|B\|\|A\|\|x\| \leq p\varepsilon\|B\|,$$

since

$$\|A\| \leq \|A\|_E = \left( \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2 \right)^{1/2} \leq (p^2\varepsilon^2)^{1/2} = p\varepsilon.$$

□

**Lemma A.1.22** *For any  $p \times p$  matrix  $A$  and for any  $p$ -dimensional vector  $v$ , there exists a  $p$ -dimensional vector  $x$ ,  $\|x\| = 1$  such that*

$$\|Av\|^2 \geq (x^T Ax)^2 \|v\|^2.$$

**Proof:** Let  $x$  be an eigenvector of norm 1 of  $A^T A$  corresponding to  $\lambda_{\min}(A^T A)$ . Then

$$\|Av\|^2 = v^T A^T A v \geq \lambda_{\min}(A^T A) \|v\|^2 = (x^T A^T A x) \|v\|^2 \geq (x^T Ax)^2 \|v\|^2,$$

by Lemma A.1.20.

□

## A.2 Limit Theorems

**Theorem A.2.1 (Lemma 2, p. 504, [19])** *Let  $\{X_i\}_{i \geq 1}$  be a sequence of independent random variables such that  $E(X_i) = 0$  and  $E(X_i^2) < \infty$  for all  $i \geq 1$ . Let  $S_n = \sum_{i=1}^n X_i$ . Let  $\{A_n\}_{n \geq 1}$  be a sequence of positive numbers such that  $A_n \rightarrow \infty$  and there exist  $\delta > 0$ ,  $c > 0$ ,  $N \geq 1$ , such that*

$$[\text{Var}(S_n)]^{1/2+\delta} \leq cA_n, \quad \forall n \geq N.$$

Then

$$\frac{1}{A_n} S_n \rightarrow 0 \text{ a.s.}$$

**Theorem A.2.2** (Theorem 3, p. 81, [16]) *Let  $(S_n)_{n \geq 1}$  be a  $p$ -dimensional zero-mean martingale and  $F_n = \text{Cov}(S_n)$ .*

(i) *If  $\|F_n\| \rightarrow \infty$  then for any  $\delta > 0$ ,*

$$\frac{1}{(\log \|F_n\|)^{1/2+\delta}} F_n^{-1/2} S_n \rightarrow 0 \text{ a.s. (and in } L^2\text{)}.$$

(ii) *If  $(\log \|F_n\|)^r / \lambda_{\min}(F_n) \rightarrow 0$  for some  $r > (2\alpha - 1)$ ,  $\alpha > 1/2$ , then*

$$F_n^{-\alpha} S_n \rightarrow 0 \text{ a.s. (and in } L^2\text{)}.$$

**Theorem A.2.3** (Theorem 2, p. 78, [16]) *Let  $(S_n)_{n \geq 1}$  be a  $p$ -dimensional zero-mean martingale and  $(A_n)_{n \geq 1}$  be a sequence of  $p \times p$  matrices such that  $\lambda_{\min}(A_n) \rightarrow \infty$  and*

$$A_n A_n^T \leq A_{n+1} A_{n+1}^T, \quad \forall n \geq 1.$$

*Let  $X_i = S_i - S_{i-1}$ ,  $i \geq 1$ . If*

$$\sum_{i=1}^{\infty} E \|A_i^{-1} X_i\|^p < \infty, \text{ for some } p \in [1, 2],$$

*then  $A_n^{-1} S_n \rightarrow 0$  a.s. (and in  $L^p$ ).*

**Theorem A.2.4** (Theorem 29.4, p. 383, [1]) *Let  $(X_n)_{n \geq 1}$  and  $X$  be  $p$ -dimensional vectors. Then*

$$X_n \rightarrow X \text{ in distribution}$$

*if and only if*

$$\lambda^T X_n \rightarrow \lambda^T X \text{ in distribution}$$

*for any  $\lambda \in R^p$ .*

**Theorem A.2.5** (Theorem 5.5.17, p. 239, [2]) *If  $X_n \rightarrow X$  in distribution and  $Y_n \rightarrow a$  in probability, where  $a$  is a constant then*

(a)  *$Y_n X_n \rightarrow aX$  in distribution.*

(b)  *$X_n + Y_n \rightarrow X + a$  in distribution.*

**Theorem A.2.6 (Theorem 2.3.12, p. 66, [2])** *Let  $(X_i)_{i \geq 1}$  be a sequence of random variables and  $M_{X_i} = E[\exp(tX_i)]$  by the m.g.f. of  $M_{X_i}(t)$ . Suppose that  $M_{X_i}(t) \rightarrow M(t)$  as  $i \rightarrow \infty$ , for all  $t$  in a neighborhood of 0, and  $M(t)$  is an m.g.f. Then there is a unique c.d.f.  $F$  whose moments are determined by  $M(t)$  such that*

$$F_{X_i}(x) \rightarrow F_X(x), \text{ as } i \rightarrow \infty$$

for all  $x$ , where  $F_X(x)$  is continuous.

**Lemma A.2.7** *Let  $X_n$  and  $Y_n$  be  $p$ -dimensional random vectors and  $A_n$  be a  $p \times p$  matrix such that  $X_n = A_n Y_n$ , for every  $n$ . If  $\sup_{i \geq 1} E(\|X_n\|^2) < \infty$  and  $\|A_n - I\| \rightarrow 0$ . Then*

$$Y_n = X_n + o_p(1),$$

i.e.  $Y_n - X_n \rightarrow 0$  in probability.

**Theorem A.2.8 (Theorem 7.21, p. 214, [5])** *Let  $\{Z_{n,i}\}_{i=1, \dots, k_n, n \geq 1}$  be a doubly indexed sequence such that  $Z_{n1}, Z_{n2}, \dots, Z_{nk_n}$  are independent  $\forall n \geq 1$  and  $S_n = \sum_{j=1}^{k_n} Z_{nj}$ . Suppose that  $E(Z_{nj}) = 0, \forall j = 1, \dots, k_n$  and  $E(S_n^2) = 1$ . Then*

$$\sum_{j=1}^{k_n} E(Z_{nj}^2 I_{\{|Z_{nj}| > \varepsilon\}}) \rightarrow 0, \forall \varepsilon > 0,$$

if and only if

$$S_n \rightarrow N(0, 1), \text{ in distribution}$$

and

$$\max_{j \leq k_n} P(|Z_{n,j}| \geq \varepsilon) \rightarrow 0, \forall \varepsilon > 0.$$

**Theorem A.2.9 (Theorem 15-6E, p. 408, [3])** *Let  $(S_n)_{n \geq 1}$  be a martingale. The following three conditions are equivalent.  $\{X_1, X_2, \dots\}$ :*

- (a)  $\{S_1, S_2, \dots\}$  is uniformly integrable.
- (b)  $\lim_n X_n$  exists in  $L^1$ .
- (c) There exists an integrable random variable  $Z$  such that  $\{X_1, X_2, \dots, Z\}$  is a martingale.

If conditions (a) through (c) are satisfied, then  $\lim_n S_n = S_\infty$  exists a.s. and  $\{S_1, S_2, \dots, S_\infty\}$  is a martingale.

If  $\sup E[|S_n|^p] < \infty$  for some  $p > 1$ , then condition (a) through (c) are satisfied and  $\lim_n S_n = S_\infty$  in  $L^p$ .

### A.3 Analysis Results

**Theorem A.3.1** *If  $g : B_\varepsilon(x_0) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^p$  is a strictly concave function such that*

$$g(x) - g(x_0) < 0, \quad \forall x \in \partial B_\varepsilon(x_0),$$

*then there exists a local maximum of  $g$  in  $B_\varepsilon(x_0)$ .*

**Lemma A.3.2 (Lemma A, p. 1162, [20])** *Let  $T$  be a smooth injection from  $\mathbb{R}^p$  to  $\mathbb{R}^p$ , with  $T(x_0) = y_0$ . Define  $\overline{B}_\delta(x_0) = \{x \in \mathbb{R}^p, \|x - x_0\| \leq \delta\}$ , and  $S_\delta(x_0) = \partial B_\delta(x_0) = \{x \in \mathbb{R}^p, \|x - x_0\| = \delta\}$ . Then  $\inf_{x \in S_\delta(x_0)} \|T(x) - y_0\| \geq r$  implies*

- (i)  $\overline{B}_r(T(y_0)) := \{y \in \mathbb{R}^p, \|y - T(x_0)\| \leq r\} \subseteq T(\overline{B}_\delta(x_0))$
- (ii)  $T^{-1}(\overline{B}_r(y_0)) \subseteq \overline{B}_\delta(x_0)$ .

**Corollary A.3.3** *Let  $T$  be a smooth injection from  $\mathbb{R}^p$  to  $\mathbb{R}^p$ . If  $\|T(x_0)\| \leq \inf_{\{x: \|x-x_0\|=\delta\}} \|T(x) - T(x_0)\|$ , then there exists  $\hat{x} \in \overline{B}_\delta(x_0)$  such that  $T(\hat{x}) = 0$ .*

**Proof:** In Lemma A.3.2, take  $r = \|T(x_0)\|$ . Then  $\overline{B}_r(T(x_0)) \subseteq T(\overline{B}_\delta(x_0))$ .

Note that  $0 \in \overline{B}_r(T(x_0))$ , because  $\|0 - T(x_0)\| = \|T(x_0)\| = r$ . Hence  $0 \in T(\overline{B}_\delta(x_0))$  i.e. there exists  $\hat{x} \in \overline{B}_\delta(x_0)$  such that  $T(\hat{x}) = 0$ .

□

# Appendix B

## Analysis of Longitudinal Data Sets Using the SAS System

In this chapter, we present the procedure implemented by SAS for finding the GEE estimator.

The steps to find the GEE solution are (p. 547, [6]):

- (1) Relate the marginal response to a linear combination of the covariates
- (2) Describe the variance of  $y_{ij}$  as a function of the mean.
- (3) Choose the form of a working correlation matrix.
- (4) Estimate the parameter vector  $\beta$  and its covariance matrix. The GEE estimator of  $\beta$  is the solution of the estimating equation

$$U(\beta) = \sum_{i=1}^n \left( \frac{\partial \mu_i}{\partial \beta} \right)' [V_i(\hat{\alpha})]^{-1} (y_i - \mu_i) = 0.$$

The estimating equation is solved by iterating between quasi-likelihood methods for estimating  $\beta$  and method of moments estimation of  $\alpha$  as a function of  $\beta$ , as follows:

- (a) Compute an initial estimate of  $\beta$  by using a GLM model or some other method.
- (b) Compute the standardized Pearson residuals  $r_{ij} = y_{ij} - \hat{\mu}_{ij} / \sqrt{v(\hat{\mu}_{ij})}$  and obtain the estimates for the nuisance parameters  $\phi$  and  $\alpha$  by using moment estimation.
- (c) Update  $\hat{\beta}$  with

$$\hat{\beta} - \left[ \sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right]^{-1} \left[ \sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i) \right].$$

(d) Iterate until convergence.

## B.1 An Example Using Logistic Regression for Binary Data

### Passive Smoking Example (p. 480, [6])

*Objective* We are given data from a hypothetical study of the effects of air pollution on children. Researchers followed 25 children and recorded whether they were exhibiting wheezing symptoms during the periods of evaluation at age 8, 9, 10 and 11. We need to investigate the data and model the dichotomous outcome with a logistic regression analysis.

*Data Description* The response is recorded as 1 for symptoms and 0 for no symptoms. Explanatory variables included age, city and a passive smoking index with values 0, 1 and 2 that reflected the degree of smoking in the home. Since there are 4 observations with explanatory variable age, we have 100 lines' records. Note that age and the passive smoking index are time-dependent explanatory variables.

*Modelling* We use GENMOD procedure to fit the GEE model and build the logistic regression model.

*Conclusion* We find that city is not a factor in wheezing, so we do the reduced model. That is the regressors are only age and smoking exposure. Smoking exposure has a nearly significant association. Age is marginally influential.

In what follows, we present the details using SAS system (outputs attached). 1-2 shows the information about the parameters, including which parameter belongs to which level of the Class variables.

1-3 contains the initial parameter estimates. To generate a starting solution of GEE, the procedure first treats all of the measurements as independent. Note that statement GENMOD automatically creates a dummy variable for each of CLASS variable. Because variable City has 2 possible values, GENMOD has created 1 dummy variable and has taken the highest value as the omitted category.

There are 25 subjects and each subject has 4 measures showed in 1-4. The data

are complete.

In 1-5, we find with Type 3 analysis results, the city is not significant. Because the p-value is 0.0583, smoking exposure has significant association. Age is marginally influential with p-value 0.0981.

1-6 displays the parameter estimates produced by GEE analysis. Since the effect reported in Type 3 analysis are single degree of freedom effects, the score statistics displayed in 1-5 are assessing the same hypotheses as the Z statistics in 1-6.

1-7 shows the empirical and model-based covariance matrix of the parameter estimates. Note that their values are often similar, so we may have some confidence that we have specified the correlation structure correctly and the estimates are relatively efficient.

We use the exchangeable working correlation matrix (output 1-8).

We do the reduced model and present the estimates in 1-9.

1-1 Basic Model Information

	Distribution	Binomial	
	Link Function	Logit	
	Dependent Variable	Symptom	
	Number of Observations Read	100	
	Number of Observations Used	100	
	Number of Events	42	
	Number of Trials	100	
Class Level Information			
Class	Levels	Values	
Id	25	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25	
City	2	greenhil steelcit	
Response Profile			
	Ordered Value	Symptom	Total Frequency
	1	1	42
	2	0	58
PROC GENMOD is modeling the probability that Symptom='1'.			

1-2 Information About Parameters

Parameter Information			
	Parameter	Effect	City
	Prm1	Intercept	
	Prm2	City	greenhil
	Prm3	City	steelcit
	Prm4	Age	
	Prm5	Smoke	

1-3 Initial Parameter Estimates

Analysis Of Initial Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	2.4161	1.8673	-1.2438	6.0760	1.67	0.1957
City greenhil	1	0.0017	0.4350	-0.8508	0.8543	0.00	0.9968
City steelcit	0	0.0000	0.0000	0.0000	0.0000	.	.
Age	1	-0.3283	0.1914	-0.7035	0.0468	2.94	0.0863
Smoke	1	0.5598	0.2952	-0.0188	1.1385	3.60	0.0579
Scale	0	1.0000	0.0000	1.0000	1.0000		
NOTE: The scale parameter was held fixed.							

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM96

1-4 General GEE Model information

GEE Model Information	
Correlation Structure	Exchangeable
Subject Effect	Id (25 levels)
Number of Clusters	25
Correlation Matrix Dimension	4
Maximum Cluster Size	4
Minimum Cluster Size	4

1-5 Type 3 Analysis

Score Statistics For Type 3 GEE Analysis				
Source	DF	Chi-Square	Pr > ChiSq	
City	1	0.01	0.9388	
Age	1	2.74	0.0981	
Smoke	1	3.59	0.0583	

1-6 GEE Parameter Estimates

Analysis Of GEE Parameter Estimates Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept	2.2615	2.0243	-1.7060	6.2290	1.12	0.2639
City greenhil	0.0418	0.5435	-1.0234	1.1070	0.08	0.9387
City steelcit	0.0000	0.0000	0.0000	0.0000	.	.
Age	-0.3201	0.1884	-0.6894	0.0492	-1.70	0.0893
Smoke	0.6506	0.2821	0.0978	1.2035	2.31	0.0211

1-7 Covariance Matrix Estimates

Covariance Matrix (Model-Based)				
	Prm1	Prm2	Prm4	Prm5
Prm1	3.26069	-0.16313	-0.32274	-0.12257
Prm2	-0.16313	0.24015	0.002520	0.03422
Prm4	-0.32274	0.002520	0.03379	0.004471
Prm5	-0.12257	0.03422	0.004471	0.09533

  

The GENMOD Procedure				
Covariance Matrix (Empirical)				
	Prm1	Prm2	Prm4	Prm5
Prm1	4.09770	-0.55261	-0.37280	-0.29397
Prm2	-0.55261	0.29538	0.03719	0.09143
Prm4	-0.37280	0.03719	0.03550	0.02064
Prm5	-0.29397	0.09143	0.02064	0.07957

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM97

1-8 Working Correlation Matirx

Working Correlation Matrix				
	Col1	Col2	Col3	Col4
Row1	1.0000	0.0883	0.0883	0.0883
Row2	0.0883	1.0000	0.0883	0.0883
Row3	0.0883	0.0883	1.0000	0.0883
Row4	0.0883	0.0883	0.0883	1.0000

Exchangeable Working  
Correlation

Correlation	0.0882765279
-------------	--------------

1-9 The Estimate For The Reduced Model

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
Age	1	2.79	0.0947
Smoke	1	4.27	0.0387

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM98

Dataset For Passive Smoking Example.

Obs	Id	Pollution City	Study Data i	Age	Smoke	Symptom
1	1	steelcit	1	8	0	1
2	1	steelcit	2	9	0	1
3	1	steelcit	3	10	0	1
4	1	steelcit	4	11	0	0
5	2	steelcit	1	8	2	1
6	2	steelcit	2	9	2	1
7	2	steelcit	3	10	2	1
8	2	steelcit	4	11	1	0
9	3	steelcit	1	8	2	1
10	3	steelcit	2	9	2	0
11	3	steelcit	3	10	1	0
12	3	steelcit	4	11	0	0
13	4	greenhil	1	8	0	0
14	4	greenhil	2	9	1	1
15	4	greenhil	3	10	1	1
16	4	greenhil	4	11	0	0
17	5	steelcit	1	8	0	0
18	5	steelcit	2	9	1	0
19	5	steelcit	3	10	1	0
20	5	steelcit	4	11	1	0
21	6	greenhil	1	8	0	1
22	6	greenhil	2	9	0	0
23	6	greenhil	3	10	0	0
24	6	greenhil	4	11	0	1
25	7	steelcit	1	8	1	1
26	7	steelcit	2	9	1	1
27	7	steelcit	3	10	0	1
28	7	steelcit	4	11	0	0
29	8	greenhil	1	8	1	0
30	8	greenhil	2	9	1	0
31	8	greenhil	3	10	1	0
32	8	greenhil	4	11	2	0
33	9	greenhil	1	8	2	1
34	9	greenhil	2	9	2	0
35	9	greenhil	3	10	1	1
36	9	greenhil	4	11	1	0
37	10	steelcit	1	8	0	0
38	10	steelcit	2	9	0	0
39	10	steelcit	3	10	0	0
40	10	steelcit	4	11	1	0
41	11	steelcit	1	8	1	1
42	11	steelcit	2	9	0	0
43	11	steelcit	3	10	0	0
44	11	steelcit	4	11	0	1
45	12	greenhil	1	8	0	0
46	12	greenhil	2	9	0	0
47	12	greenhil	3	10	0	0
48	12	greenhil	4	11	0	0
49	13	steelcit	1	8	2	1
50	13	steelcit	2	9	2	1
51	13	steelcit	3	10	1	0
52	13	steelcit	4	11	0	1

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM 99

Dataset For Passive Smoking Example.

Obs	Id	City	i	Age	Smoke	Symptom
53	14	greenhil	1	8	0	1
54	14	greenhil	2	9	0	1
55	14	greenhil	3	10	0	0
56	14	greenhil	4	11	0	0
57	15	steelcit	1	8	2	0
58	15	steelcit	2	9	0	0
59	15	steelcit	3	10	0	0
60	15	steelcit	4	11	2	1
61	16	greenhil	1	8	1	0
62	16	greenhil	2	9	1	0
63	16	greenhil	3	10	0	0
64	16	greenhil	4	11	1	0
65	17	greenhil	1	8	0	0
66	17	greenhil	2	9	0	1
67	17	greenhil	3	10	0	1
68	17	greenhil	4	11	1	1
69	18	steelcit	1	8	1	1
70	18	steelcit	2	9	2	1
71	18	steelcit	3	10	0	0
72	18	steelcit	4	11	1	0
73	19	steelcit	1	8	2	1
74	19	steelcit	2	9	1	0
75	19	steelcit	3	10	0	1
76	19	steelcit	4	11	0	0
77	20	greenhil	1	8	0	0
78	20	greenhil	2	9	0	1
79	20	greenhil	3	10	0	1
80	20	greenhil	4	11	0	0
81	21	steelcit	1	8	1	0
82	21	steelcit	2	9	1	0
83	21	steelcit	3	10	1	0
84	21	steelcit	4	11	2	1
85	22	greenhil	1	8	0	1
86	22	greenhil	2	9	0	1
87	22	greenhil	3	10	0	0
88	22	greenhil	4	11	0	0
89	23	steelcit	1	8	1	1
90	23	steelcit	2	9	1	0
91	23	steelcit	3	10	0	1
92	23	steelcit	4	11	0	0
93	24	greenhil	1	8	1	0
94	24	greenhil	2	9	1	1
95	24	greenhil	3	10	1	1
96	24	greenhil	4	11	2	1
97	25	greenhil	1	8	0	1
98	25	greenhil	2	9	0	0
99	25	greenhil	3	10	0	0
100	25	greenhil	4	11	0	0

**SAS Code**

```
LIBNAME my 'd:\MyRawData';
DATA my.children;
  INFILE 'd:\MyRawData\polution.dat';
  INPUT Id City$ @@;
  DO i=1 to 4;
    INPUT Age Smoke Symptom@@;
    OUTPUT;
  END;
  DATALINES;
PROC PRINT DATA=my.children;
  TITLE 'Pollution Study Data';
RUN;
PROC GENMOD DATA=my.children descending;
  CLASS Id City;
  MODEL Symptom = City Age Smoke /
link=logit dist=bin type3;
  REPEATED subject=Id / type= exch covb
corrw;
RUN;
```

## B.2 An Example Using Log-linear Poisson Regression For Count Data

### **Incidence of Lower Respiratory Illness Example** (p. 542, [6])

*Objective* In order to study the incidence of lower respiratory illness, researchers take repeated observations of infants over one year. They study 284 children and examine them every two weeks. We need to investigate the data and model outcome with a Poisson regression analysis.

*Data Description* One outcome of interest is the total number of times of lower respiratory infection recorded for the year. Explanatory variables include Risk (the number of weeks during that year for which the child is considered at risk, when a lower respiratory infection is ongoing, the child is not considered to be at risk for a new one), CROWDING (an indicator variable for whether crowded conditions occur in the household), SES (an indicator variable for whether the family's socioeconomic status is considered low(0), medium(1), or high(2) ), RACE (an indicator variable for whether the child is white(1) or not(0) ), PASSIVE (an indicator variable for whether the child is exposed to cigarette smoking), and AGEGROUP (taking the values 1,2 and 3 for under four, four to six , or more that six months).

*Modelling* We model the total number of infections with Poisson regression. Since the children that have an infection are more likely to have other infections, it will cause overdispersion. The variance is not "acting" as it should; it does not take the form for data from a Poisson distribution. With GEE estimation, we are using a subject-to-subject measure for variance estimation instead of a model-based one, so we use GENMOD with GEE to fix the overdispersion.

*Conclusion* We conclude that Crowding and Smoking exposure are significant.

In what follows, we present the details using SAS system (outputs attached).

2-2 shows the information about the parameters, including which parameter belongs to which level of the Class variables.

In 2-3, comparing values of 1.4806 with 1.7877, there is evidence of overdispersion. The model-based estimates of standard errors may not be appropriate. We account for this overdispersion with the GEE-generated robust covariances.

2-4 presents the initial parameter estimates. To generate a starting solution of GEE, the procedure treats all of the measurements as independent. Note that SES and AGEGROUP are CLASS variables.

There are 284 subjects and each subject has 1 measure showed in 2-5. The data are complete.

In 2-6, parameter estimates are the same as displayed in 2-4, but the standard errors are different. They are larger than the corresponding standard errors in 2-4, because overdispersion means that the data are exhibiting additional variance.

Form 2-7, we conclude that Crowding and Smoking exposure are significant.

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM103

2-1 Basic Model Information

Model Information				
Data Set	MY.LRI			
Distribution	Poisson			
Link Function	Log			
Dependent Variable	Count			
Offset Variable	Logrisk			
Number of Observations Read	284			
Number of Observations Used	284			
Class Level Information				
Class	Levels	Values		
Id	284	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20		
		21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37		
		38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54		
		55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71		
		72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87		
		...		
		Ses	3	0 1 2
		Race	2	0 1
		Agegroup	3	1 2 3

2-2 Information About Parameter

Parameter Information				
Parameter	Effect	Ses	Race	Agegroup
Prm1	Intercept			
Prm2	Passive			
Prm3	Crowding			
Prm4	Ses	0		
Prm5	Ses	1		
Prm6	Ses	2		
Prm7	Race		0	
Prm8	Race		1	
Prm9	Agegroup			1
Prm10	Agegroup			2
Prm11	Agegroup			3

2-3 Goodness-of-Fit Statistics

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	276	408.6556	1.4806
Scaled Deviance	276	408.6556	1.4806
Pearson Chi-Square	276	493.4004	1.7877
Scaled Pearson X2	276	493.4004	1.7877
Log Likelihood		-260.6621	
Algorithm converged			

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM104

2-4 Initial Parameter Estimates

Analysis Of Initial Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr >
Intercept	1	0.6322	0.5434	-0.4330	1.6973	1.35	0.2447
Passive	1	0.4243	0.1649	0.1010	0.7476	6.62	0.0101
Crowding	1	0.4985	0.1613	0.1825	0.8146	9.56	0.0020
Ses	0 1	-0.4181	0.2157	-0.8408	0.0046	3.76	0.0525
Ses	1 1	-0.0584	0.1964	-0.4433	0.3264	0.09	0.7661
Ses	2 0	0.0000	0.0000	0.0000	0.0000	.	.
Race	0 1	0.2038	0.1754	-0.1400	0.5476	1.35	0.2453
Race	1 0	0.0000	0.0000	0.0000	0.0000	.	.
Agegroup	1 1	-0.5185	0.6753	-1.8420	0.8050	0.59	0.4426
Agegroup	2 1	-1.0500	0.5122	-2.0539	-0.0461	4.20	0.0404
Agegroup	3 0	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000	.	.

2-5 General GEE Model information

GEE Model Information	
Correlation Structure	Independent
Subject Effect	Id (284 levels)
Number of Clusters	284
Correlation Matrix Dimension	1
Maximum Cluster Size	1
Minimum Cluster Size	1

Algorithm converged.

2-6 GEE Parameter Estimates

The GENMOD Procedure							
Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z	
Intercept	0.6322	0.5477	-0.4413	1.7056	1.15	0.2484	
Passive	0.4243	0.2103	0.0120	0.8366	2.02	0.0437	
Crowding	0.4985	0.2336	0.0406	0.9565	2.13	0.0329	
Ses	0 -0.4181	0.2978	-1.0018	0.1657	-1.40	0.1604	
Ses	1 -0.0584	0.2539	-0.5561	0.4393	-0.23	0.8180	
Ses	2 0.0000	0.0000	0.0000	0.0000	.	.	
Race	0 0.2038	0.2181	-0.2237	0.6313	0.93	0.3501	
Race	1 0.0000	0.0000	0.0000	0.0000	.	.	
Agegroup	1 -0.5185	0.6038	-1.7019	0.6649	-0.86	0.3905	
Agegroup	2 -1.0500	0.4646	-1.9605	-0.1394	-2.26	0.0238	
Agegroup	3 0.0000	0.0000	0.0000	0.0000	.	.	

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM105

2-7 Type 3 Analysis

Score Statistics For Type 3 GEE Analysis				
Source	DF	Chi-Square	Pr > ChiSq	
Passive	1	3.79	0.0516	
Crowding	1	4.49	0.0341	
Ses	2	2.48	0.2888	
Race	1	0.92	0.3364	
Agegroup	2	2.96	0.2274	

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM106

Dataset for incidence of lower respiratory illness example

LRI Dataset									
Obs	Id	Count	Risk	Passive	Crowding	Ses	Agegroup	Race	Logrisk
1	1	0	42	1	0	2	2	0	-0.21357
2	2	0	43	1	0	0	2	0	-0.19004
3	3	0	41	1	0	1	2	0	-0.23767
4	4	1	36	0	1	0	2	0	-0.36772
5	5	1	31	0	0	0	2	0	-0.51726
6	6	0	43	1	0	0	2	0	-0.19004
7	7	0	45	0	0	0	2	0	-0.14458
8	8	0	42	0	0	0	2	1	-0.21357
9	9	0	45	0	0	0	2	1	-0.14458
10	10	0	35	1	1	0	2	0	-0.39590
11	11	0	43	0	0	0	2	0	-0.19004
12	12	2	38	0	0	0	2	0	-0.31366
13	13	0	41	0	0	0	2	0	-0.23767
14	14	0	12	1	1	0	1	0	-1.46634
15	15	0	6	0	0	0	3	0	-2.15948
16	16	0	43	0	0	0	2	0	-0.19004
17	17	2	39	1	0	1	2	0	-0.28768
18	18	0	43	0	1	0	2	0	-0.19004
19	19	2	37	0	0	0	2	1	-0.34033
20	20	0	31	1	1	1	2	0	-0.51726
21	21	0	45	0	1	0	2	0	-0.14458
22	22	1	29	1	1	1	2	1	-0.58395
23	23	1	35	1	1	1	2	0	-0.39590
24	24	3	20	1	1	2	2	0	-0.95551
25	25	1	23	1	1	1	2	0	-0.81575
26	26	1	37	1	0	0	2	0	-0.34033
27	27	0	49	1	0	0	2	0	-0.05942
28	28	0	35	0	0	0	2	0	-0.39590
29	29	3	44	1	1	1	2	0	-0.16705
30	30	0	37	1	1	0	2	0	-0.34033
31	31	2	39	0	1	1	2	0	-0.28768
32	32	0	41	0	0	0	2	0	-0.23767
33	33	1	46	1	1	2	2	0	-0.12260
34	34	0	5	1	1	2	3	1	-2.34181
35	35	1	29	0	0	0	2	0	-0.58395
36	36	0	31	0	1	0	2	0	-0.51726
37	37	0	22	1	1	2	2	0	-0.86020
38	38	1	22	1	1	2	2	1	-0.86020
39	39	0	47	0	0	0	2	0	-0.10110
40	40	1	46	1	1	1	2	1	-0.12260
41	41	0	37	0	0	0	2	0	-0.34033
42	42	1	39	0	0	0	2	0	-0.28768
43	43	0	33	0	1	1	2	1	-0.45474
44	44	0	34	1	0	1	2	0	-0.42488
45	45	3	32	1	1	1	2	0	-0.48551
46	46	3	22	0	0	0	2	0	-0.86020
47	47	1	6	1	0	2	3	0	-2.15948
48	48	0	38	0	0	0	2	0	-0.31366
49	49	1	43	0	1	0	2	0	-0.19004
50	50	2	36	0	1	0	2	0	-0.36772
51	51	0	43	0	0	0	2	0	-0.19004
52	52	0	24	1	0	0	2	0	-0.77319

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM107

Dataset for incidence of lower respiratory illness example

LRI Dataset									
Obs	Id	Count	Risk	Passive	Crowding	Ses	Agegroup	Race	Logrisk
53	53	0	25	1	0	1	2	1	-0.73237
54	54	0	41	0	0	0	2	0	-0.23767
55	55	0	43	0	0	0	2	0	-0.19004
56	56	2	31	0	1	1	2	0	-0.51726
57	57	3	28	1	1	1	2	0	-0.61904
58	58	1	22	0	0	1	2	1	-0.86020
59	59	1	11	1	1	1	1	0	-1.55335
60	60	3	41	0	1	1	2	0	-0.23767
61	61	0	31	0	0	1	2	0	-0.51726
62	62	0	11	0	0	1	1	1	-1.55335
63	63	0	44	0	1	0	2	0	-0.16705
64	64	0	9	1	0	0	3	1	-1.75402
65	65	0	36	1	1	1	2	0	-0.36772
66	66	0	29	1	0	0	2	0	-0.58395
67	67	0	27	0	1	0	2	1	-0.65541
68	68	0	36	0	1	0	2	0	-0.36772
69	69	1	33	1	0	0	2	0	-0.45474
70	70	2	13	1	1	2	1	1	-1.38629
71	71	0	38	0	0	0	2	0	-0.31366
72	72	0	41	0	0	0	2	1	-0.23767
73	73	0	41	1	0	2	2	0	-0.23767
74	74	0	35	0	0	1	2	0	-0.39590
75	75	0	45	0	0	0	2	0	-0.14458
76	76	4	38	1	0	2	2	1	-0.31366
77	77	1	42	1	0	0	2	1	-0.21357
78	78	1	42	1	1	2	2	1	-0.21357
79	79	6	36	1	1	0	2	0	-0.36772
80	80	2	23	1	1	1	2	1	-0.81575
81	81	1	32	0	0	1	2	0	-0.48551
82	82	0	41	0	1	0	2	0	-0.23767
83	83	0	50	0	0	0	2	0	-0.03922
84	84	0	42	1	1	1	2	1	-0.21357
85	85	1	30	0	0	0	2	0	-0.55005
86	86	2	47	0	1	0	2	0	-0.10110
87	87	1	35	1	1	2	2	0	-0.39590
88	88	1	38	1	0	1	2	1	-0.31366
89	89	1	38	1	1	1	2	1	-0.31366
90	90	1	38	1	1	1	2	1	-0.31366
91	91	0	32	1	1	1	2	0	-0.48551
92	92	1	3	1	0	1	3	1	-2.85263
93	93	0	26	1	0	0	2	1	-0.69315
94	94	0	35	1	0	0	2	0	-0.39590
95	95	3	37	1	0	0	2	0	-0.34033
96	96	1	41	1	0	1	2	0	-0.23767
97	97	1	26	1	1	2	2	0	-0.69315
98	98	0	36	0	0	0	2	0	-0.36772
99	99	0	34	0	0	0	2	0	-0.42488
100	100	1	3	1	1	2	3	1	-2.85263
101	101	0	45	1	0	0	2	0	-0.14458
102	102	0	38	0	0	1	2	0	-0.31366
103	103	0	41	1	1	1	2	1	-0.23767
104	104	1	37	0	1	0	2	0	-0.34033

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM108

Dataset for incidence of lower respiratory illness example

LRI Dataset									
Obs	Id	Count	Risk	Passive	Crowding	Ses	Agegroup	Race	Logrisk
105	105	0	40	0	0	0	2	0	-0.26236
106	106	1	35	1	0	0	2	0	-0.39590
107	107	0	28	0	1	2	2	0	-0.61904
108	108	3	33	0	1	2	2	0	-0.45474
109	109	0	38	0	0	0	2	0	-0.31366
110	110	0	42	1	1	2	2	1	-0.21357
111	111	0	40	1	1	2	2	0	-0.26236
112	112	0	38	0	0	0	2	0	-0.31366
113	113	2	37	0	1	1	2	0	-0.34033
114	114	1	42	0	1	0	2	0	-0.21357
115	115	5	37	1	1	1	2	1	-0.34033
116	116	0	38	0	0	0	2	0	-0.31366
117	117	0	4	0	0	0	3	0	-2.56495
118	118	2	37	1	1	1	2	0	-0.34033
119	119	0	39	1	0	1	2	0	-0.28768
120	120	0	42	1	1	0	2	0	-0.21357
121	121	0	40	1	0	0	2	0	-0.26236
122	122	0	36	1	0	0	2	0	-0.36772
123	123	1	42	0	1	1	2	0	-0.21357
124	124	1	39	0	0	0	2	0	-0.28768
125	125	2	29	0	0	0	2	0	-0.58395
126	126	3	37	1	1	2	2	1	-0.34033
127	127	0	40	1	0	0	2	0	-0.26236
128	128	0	40	0	0	0	2	0	-0.26236
129	129	0	39	0	0	0	2	0	-0.28768
130	130	0	40	1	0	1	2	0	-0.26236
131	131	1	32	0	0	0	2	0	-0.48551
132	132	0	46	1	0	1	2	0	-0.12260
133	133	4	39	1	1	0	2	0	-0.28768
134	134	0	37	0	0	0	2	0	-0.34033
135	135	0	51	0	0	1	2	0	-0.01942
136	136	1	39	1	1	0	2	0	-0.28768
137	137	1	34	1	1	0	2	0	-0.42488
138	138	1	14	0	1	0	1	0	-1.31219
139	139	2	15	1	0	0	2	0	-1.24319
140	140	1	34	1	1	0	2	1	-0.42488
141	141	0	43	0	1	0	2	0	-0.19004
142	142	1	33	0	0	0	2	0	-0.45474
143	143	3	34	1	0	0	2	0	-0.42488
144	144	0	48	0	0	0	2	0	-0.08004
145	145	4	26	1	1	0	2	0	-0.69315
146	146	0	30	0	1	2	2	1	-0.55005
147	147	0	41	1	1	1	2	0	-0.23767
148	148	0	34	0	1	1	2	0	-0.42488
149	149	0	43	0	1	0	2	0	-0.19004
150	150	1	31	1	0	1	2	0	-0.51726
151	151	0	26	1	0	1	2	0	-0.69315
152	152	0	37	0	0	0	2	0	-0.34033
153	153	0	44	0	0	0	2	0	-0.16705
154	154	0	40	1	0	0	2	0	-0.26236
155	155	0	8	1	1	1	3	1	-1.87180
156	156	0	40	1	1	1	2	1	-0.26236

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM109

Dataset for incidence of lower respiratory illness example

Obs	Id	Count	Risk	Passive	Crowding	Ses	Agegroup	Race	Logrisk
157	157	1	45	0	0	0	2	0	-0.14458
158	158	0	4	0	0	2	3	0	-2.56495
159	159	1	36	0	1	0	2	0	-0.36772
160	160	3	37	1	1	1	2	0	-0.34033
161	161	0	15	1	0	0	1	0	-1.24319
162	162	1	27	1	0	1	2	1	-0.65541
163	163	2	31	0	1	0	2	0	-0.51726
164	164	0	42	0	0	0	2	0	-0.21357
165	165	0	42	1	0	0	2	0	-0.21357
166	166	1	38	0	0	0	2	0	-0.31366
167	167	0	44	1	0	0	2	0	-0.16705
168	168	0	45	0	0	0	2	0	-0.14458
169	169	0	34	0	1	0	2	0	-0.42488
170	170	2	41	0	0	0	2	0	-0.23767
171	171	2	30	1	1	1	2	0	-0.55005
172	172	0	44	0	0	0	2	0	-0.16705
173	173	0	40	1	0	0	2	0	-0.26236
174	174	2	31	0	0	0	2	0	-0.51726
175	175	0	41	1	0	0	2	0	-0.23767
176	176	0	41	0	0	0	2	0	-0.23767
177	177	0	39	1	0	0	2	0	-0.28768
178	178	0	40	1	0	0	2	0	-0.26236
179	179	2	35	1	0	2	2	0	-0.39590
180	180	1	43	1	0	0	2	0	-0.19004
181	181	2	39	0	0	0	2	0	-0.28768
182	182	0	35	1	1	0	2	0	-0.39590
183	183	0	37	0	0	0	2	0	-0.34033
184	184	3	37	0	0	0	2	0	-0.34033
185	185	0	43	0	0	0	2	0	-0.19004
186	186	0	42	0	0	0	2	0	-0.21357
187	187	0	42	0	0	0	2	0	-0.21357
188	188	0	38	0	0	0	2	0	-0.31366
189	189	0	36	1	0	0	2	0	-0.36772
190	190	0	39	0	1	0	2	0	-0.28768
191	191	0	44	1	0	0	2	0	-0.16705
192	192	0	45	0	0	0	2	1	-0.14458
193	193	0	42	0	0	0	2	0	-0.21357
194	194	1	31	0	0	0	2	1	-0.51726
195	195	0	35	0	0	0	2	0	-0.39590
196	196	1	35	1	0	0	2	0	-0.39590
197	197	1	27	1	0	1	2	0	-0.65541
198	198	1	33	0	0	0	2	0	-0.45474
199	199	0	39	1	0	1	2	0	-0.28768
200	200	3	40	0	1	2	2	0	-0.26236
201	201	4	26	1	0	1	2	0	-0.69315
202	202	0	14	1	1	1	1	1	-1.31219
203	203	0	39	0	1	1	2	0	-0.28768
204	204	0	4	1	1	1	3	0	-2.56495
205	205	1	27	1	1	1	2	1	-0.65541
206	206	0	36	1	0	0	2	1	-0.36772
207	207	0	30	1	0	2	2	1	-0.55005
208	208	0	34	0	1	0	2	0	-0.42488

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM110

Dataset for incidence of lower respiratory illness example

Obs	Id	Count	Risk	Passive	Crowding	Ses	Agegroup	Race	Logrisk
209	209	1	40	1	1	1	2	0	-0.26236
210	210	0	6	1	0	1	1	1	-2.15948
211	211	1	40	1	1	1	2	0	-0.26236
212	212	2	43	0	1	0	2	0	-0.19004
213	213	0	36	1	1	1	2	0	-0.36772
214	214	0	35	1	1	1	2	1	-0.39590
215	215	1	35	1	1	2	2	0	-0.39590
216	216	0	43	1	0	1	2	0	-0.19004
217	217	0	33	1	1	2	2	0	-0.45474
218	218	0	36	0	1	1	2	1	-0.36772
219	219	1	41	0	0	0	2	0	-0.23767
220	220	0	41	1	1	0	2	1	-0.23767
221	221	1	42	0	0	0	2	1	-0.21357
222	222	0	33	0	1	2	2	1	-0.45474
223	223	0	40	1	1	2	2	0	-0.26236
224	224	0	40	1	1	1	2	1	-0.26236
225	225	0	40	0	0	2	2	0	-0.26236
226	226	0	28	1	0	1	2	0	-0.61904
227	227	0	47	0	0	0	2	1	-0.10110
228	228	0	18	1	1	2	2	1	-1.06087
229	229	0	45	0	0	0	2	0	-0.14458
230	230	0	35	0	0	0	2	0	-0.39590
231	231	1	17	1	0	1	1	1	-1.11803
232	232	0	40	0	0	0	2	0	-0.26236
233	233	0	29	1	1	2	2	0	-0.58395
234	234	1	35	1	1	1	2	0	-0.39590
235	235	0	40	0	0	2	2	0	-0.26236
236	236	1	22	1	1	1	2	0	-0.86020
237	237	0	42	0	0	0	2	0	-0.21357
238	238	0	34	1	1	1	2	1	-0.42488
239	239	6	38	1	0	1	2	0	-0.31366
240	240	0	25	0	0	1	2	1	-0.73237
241	241	0	39	0	1	0	2	0	-0.28768
242	242	1	35	0	1	2	2	1	-0.39590
243	243	1	36	1	1	1	2	1	-0.36772
244	244	0	23	1	0	0	2	0	-0.81575
245	245	4	30	1	1	1	2	0	-0.55005
246	246	1	41	1	1	1	2	1	-0.23767
247	247	0	37	0	1	1	2	0	-0.34033
248	248	0	46	1	1	0	2	0	-0.12260
249	249	0	45	1	1	0	2	1	-0.14458
250	250	1	38	1	1	1	2	0	-0.31366
251	251	0	10	1	1	1	1	0	-1.64866
252	252	0	30	1	1	2	2	0	-0.55005
253	253	0	32	0	1	2	2	0	-0.48551
254	254	0	46	1	0	0	2	0	-0.12260
255	255	5	35	1	1	2	2	1	-0.39590
256	256	0	44	0	0	0	2	0	-0.16705
257	257	0	41	0	1	1	2	0	-0.23767
258	258	2	36	1	0	1	2	0	-0.36772
259	259	0	34	1	1	1	2	1	-0.42488
260	260	1	30	0	1	0	2	1	-0.55005

APPENDIX B. ANALYSIS OF LONGITUDINAL DATA SETS USING THE SAS SYSTEM111

Dataset for incidence of lower respiratory illness example

Obs	Id	Count	Risk	Passive	Crowding	Ses	Agegroup	Race	Logrisk
261	261	1	27	1	0	0	2	0	-0.65541
262	262	0	48	1	0	0	2	0	-0.08004
263	263	1	6	0	1	2	3	1	-2.15948
264	264	0	38	1	1	0	2	1	-0.31366
265	265	0	29	1	1	1	2	1	-0.58395
266	266	1	43	0	1	2	2	1	-0.19004
267	267	0	43	0	1	0	2	0	-0.19004
268	268	0	37	1	0	2	2	0	-0.34033
269	269	1	23	1	1	0	2	1	-0.81575
270	270	0	44	0	0	1	2	0	-0.16705
271	271	0	5	0	0	1	3	1	-2.34181
272	272	0	25	1	1	2	2	0	-0.73237
273	273	0	25	1	0	1	2	0	-0.73237
274	274	1	28	1	1	1	2	1	-0.61904
275	275	0	7	0	1	0	3	1	-2.00533
276	276	0	32	0	0	0	2	0	-0.48551
277	277	0	41	0	0	0	2	0	-0.23767
278	278	1	33	1	1	2	2	1	-0.45474
279	279	2	36	1	1	2	2	0	-0.36772
280	280	0	31	0	0	0	2	0	-0.51726
281	281	0	18	0	0	0	2	0	-1.06087
282	282	1	32	1	0	2	2	0	-0.48551
283	283	0	22	1	1	2	2	1	-0.86020
284	284	0	35	0	0	0	2	1	-0.39590

**SAS Code**

```
LIBNAME my 'd:\MyRawData';
DATA my.lri;
  INFILE 'd:\MyRawData\lri.dat';
  INPUT Id Count Risk Passive Crowding Ses
Agegroup Race @@;
  Logrisk=log(Risk/52);
  DATALINES;
PROC PRINT DATA=my.lri;
  TITLE 'LRI Dataset';
RUN;
PROC GENMOD;
  CLASS Id Ses Race Agegroup;
  MODEL Count= Passive Crowding Ses Race
Agegroup /
  dist=poisson link=log
offset=logrisk type3;
  REPEATED subject=Id / type=ind;
RUN;
```

# Bibliography

- [1] Billingsley, P. (1995). *Probability and Measure*, Third Edition, John Wiley, New York.
- [2] Berger, R.L. and Casella, G. (2002). *Statistical Inference*, Second Edition, Duxbury Press, Pacific Grove, California.
- [3] Burrell, C.W. (1972). *Measure, Integration, and Probability*, Mc-Gram Hill, New York.
- [4] Chen, K., Hu, I. and Ying, Z. (1999). Strong Consistency of Maximum Quasi-likelihood Estimators in Generalized Linear Models With Fixed and Adaptive Designs. *Ann. Stat.*, **27**, pages 1155-1163.
- [5] Chung, K.L. (2001). *A Course to Probability Theory*, Third Edition, Academic Press, San Diego.
- [6] Davis, C.S., Stokes, M.E. and Koch, G.G. (2000). *Categorical Data Analysis Using the SAS System*, Second Edition, SAS Institute and John Wiley, Cary, North Carolina.
- [7] Diggle, P.J., Liang K.-Y. and Zeger, S.L. (1996). *Analysis of Longitudinal Data*, Clarendon Press, Oxford.
- [8] Drygas, H. (1976) Weak and Strong Consistency of Least Squares Estimators in Regression Models, *Z. Wahr. verw. Gebiete*, **34**, pages 119-127.

- [9] Fahrmeir, L. and Kaufmann, H. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *Ann. Stat.*, **13**, pages 342-368.
- [10] James, W. H. and Joseph, M.H. (2003) *Generalized Estimating Equations*, Chapman and Hall/CRC, London.
- [11] Jennrich, R.I. and Yuan, K.-H. (1998). Asymptotic of Estimating Equations Under Natural Conditions. *J. Multivariate Anal.*, **65**, pages 245-260.
- [12] Lai T.L., Robbins, H. and Wei C.Z. (1979). Strong Consistency of Least Squares Estimates in Multiple Regression II. *J. Multiv. Anal.*, **9** pages 343-361.
- [13] Lai, T.L. and Wei, C.Z. (1982). Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Danamic Systems. *Ann. Stat.*, **10**, pages 154-156.
- [14] Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, pages 13-22.
- [15] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition, Chapman and Hall, London.
- [16] Kaufmann, H. (1987). On the Strong Law of Large Numbers for Multivariate Martingales. *Stoch. Proc. Appl.*, **26**, pages 73-85.
- [17] Schott, J.R. (2005). *Matrix Analysis for Statistics*, Second Edition, Wiley, Hoboken New York.
- [18] Shao, J. (1992). Asymptotic Theory in Generalized Linear Models with Nuisance Scale Parameters. *Probab. Theory Rel. Fields*, **91**, pages 25-41.
- [19] Wu, C.F. (1981). Asymptotic Theory of Nonlinear Least Squares Estimation. *Ann. Stat.*, **9** pages 501-513.
- [20] Xie, M. and Yang, Y. (2003). Asymptotics for Generalized Estimating Equations with Large Cluster Sizes. *Ann. Stat.*, **31**, pages 310-347.