LOGISTIC REGRESSION WITH MISCLASSIFIED COVARIATES USING AUXILIARY DATA

by

NATHAN N. DONG

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2009

ACKNOWLEDGMENTS

I would like to express my deep appreciation to my supervising professor Dr. D.L. Hawkins for his constant patience, motivation, and encouragement with me. His knowledge, insight, and invaluable advice have guided me through the course of my doctoral studies. He is always devoted to his students, even during the hardest times of his life. I am so lucky to be one of his students.

I also would like to thank Dr. Chien-Pai Han, Dr. Danny Dyer, Dr. Shan Sun-Mitchell, and Dr. Andrzej Korzeniowski for taking time from their busy schedules to be on my committee. I am especially grateful to Dr. Korzeniowski for his help and invaluable advice with my paper. Of course, the financial support from Department of Mathematics provides me a peaceful mind to concentrate on the study, the GAANN fellowship from U.S. Department of Education is one of those. I would like to thank Dr. Tie Luo, Dr. Minerva Cordero, Dr. Guojun Liao, Dr. Larry Heath, Dr. Tuncay Aktosun, Dr. Jianzhong Su, Dr. Jianping Zhu …for introducing and keeping me in the math department over those years. Finally, I would like to thank my entire family, especially my parents and my wife, Lien for the sacrifice, support, and understanding. Thank you for believing in me.

July 4, 2009

ABSTRACT

LOGISTIC REGRESSION WITH MISCLASSIFIED COVARIATES USING AUXILIARY DATA

Nathan N. Dong, Ph.D.

The University of Texas at Arlington, 2009

Supervising Professor: Doyle L. Hawkins, Jr.

When standard regression methods are used, measurement errors cause a bias in parameters estimation. In dealing with discrete covariates, such measurement errors are known as misclassification and the corresponding discrete covariate is said to be misclassified. Even though the collected data may not be fully reliable, it may be possible to collect some sub-data with full precision called auxiliary data and the remaining data called primary data may contain misclassification.

In this paper, in order to improve predictions from the primary data with misclassification, we propose a method based on the maximum likelihood

approach; by using the prediction with auxiliary data we are able to improve the prediction from the primary data and to correct the bias in parameters estimation.

For the simplified model, we consider a primary data sample with binary response Y and discrete covariate W which is a misclassified version of true latent variable X. In addition, an auxiliary data in which both W and X can be observed used to adjust for the bias due to misclassification in W. First the model parameters are shown to be non-identifiable. To resolve around the problem, we replace the estimated misclassification probability between X and W in the model. The estimator is shown to be consistent and asymptotic normality under some regularity conditions. Testing of our method by using Monte Carlo simulations indicate that the method works well with finite samples.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

ABSTRACT.. iv

LIST OF TABLES…………………………………………………………… viii

Chapter Page

1. INTRODUCTION.. 1

1.1 Introduction 1

1.2 Literature Review 2

1.3 Development of the $(\underset{\sim}{\beta}, \underset{\sim}{P})$ Likelihood Function 6

1.4 Non-Identifiability of the $(\underset{\sim}{\beta}, \underset{\sim}{P})$ Likelihood Function 9

2. THE PROPOSED LOGISTIC MODEL... 12

2.1 The Auxiliary Sample 12

2.2 The Approximate Likelihood Function..................................... 13

2.3 Large Sample Performance of $\underset{\sim}{\beta}_{N,M}$ 13

2.4 Monte Carlo Study of the Performance of $\underset{\sim}{\beta}_{N,M}$ 14

3. PROOFS OF RESULTS.. 20

3.1 Notations and Preliminaries.. 20

3.2 Assumptions 21

3.3 Consistency of $\hat{\underset{\sim}{\beta}}_{N,M}$ 21

3.4 Asymptotic Normality of $\hat{\underset{\sim}{\beta}}_{N,M}$ 24

4. EXTENSION AND DISCUSSION 27

4.1 Model for Including Continuous Variables.............................. 27

4.2 Model for Multiple Misclassified Variables 28

4.3 Future Work 29

Appendix

A. SOME USEFUL INEQUALITIES... 30

B. PROOFS OF MAIN RESULTS 32

C. PROOFS OF CONSISTENCY CONDITIONS C1-C5 47

D. PROOFS OF ASYMPTOTIC NORMALITY CONDITIONS N1-N9... 56

E. SAS PROGRAM FOR MONTE CARLO SIMULATION STUDY....... 80

REFERENCES 94

BIOGRAPHICAL INFORMATION... 96

LIST OF TABLES

Table Page

1 Monte Carlo Simulation Study 1 16

2 Monte Carlo Simulation Study 2 17

3 Monte Carlo Simulation Study 3 18

4 Monte Carlo Simulation Study 4 19

CHAPTER 1

INTRODUCTION

1.1 Introduction

Measurement error is recognized as a common problem in many fields of science. There are many possible sources of data errors. Examples include the high cost of measurement procedure to get exact value, tendency of study subjects not to report the true etc. It is well known that measurement in predictor biases effect estimate in regression modeling, losses of power in statistical testing, and makes graphical model analysis difficult. Fuller (1987), Chen and Van Ness (1999) provide the standard reference in linear models, while Carroll et al. (2006) provides an excellent overview of methods in non-linear models.

Variables that can take values on continuous scale are thought to be continuous variables, whereas those variables takes on only discrete set of values are thought to be discrete variables. Error in continuous variable is called measurement error and in discrete variable is called misclassification. Gustafson (2005) shows that bias effect estimate due to error in discrete variable even more profound than that of continuous variable.

While measurement error models have received considerable attention over the past few decades, there are only few papers on misclassification in the context of regression models. This is particularly true for multiple-value discrete variable. The model considered in this study contains two variables: Y and X. The variable Y is a dependent variable which takes on value 0 or 1, X is the latent true discrete variable which is subject to misclassification error. The objective is to estimate the coefficient of the logistic model of the form taking into account of the misclassification in X:

$$Log\left(\frac{P(Y_i=1|X_i)}{P(Y_i=0|X_i)}\right)=\beta_0+\beta_1(X_i=1)+\beta_2(X_i=2)+...+\beta_{K-1}(X_i=(K-1))$$, where

$(X_i = j) = \begin{cases} 1 \ if \ \ X_i = j \\ 0 \ \ otherwise \end{cases}$ for $1 \le j \le (K-1)$, $\beta_0,...,\beta_{K-1}$ are regression coefficients for correspondence level of X.

One motivation for this model is to study the effect of attendance on the student's pass/fail rate. Suppose we can observe N sampling units **(called primary data)** of response variable Y (= 1 for passing grade, 0 for fail) and attendance level variable W (=1 for poor, 2 for acceptable, and 3 for good) which is a misclassified version of a true latent attendance level variable X. Besides primary data, we can observe M sampling units **(called auxiliary data)** of true attendance level X and its misclassified version W. We will use this information to approximate the misclassification probability matrix between X and W, denoted by $\underset{\sim}{P}=\left[p_{ij}\right]_{3\times 3}$, where $p_{ij}=P(X=i\,|\,W=j)$ and then obtain an approximate likelihood function for logistic regression model (see the development for the likelihood function beginning section 1.3).

1.2 Literature Review

The misclassification error in a binary independent variable has been analyzed in a few studies. Aigner (1973) shows that the error is correlated with the independent variable; moreover, without additional information the regression models cannot be identified. Bollinger (1996) considers regression models with misclassified covariates and provides a bound for the estimate for the interested parameter. Mahajan (2006) uses an instrumental variable together with certain restriction to provide the identification in a nonlinear model.

For the case of a general discrete variable, Swartz et al. (2004) discuss identification problems due to misclassification from a Bayesian perspective. They start with the basic multinomial misclassification model for which N sampling units are observed that can take on K possible outcomes ($W_l,\ l=1,...,N$). Define $q_i = P(W_l = i)$ and $p_i = P(X_l = i)$, for $i = 1,...,K; l = 1,...,N$. The misclassification probabilities is defined as

$$\prod = \left[\pi_{ji} \right]_{K\times K},\ \pi_{ji} = P(W_l = i \mid X_l = j)$$

The likelihood is then $L = L(p_1,...,p_K,\pi_1,...,\pi_K) = \prod_{l=1}^{K} q_l^{N_l} = \prod_{l=1}^{K}\sum_{h=1}^{K}\left(p_h\pi_{hl}\right)^{N_l}$ where π_j is the j^{th} row of $\prod$, n_i is the number of sampling units such that $W_l = i$. They discuss two types of nonidentifiability. The permutation type is the form of swapping positions between p_i and p_j, π_i and π_j but the likelihood L stays unchanged. The second takes the form of contours when the probability distribution of data is viewed as a function of the parameters. To overcome the permutation-type nonidentifiabilities, they introduce a various constraints on the matrix of misclassification probabilities:

a) $\pi_{j,1} < ... < \pi_{j,j-1} < \pi_{j,j} > \pi_{j,j+1} > ... > \pi_{j,K}$ for $j = 1,...,K$

b) $\pi_{j,i} < \pi_{j,j}$ for all $i \neq j$

c) $\pi_{i,j} + \pi_{j,i} < \pi_{i,i} + \pi_{j,j}$ for all $i < j$

d) $\pi_{j,i} < \pi_{i,i}$ for all $i \neq j$

The point identification then is achieved by specifying a Dirichlet distribution prior on the misclassification probabilities and on $P(X)$.

Molinari (2008) proposes the so-called direct misclassification approach which formalizes the misclassification in matrix notation. She defines the relation between the observable distribution of W and the unobservable distribution of X as follow:

$$\begin{bmatrix} P(W=1) \\ \dots \\ P(W=k) \end{bmatrix} = \begin{bmatrix} P(W=1 \mid X=1) & \dots & P(W=1 \mid X=k) \\ \dots & \dots & \dots \\ P(W=k \mid X=1) & \dots & P(W=k \mid X=k) \end{bmatrix} \begin{bmatrix} P(X=1) \\ \dots \\ P(X=k) \end{bmatrix}$$

Denoted by $P^w = \prod P^x$ (*)

Let $H\left[\prod\right]$ be the set containing $\prod = \left\{\pi_{ij}\right\}_{i,j \in X}$

The identification region H_{P^x} is defined as the set of column vector P^x such that given $\prod \in H\left[\prod\right]$, P^x solves system (*) above.

$$H_{P^x} = \left\{P^x : P^w = \prod P^x, \prod \in H\left[\prod\right]\right\}$$

Under variety of assumptions on $\prod$, she provides the interval identification for P^x.

Hu (2008) introduces the matrix diagonalization technique which is a close study to that of Mahajan (2006) but dealing with multi-value discrete variable.

In his paper, besides variables Y and W, an accurately measured variable Z and a variable V which is used as an instrument for latent variable X are observed in an i.i.d. sample satisfying:

Assumption 1: $f_{Y|XWVZ}(y \mid x, w, v, z) = f_{Y|XZ}(y \mid x, z)$

Assumption 2: $f_{W|XVZ}(w \mid x, v, z) = f_{W|XZ}(w \mid x, z)$

These assumptions are called non-differential measurement error and are widely used in relevant literature; it implies that W and V contain no information about Y once X is available and V contains no information about W once X is available.

He then defines:

$$F_{y|xz_{(K \times K)}} = \begin{bmatrix} f_{y|xz}(y \mid 1, z) & 0 & \dots & 0 \\ 0 & f_{y|xz}(y \mid 2, z) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & f_{y|xz}(y \mid K, z) \end{bmatrix}$$

$F_{yw|vz_{(K\times K)}}$, $F_{x|vz_{(K\times K)}}$, $F_{w|xz_{(K\times K)}}$, and $F_{y|vz_{(K\times 1)}}$ are defined similarly over all possible values of X,W,V.
Using assumption 1, 2, together with:
Assumption 3: Rank($F_{x|vz_{(K\times K)}}$) = K
Assumption 4: $F_{w|xz_{(K\times K)}}$ is invertible

He shows $F_{y|xz_{(K\times K)}} = F_{w|xz} \times A \times F_{w|xz}^{-1}, \; where \; A = F_{w|vz}^{-1} \times F_{yw|vz}$

This is a form of similar matrix he exploits in the proof of identifiability. With the following assumption, he ensures the eigenvalues are unique:
Assumption 5: There exists a function $\varpi(.)$ such that

$$E\left[\varpi(y) \mid x = i, z\right] \neq E\left[\varpi(y) \mid x = j, z\right] \text{ for all } i \neq j .$$

Together with various assumptions similar to that of Swartz et. al. (2004) he shows the model $f_{y|xz}$, together with $f_{w|xz}$ and $f_{x|vz}$ is nonparametrically identifiable and directly estimable.

Kuchenhoff (2006) develops the MC-SIMEX (Misclassification-Simulation and Extrapolation) method as a general regression method for misclassification error. His method is an extension of SIMEX method which developed by Cook and Stefanski (1994) for dealing with additive measurement error in continuous variable. Given data $(Y_i, W_i, Z_i)_{i=1}^N$ and misclassification matrix between W and X, $\prod_{K\times K}$, which is defined by its components $\pi_{ij} = P(W = i \mid X = j)$. Suppose β is the parameter of interest. If the misclassification is ignored, then the estimator using standard regression is called the naïve estimator. Denote $\hat{\beta}_{na}\left[(Y_i, W_i, Z_i)_{i=1}^N\right] := \beta^*(\prod)$ and assume the consistent estimator $\beta = \beta^*(I_{K\times K})$, where $I_{K\times K}$ is the identity matrix. The MC-SIMEX procedure consists of two steps:
Simulation step:
For a fixed grid of values $\lambda_1, ..., \lambda_M (\geq 0)$ B new pseudo data set are simulated by $W_{b,i}(\lambda_k) := MC\left[\prod^{\lambda_k}\right](W_i)$, i=1,…,N; b=1,…,B; k=1,…,M.

where $\prod^{\lambda_k} := E\Lambda^{\lambda_k}E^{-1}$, Λ is diagonal matrix of eigenvalues and E the corresponding matrix of eigenvectors; and $MC\left[\prod^{\lambda}\right](W_i)$ denotes the simulation of a variable given W_i with misclassification matrix $\prod^{\lambda}$. For each λ_k the estimator is approximated by $\hat{\beta}_{\lambda_k} := \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{na}\left[(Y_i, W_{b,i}(\lambda_k)_i, Z_i)_{i=1}^{N}\right]$, k=1,…,M.

According to this setup, W has misclassification matrix Π in relation to X, $W_b(\lambda_k)$ has misclassification matrix $\prod^{\lambda_k}$ in relation to W, then $W_b(\lambda_k)$ has misclassification matrix $\prod^{1+\lambda_k}$ in relation to X .

Extrapolation step:

In this step a parametric curve $\beta^*(\prod^{\lambda}) \approx g(\lambda, \Gamma)$ is fitted by least squares on $\left[1+\lambda_k, \hat{\beta}_{\lambda_k}\right]_{k=0}^{M}$ to estimate parameter $\hat{\Gamma}$. The MC-SIMEX estimator is then given by $\hat{\beta}_{SIMEX} := g(0, \hat{\Gamma})$.

1.3 Development of the $(\underset{\sim}{\beta}, \underset{\sim}{P})$ Likelihood Function

Following the motivation example in section 1.1, we consider a model with response variable Y which takes on values either 0 or 1 and a true latent variable X which has values 1, 2, or 3. W is an observable variable which is a misclassified version of X and also can take on values 1, 2, or 3. This misclassification probability between X and W is denoted by $\underset{\sim}{P} = \left[p_{ij}\right]_{3\times 3}$, where $p_{ij} = P(X = i \mid W = j)$.

Considering each observation of the response Y_i = 0, 1 to be a Bernoulli random variable for which we can state the probability distribution as follow:

Y_i	Probability
1	P(Y_i=1| X_i) = π_i
0	P(Y_i=0| X_i) = 1-π_i

Then E[Y_i | X_i] = 1(π_i) + 0(1-π_i) = π_i

Using the result from Logistic Mean Response Function:

E[Y_i | X_i] = P(Y_i=1| X_i) = π_i

$$= \frac{e^{\beta_0+\beta_1(X_i=1)+\beta_2(X_i=2)}}{1+e^{\beta_0+\beta_1(X_i=1)+\beta_2(X_i=2)}} \qquad (1.1)$$

Where $(X_i = k) = \begin{cases} 1 \ \ if \ \ X_i = k \\ 0 \ \ otherwise \end{cases}$

So, $P(Y_i = 1 | X_i = 1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$ *denoted by* η_1

$P(Y_i = 1 | X_i = 2) = \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}$ *denoted by* η_2

And $P(Y_i = 1 | X_i = 3) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$ *denoted by* η_0

The distribution of Y_i given X_i at any particular value is then,

$$f_{Y_i|X_i=1}(y_i | x_i = 1) = P(Y_i = 1 | X_i = 1)^{y_i} \left(1 - P(Y_i = 1 | X_i = 1)\right)^{1-y_i} = \eta_1^{y_i}[1-\eta_1]^{1-y_i}$$

$$f_{Y_i|X_i=2}(y_i | x_i = 2) = P(Y_i = 1 | X_i = 2)^{y_i} \left(1 - P(Y_i = 1 | X_i = 2)\right)^{1-y_i} = \eta_2^{y_i}[1-\eta_2]^{1-y_i} \qquad (1.2)$$

$$f_{Y_i|X_i=3}(y_i | x_i = 3) = P(Y_i = 1 | X_i = 3)^{y_i} \left(1 - P(Y_i = 1 | X_i = 3)\right)^{1-y_i} = \eta_0^{y_i}[1-\eta_0]^{1-y_i}$$

We now start with the observed data:

$$P(Y_i | W_i) = \frac{P(Y_i, W_i)}{P(W_i)} = \frac{\sum_{j=1}^{3} P(Y_i, W_i, X_i = j)}{P(W_i)}$$

$$= \frac{\sum_{j=1}^{3} P(Y_i | W_i, X_i = j) P(W_i, X_i = j)}{P(W_i)}$$

$$= \frac{\sum_{j=1}^{3} P(Y_i | X_i = j) P(W_i, X_i = j)}{P(W_i)}$$ by non-differential error assumption

$$= \frac{\sum_{j=1}^{3} P(Y_i | X_i = j) P(X_i = j | W_i) P(W_i)}{P(W_i)}$$

$$= \sum_{j=1}^{3} P(Y_i | X_i = j) P(X_i = j | W_i) \qquad (1.3)$$

And the distribution of Y_i given W_i in term of parameters $(\underset{\sim}{\beta}, \underset{\sim}{P})$ takes on the form:

$$f_i(y_i | w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \sum_{j=1}^{3} f_{Y_i | X_i = j}(y_i | x_i = j) P(X_i = j | W_i) \qquad (1.4)$$

$$= \eta_1^{y_i} (1-\eta_1)^{1-y_i} P(X_i = 1 | W_i) + \eta_2^{y_i} (1-\eta_2)^{1-y_i} P(X_i = 2 | W_i) + \eta_0^{y_i} (1-\eta_0)^{1-y_i} P(X_i = 3 | W_i)$$

The Log-Likelihood function is then,

$$LnL(\underset{\sim}{\beta}, \underset{\sim}{P}) = \sum_{i=1}^{N} L\text{n } f_i(y_i | w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \qquad (1.5)$$

$$= \sum_{i=1}^{N} Ln\left\{\eta_1^{y_i} (1-\eta_1)^{1-y_i} P(X_i = 1 | W_i) + \eta_2^{y_i} (1-\eta_2)^{1-y_i} P(X_i = 2 | W_i) + \eta_0^{y_i} (1-\eta_0)^{1-y_i} P(X_i = 3 | W_i)\right\}$$

Without lost of generality, suppose there are

$$\begin{cases} n_{1k} \ i's \ with \ y_i = 1 \ given \ w_i = k, \ 1 \le k \le 3 \\ \quad n_{0k} \ i's \ with \ y_i = 0 \ given \ w_i = k \end{cases}$$

Note that: $n_{1k} + n_{0k} = n_k \ and \ n_1 + n_2 + n_3 = N$

Then the Log-Likelihood function becomes,

$$LnL(\underset{\sim}{\beta}, \underset{\sim}{P}) = n_{11}Ln(\eta_1 p_{11} + \eta_2 p_{21} + \eta_0 p_{31}) + (n_1 - n_{11})Ln[(1-\eta_1)p_{11} + (1-\eta_2)p_{21} + (1-\eta_0)p_{31}]$$

$$+ \ n_{12}Ln(\eta_1 p_{12} + \eta_2 p_{22} + \eta_0 p_{32}) + (n_2 - n_{12})Ln[(1-\eta_1)p_{12} + (1-\eta_2)p_{22} + (1-\eta_0)p_{32}]$$

$$+ \ n_{13}Ln(\eta_1 p_{13} + \eta_2 p_{23} + \eta_0 p_{33}) + (n_3 - n_{13})Ln[(1-\eta_1)p_{13} + (1-\eta_2)p_{23} + (1-\eta_0)p_{33}]$$

$$= n_{11}Lna_1 + (n_1 - n_{11})Ln(1-a_1) + n_{12}Lna_2 + (n_2 - n_{12})Ln(1-a_2) + n_{13}Lna_3 + (n_3 - n_{13})Ln(1-a_3) \tag{1.6}$$

$where, \ a_i = \eta_1 p_{1i} + \eta_2 p_{2i} + \eta_0 p_{3i}, \ 1 \le i \le 3, and \ \sum_{l=1}^{3} p_{li} = 1$

Or the likelihood function could be written as

$$L(\underset{\sim}{\beta}, \underset{\sim}{P}) = a_1^{n_{11}}(1-a_1)^{n_1-n_{11}} a_2^{n_{12}}(1-a_2)^{n_2-n_{12}} a_3^{n_{13}}(1-a_3)^{n_3-n_{13}} \tag{1.7}$$

Following the notation from Hoadley (1971), whose result will be used in the proofs of consistency and asymptotic normality, we write:

$$L(\underset{\sim}{\beta}, \underset{\sim}{P}) = L_N(\underset{\sim}{\beta}, \underset{\sim}{P}) = \prod_{i=1}^{N} f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

Let $\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \ = \ Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$;

$$\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \tfrac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \tfrac{\partial}{\partial \underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

The $\underset{\sim}{\beta} - Score$ function is $S_N(\underset{\sim}{\beta}, \underset{\sim}{P}) = \sum_{i=1}^{N} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ and (1.8)

the $\underset{\sim}{\beta} - Hessian$ function is $H_N(\underset{\sim}{\beta}, \underset{\sim}{P}) = \sum_{i=1}^{N} \ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ (1.9)

1.4 Non-Identifiability of the $(\underset{\sim}{\beta}, \underset{\sim}{P})$ Likelihood Function

Whether a consistent estimator for a parameter of interest exists, we need to determine if the parameter in the model is identifiable.

Definition (of identifiability):

Given a model with likelihood function $L(\underset{\sim}{\beta})$, the parameter $\underset{\sim}{\beta}$ is said to be identifiable if for each $\underset{\sim}{\beta}_1 \neq \underset{\sim}{\beta}_2$ in the parameter space we have $L(\underset{\sim}{\beta}_1) \neq L(\underset{\sim}{\beta}_2)$. The following lemma will show the parameter $\underset{\sim}{P}$ from the above likelihood function is not identifiable.

Lemma 1: The parameter $\underset{\sim}{P}$ in model of equation (1.7) is not identifiable.

Proof:

In order to prove the non-identifiability of $\underset{\sim}{P}$ we need to find $\underset{\sim}{P} \neq \underset{\sim}{P}^*$ such that

$L(\underset{\sim}{\beta}, \underset{\sim}{P}) = L(\underset{\sim}{\beta}, \underset{\sim}{P}^*)$

If $L(\underset{\sim}{\beta}, \underset{\sim}{P}) = L(\underset{\sim}{\beta}, \underset{\sim}{P}^*)$, then from equation (1.7)

$$a_1^{n_{11}}(1-a_1)^{n_1-n_{11}} a_2^{n_{12}}(1-a_2)^{n_2-n_{12}} a_3^{n_{13}}(1-a_3)^{n_3-n_{13}} = (a_1^*)^{n_{11}}(1-a_1^*)^{n_1-n_{11}}(a_2^*)(1-a_2^*)^{n_2-n_{12}}(a_3^*)(1-a_3^*)^{n_3-n_{13}}$$

Or $\begin{cases} a_1 = a_1^* \\ a_2 = a_2^* \\ a_3 = a_3^* \end{cases} \Leftrightarrow \begin{cases} \eta_1 p_{1i} + \eta_2 p_{2i} + \eta_0 p_{3i} = \eta_1 p_{1i}^* + \eta_2 p_{2i}^* + \eta_0 p_{3i}^* \\ \quad for \\ 1 \le i \le 3 \end{cases}$

Take $i = 1$, then $\eta_1 p_{11} + \eta_2 p_{21} + \eta_0 p_{31} = \eta_1 p_{11}^* + \eta_2 p_{21}^* + \eta_0 p_{31}^*$

$$\Leftrightarrow \eta_1(p_{11} - p_{11}^*) + \eta_2(p_{21} - p_{21}^*) + \eta_0(p_{31} - p_{31}^*) = 0$$

Let $d_i = p_{i1} - p_{i1}^*$, $1 \le i \le 3$ and the fact $\sum_{l=1}^{3} p_{li} = \sum_{l=1}^{3} p_{li}^* = 1$ we have:

$$\begin{cases} \eta_1 d_1 + \eta_2 d_2 + \eta_0 d_3 = 0 \\ d_1 + d_2 + d_3 = 0 \\ -1 \le d_i \le 1 \\ 0 < \eta_i < 1, \ 1 \le i \le 3 \end{cases}$$

This is a system of two equations with three unknowns and we would always come up with more than one solution. Eliminating d_3 , then

$(\eta_1 - \eta_0)d_1 + (\eta_2 - \eta_0)d_2 = 0$ or $d_2 = -\dfrac{(\eta_1 - \eta_0)}{(\eta_2 - \eta_0)} d_1$ with $\eta_2 \neq \eta_0$

Suppose $p_{11} = 0.7, p_{21} = 0.2, p_{31} = 0.1$; let $c = \min\{1 - p_{11}, p_{31}\}$ = $\min\{1 - 0.7, 0.1\}$ = 0.1

Then $\begin{cases} d_1 = -\dfrac{c}{2} = -0.05 \\ p_{11}^* = p_{11} - d_1 = 0.7 - (-0.05) = 0.75 \end{cases}$; $\begin{cases} d_2 = -\dfrac{d_1}{2} = -\dfrac{-0.05}{2} = 0.025 \\ p_{21}^* = p_{21} - d_2 = 0.2 - 0.025 = 0.175 \end{cases}$

$$\begin{cases} d_3 = -(d_1 + d_2) = -(-0.05 + 0.025) = 0.025 \\ p_{31}^* = p_{31} - d_3 = 0.1 - 0.025 = 0.075 \end{cases}$$

Note that $p_{11}^* + p_{21}^* + p_{31}^* = 1$, $\underset{\sim}{P} \neq \underset{\sim}{P}^*$ and the proof is completed.

CHAPTER 2

THE PROPOSED LOGISTIC MODEL

2.1 The Auxiliary Sample

Having shown the parameter $\underset{\sim}{P}$ is not identifiable, we propose, besides the primary data, we also have an auxiliary sample of data which both X and its misclassified version W are observed for sample size M. We use this auxiliary data to estimate $\underset{\sim}{P}$ which is denoted by $\hat{\underset{\sim}{P}}_M$ and is computed as follows:

$$\hat{p}_{11} = estimated\ of\ P(X_i = 1 | W_i = 1) = \frac{Number\ of\ X_i=1\ given\ W_i=1}{Number\ of\ W_i=1}$$

$$\hat{p}_{21} = estimated\ of\ P(X_i = 2 | W_i = 1) = \frac{Number\ of\ X_i=2\ given\ W_i=1}{Number\ of\ W_i=1}$$

$$\hat{p}_{31} = estimated\ of\ P(X_i = 3 | W_i = 1) = \frac{Number\ of\ X_i=3\ given\ W_i=1}{Number\ of\ W_i=1}$$

$$\hat{p}_{12} = estimated\ of\ P(X_i = 1 | W_i = 2) = \frac{Number\ of\ X_i=1\ given\ W_i=2}{Number\ of\ W_i=2}$$

$$\hat{p}_{22} = estimated\ of\ P(X_i = 2 | W_i = 2) = \frac{Number\ of\ X_i=2\ given\ W_i=2}{Number\ of\ W_i=2}$$

$$\hat{p}_{32} = estimated\ of\ P(X_i = 3 | W_i = 2) = \frac{Number\ of\ X_i=3\ given\ W_i=2}{Number\ of\ W_i=2}$$

$$\hat{p}_{13} = estimated\ of\ P(X_i = 1 | W_i = 3) = \frac{Number\ of\ X_i=1\ given\ W_i=3}{Number\ of\ W_i=3}$$

$$\hat{p}_{23} = estimated\ of\ P(X_i = 2 | W_i = 3) = \frac{Number\ of\ X_i=2\ given\ W_i=3}{Number\ of\ W_i=3}$$

$$\hat{p}_{33} = estimated\ of\ P(X_i = 3 | W_i = 3) = \frac{Number\ of\ X_i=3\ given\ W_i=3}{Number\ of\ W_i=3}$$

With sample size $M \to \infty$, it is known that $\hat{\underset{\sim}{P}}_M \xrightarrow{p} \underset{\sim}{P}$, where $\underset{\sim}{P}$ is the true misclassification matrix between X and W.

2.2 The Approximate Likelihood Function

Now we use the estimated $\hat{\underset{\sim}{P}}_M$ to plug into the likelihood function to get an "approximate" likelihood function:

$$L^{(2)}(\underset{\sim}{\beta}) = L(\underset{\sim}{\beta}, \hat{\underset{\sim}{P}}_M) = \prod_{i=1}^{N} f_i(y_i \mid w_i, \underset{\sim}{\beta}, \hat{\underset{\sim}{P}}_M) \quad (2.1)$$

$$= \hat{a}_1^{n_{11}} (1-\hat{a}_1)^{n_1 - n_{11}} \hat{a}_2^{n_{12}} (1-\hat{a}_2)^{n_2 - n_{12}} \hat{a}_3^{n_{13}} (1-\hat{a}_3)^{n_3 - n_{13}}$$

$$where,\ \hat{a}_i = \eta_1 \hat{p}_{1i} + \eta_2 \hat{p}_{2i} + \eta_0 \hat{p}_{3i},\ 1 \le i \le 3,\ and\ \sum_{l=1}^{3} \hat{p}_{li} = 1$$

We will try to estimate $\underset{\sim}{\beta}$ such that $LnL^{(2)}(\underset{\sim}{\beta})$ maximized. In doing so, we solve for $\hat{\underset{\sim}{\beta}}$ of the approximate score vector function

$$\underset{\sim}{S}^{(2)}_{N,M}(\hat{\underset{\sim}{\beta}}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} LnL^{(2)}(\hat{\underset{\sim}{\beta}}) = \left(\tfrac{\partial}{\partial \beta_0} LnL^{(2)}(\hat{\underset{\sim}{\beta}}), \tfrac{\partial}{\partial \beta_1} LnL^{(2)}(\hat{\underset{\sim}{\beta}}), \tfrac{\partial}{\partial \beta_2} LnL^{(2)}(\hat{\underset{\sim}{\beta}}) \right)^T \quad (2.2)$$

$$= \underset{\sim}{0}$$

We call the resulting solution an "approximate likelihood estimator" and is denoted by $\hat{\underset{\sim}{\beta}}_{N,M}$.

2.3 Large Sample Performance of $\hat{\underset{\sim}{\beta}}_{N,M}$

Let $\underset{\sim}{\beta}_0$ be the true value of parameter $\underset{\sim}{\beta}$. Then under the assumptions listed in section 3.2, we can show the following result when sample sizes N and M are large.

Theorem 1: $\hat{\underset{\sim}{\beta}}_{N,M} \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N, M \to \infty$

Theorem 2: $\sqrt{N}(\hat{\underset{\sim}{\beta}}_{N,M} - \underset{\sim}{\beta}_0) \xrightarrow{L} N(\underset{\sim}{0}, \overline{\Gamma}^{-1}(\underset{\sim}{\beta}_0))$ as $N, M \to \infty$ **,**

$$where\ \overline{\Gamma}(\underset{\sim}{\beta}_0) = \lim_{N \to \infty} \left(-\tfrac{1}{N} \sum_{i=1}^{N} E[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P}) \mid \underset{\sim}{\beta}_0] \right)$$

2.4 Monte Carlo Study of the Performance of $\hat{\underset{\sim}{\beta}}_{N,M}$

Table 1 to 4 show the result of simulation for different combination of primary sample size N and auxiliary sample size M. In table 1, the true parameter values $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (0, 1, -1.5)$, together with misclassification probability matrix $\underset{\sim}{P} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$ used to generate data for Y and W in primary data set and W and X in auxiliary data set. In table 2, the same $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (0, 1, -1.5)$ as in table 1 is used but the misclassification probability matrix now is

$\underset{\sim}{P} = \begin{bmatrix} 0.8 & 0.15 & 0.0 \\ 0.2 & 0.7 & 0.2 \\ 0.0 & 0.15 & 0.8 \end{bmatrix}$. The same setup is used for table 3 and 4 but now the true parameter values $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (1, 1, -1.5)$

For each experiment, 500 replications are run. First we estimate the misclassification probability $\hat{\underset{\sim}{P}}_M$ of $\underset{\sim}{P}$ using the auxiliary data. Parameter $\underset{\sim}{\beta} = (\beta_0, \beta_1, \beta_2)$ is estimated by maximizing the approximate likelihood function using function call NLPTR (…) in SAS which implements the Trust Region Method for nonlinear optimization and denoted by $\hat{\underset{\sim}{\beta}}_{N,M}$. The estimated parameter is the average over 500 estimates above. The estimated asymptotic variance is the average over 500 asymptotic variances estimated as the diagonal elements of the negative inverse Hessian matrix, $-\hat{H}_N^{-1}(\hat{\underset{\sim}{\beta}}_{N,M})$ where $\hat{H}_N(\underset{\sim}{\beta}) = \sum_{i=1}^{N} \ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \hat{\underset{\sim}{P}}_M)$.

In those tables, the simulation results reported are parameter estimated, bias from true value, asymptotic standard error, lower and upper limits of 95% confidence interval. By looking at the tables, in general we see that:

1. The bias is small in for large enough sample size N and M.
2. Larger sample sizes can predict the true parameter with slightly more accuracy.
3. Standard errors are clearly reduced with larger sample sizes.
4. The more misclassification between X and W, the bigger sample sizes N and M are needed in order to achieve the same accuracy.

Table 1: Monte Carlo Simulation Study with $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (0, 1, -1.5)$

N	M	Para.	Estimated	Bias	Std Error	Lower	Upper
200	100	β_0	-0.00796	-0.00796	0.27844	-0.5537	0.53777
		β_1	1.06547	0.06547	0.46454	0.15498	1.97597
		β_2	-1.55625	-0.05625	0.53045	-2.5959	-0.5165
200	200		-0.0027	-0.0027	0.27772	-0.5471	0.54152
			1.02079	0.02079	0.45599	0.12705	1.91452
			-1.5317	-0.0317	0.52596	-2.5653	-0.4980
500	100		0.00759	0.00759	0.17548	-0.3363	0.35153
			1.00441	0.00441	0.28688	0.44214	1.56669
			-1.5329	-0.0329	0.31953	-2.1591	-0.9066
500	200		0.00264	0.00264	0.17513	-0.3406	0.34588
			0.99978	-0.0002	0.28643	0.43838	1.56118
			-1.5550	-0.0550	0.32575	-2.1935	-0.91654
500	500		0.00730	0.00730	0.17513	-0.3359	0.35055
			1.00035	0.00035	0.28492	0.44191	1.55878
			-1.51511	-0.0151	0.31929	-2.1109	-0.8893
1000	100		-0.0082	-0.0082	0.12403	-0.2513	0.23484
			1.01493	0.01493	0.20241	0.61821	1.41165
			-1.5038	-0.0038	0.22683	-1.9484	-1.0593
1000	200		0.00250	0.00250	0.12341	-0.2393	0.24438
			1.00888	0.00888	0.20026	0.61637	1.40139
			-1.5087	-0.0087	0.22485	-1.9494	-1.0680
1000	500		-0.0030	-0.0030	0.12304	-0.2442	0.23808
			1.01706	0.01706	0.20011	0.62483	1.40928
			-1.4962	0.0037	0.22394	-1.9351	-1.0573

Table 2: Monte Carlo Simulation Study with $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (0, 1, -1.5)$

N	M	Para.	Estimated	Bias	Std Error	Lower	Upper
200	100	β_0	-0.0048	-0.0048	0.35661	-0.7038	0.69406
		β_1	1.01686	0.01686	0.53144	-0.0247	2.05847
		β_2	-1.6288	-0.1288	1.23977	-4.0587	0.80108
200	200		-0.0068	-0.0068	0.34846	-0.6897	0.67615
			1.03761	0.03761	0.52751	0.00370	2.07153
			-1.5992	-0.0992	1.08720	-3.7301	0.53163
500	100		-0.0027	-0.0027	0.22003	-0.4340	0.42846
			1.00907	0.00907	0.32262	0.37674	1.64139
			-1.5798	-0.0798	0.63327	-2.8210	-0.33861
500	200		-0.0116	-0.0116	0.21734	-0.4376	0.41432
			1.02265	0.02265	0.31788	0.39961	1.64568
			-1.5326	-0.0326	0.59357	-2.6960	-0.36929
500	500		-0.0137	-0.0137	0.21352	-0.4322	0.40476
			1.03301	0.03301	0.31626	0.41314	1.65289
			-1.5435	-0.0435	0.59643	-2.7124	-0.3745
1000	100		-0.0153	-0.0153	0.15614	-0.3213	0.29070
			1.02273	0.02273	0.22647	0.57885	1.46662
			-1.5086	-0.0086	0.42020	-2.3322	-0.6850
1000	200		-0.0031	-0.0031	0.15348	-0.3039	0.29770
			1.00028	0.00028	0.22157	0.56599	1.43457
			-1.5171	-0.0171	0.40416	-2.3092	-0.7249
1000	500		0.00071	0.00071	0.15139	-0.2960	0.29744
			1.00539	0.00539	0.22083	0.57257	1.43822
			-1.5363	-0.0363	0.39142	-2.3035	-0.7691

Table 3: Monte Carlo Simulation Study with $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (1, 1, -1.5)$

N	M	Para.	Estimated	Bias	Std Error	Lower	Upper
200	100	β_0	1.02170	0.02170	0.32147	0.39163	1.65178
		β_1	1.00738	0.00738	0.64514	-0.2570	2.27185
		β_2	-1.5019	-0.0019	0.45615	-2.3959	-0.6078
200	200		1.04882	0.04882	0.32608	0.40970	1.68795
			1.05956	0.05956	0.67877	-0.2708	2.38994
			-1.5700	-0.0700	0.46445	-2.4803	-0.6596
500	100		1.01139	0.01139	0.20170	0.61606	1.40673
			0.99014	0.00985	0.38778	0.23009	1.75019
			-1.5247	-0.0247	0.28908	-2.0913	-0.9581
500	200		1.01431	0.01431	0.20177	0.61884	1.40978
			1.00146	0.00146	0.39247	0.23222	1.77069
			-1.5108	-0.0108	0.28727	-2.07391	-0.9478
500	500		1.00893	0.00893	0.19998	0.61714	1.40072
			1.04417	0.04417	0.39163	0.27658	1.81176
			-1.5180	-0.0180	0.28546	-2.0775	-0.9585
1000	100		1.00753	0.00753	0.14239	0.72844	1.28661
			1.03769	0.03769	0.27614	0.49645	1.57893
			-1.5096	-0.0096	0.20330	-1.9080	-1.1111
1000	200		1.01106	0.01106	0.14187	0.73299	1.28913
			1.01084	0.01084	0.27060	0.48047	1.54121
			-1.5181	-0.0181	0.20231	-1.9147	-1.1216
1000	500		1.00582	0.00582	0.14109	0.72928	1.28237
			1.01744	0.01744	0.27009	0.48806	1.54683
			-1.5090	-0.0090	0.20145	-1.9038	-1.1141

Table 4: Monte Carlo Simulation Study with $\underset{\sim}{\beta}_0 = \left(\beta_0^0, \beta_1^0, \beta_2^0\right) = (1, 1, -1.5)$

N	M	Para.	Estimated	Bias	Std Error	Lower	Upper
200	100	β_0	1.04226	0.04226	0.45934	0.14195	1.94256
		β_1	0.99773	0.00227	0.99249	-0.9475	2.94297
		β_2	-1.5114	-0.0114	0.80459	-3.0884	0.06548
200	200		1.03455	0.03455	0.46016	0.13264	1.93645
			1.03378	0.03378	0.95244	-0.8329	2.90053
			-1.5193	-0.0193	0.79424	-3.0759	0.03738
500	100		1.02525	0.02525	0.27839	0.47961	1.57089
			1.05885	0.05885	0.58101	-0.0799	2.19768
			-1.5228	-0.0228	0.48850	-2.4802	-0.5653
500	200		1.01959	0.01959	0.27297	0.48457	1.55462
			1.02683	0.02683	0.53929	-0.0301	2.08384
			-1.5181	-0.0181	0.48241	-2.4636	-0.5725
500	500		1.00628	0.00628	0.26738	0.48222	1.53034
			1.07148	0.07148	0.54185	0.00945	2.13350
			-1.5243	-0.0243	0.47098	-2.4474	-0.6012
1000	100		1.01736	0.01736	0.19164	0.64176	1.39297
			1.02245	0.02245	0.38393	0.26995	1.77496
			-1.5145	-0.0145	0.33694	-2.1749	-0.8542
1000	200		1.00431	0.00431	0.18748	0.63685	1.37177
			1.02833	0.02833	0.35276	0.33693	1.71974
			-1.5011	-0.0011	0.33085	-2.1496	-0.8527
1000	500		1.00401	0.00401	0.18634	0.63878	1.36925
			1.03490	0.03490	0.35345	0.34214	1.72766
			-1.5202	-0.0202	0.32695	-2.1610	-0.7894

CHAPTER 3

PROOFS OF RESULTS

3.1 Notations and Preliminaries

From the likelihood function (equation 1.7), we introduce the "artificial" likelihood function: $L^{(1)}(\underset{\sim}{\beta}) = L(\underset{\sim}{\beta}, \underset{\sim}{P})$, where $\underset{\sim}{P}$ is the misclassification matrix.

Define $\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta}) = \frac{\partial}{\partial \underset{\sim}{\beta}} LnL^{(1)}(\underset{\sim}{\beta})$. Let $\hat{\underset{\sim}{\beta}}_N$ be such that $\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N) = \underset{\sim}{0}$. Then we call this $\hat{\underset{\sim}{\beta}}_N$ an "artificial" likelihood estimator of the model.

From the $\underset{\sim}{\beta} - Hessian$ matrix (equation 1.9) $H_N(\underset{\sim}{\beta}, \underset{\sim}{P}) = \sum_{i=1}^{N} \ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$,

we define two related Hessian matrices: $H_N(\underset{\sim}{\beta}) = \sum_{i=1}^{N} \ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$

and $\hat{H}_N(\underset{\sim}{\beta}) = \sum_{i=1}^{N} \ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \hat{\underset{\sim}{P}}_M)$.

The notation below used by Hoadley (1971) will be used in the proofs of consistency and asymptotic normality of the "artificial" likelihood estimator $\hat{\underset{\sim}{\beta}}_N$:

$$R_i(\underset{\sim}{\beta}) = \begin{cases} Ln \dfrac{f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})}{f_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P})} & if \quad f_i(y_i \mid w_i, \underset{\sim}{\beta}_0) > 0 \\ 0 \; otherwise; \end{cases}$$

$$R_i(\underset{\sim}{\beta}, \rho) = \sup\left\{ R_i(\underset{\sim}{\beta}^*) : \left\| \underset{\sim}{\beta}^* - \underset{\sim}{\beta} \right\| < \rho \right\}$$

$$V_i(r) = \sup\left\{R_i(\underset{\sim}{\beta}) : \left\|\underset{\sim}{\beta}\right\| > r\right\}$$

$$r_i(\underset{\sim}{\beta}) = E_{\underset{\sim}{\beta}_0}\left\{R_i(\underset{\sim}{\beta})\right\}$$

$$\bar{r}_N(\underset{\sim}{\beta}) = \frac{1}{N}\sum_{i=1}^{N} r_i(\underset{\sim}{\beta})$$

$$v_i(r) = E_{\underset{\sim}{\beta}_0}\left\{V_i(r)\right\}$$

$$\bar{v}_N(r) = \frac{1}{N}\sum_{i=1}^{N} v_i(r)$$

3.2 Assumptions

To prove consistency and asymptotic normality of $\hat{\underset{\sim}{\beta}}_{N,M}$ **,** we assume the following:

Assumption A1: $\hat{\underset{\sim}{P}}_M \xrightarrow{p} \underset{\sim}{P}$ as $M \to \infty$, where M is the auxiliary sample size and $\underset{\sim}{P}$ is the true misclassification matrix.

Assumption A2: $\frac{N}{\sqrt{M}} \leq K < \infty \ as \ N, M \to \infty$; where N is the primary sample size and M is the auxiliary sample size.

Assumption A3: Θ is a compact subset of R^p .

Assumption A4: $\underset{\sim}{\beta}_0 \in \mathrm{int}(\Theta)$.

Assumption A5: $\underset{\sim}{P}$ is nonsingular.

3.3 Consistency of $\hat{\underset{\sim}{\beta}}_{N,M}$

Theorem 1: Under assumptions A1 and A3, $\hat{\underset{\sim}{\beta}}_{N,M} \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N, M \to \infty$

Theorem 1 will be proved using result of the following two lemmas and application of triangle inequality:

a) **Lemma 2:** $\hat{\underset{\sim}{\beta}}_{M,N} \xrightarrow{p} \hat{\underset{\sim}{\beta}}_N$ as $N, M \to \infty$

b) Lemma 3: $\hat{\underset{\sim}{\beta}}_N \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N \to \infty$

Proof of Lemma 2:

The following claims M1, M4, and M5, which are proved in appendix B, are used for lemma 2.

Claim M1: $\max_{\underset{\sim}{\beta}\in\Theta}\left[\frac{1}{N}\left\|\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta}) - \underset{\sim}{S}_{N,M}^{(2)}(\underset{\sim}{\beta})\right\|\right] \xrightarrow{p} 0$ as $N, M \to \infty$.

Claim M4: If $x_n^T A_n x_n \xrightarrow{p} 0$ and $\lambda_{\min}(A_n) \geq \delta > 0,\ \forall n \geq N$, then $x_n \xrightarrow{p} 0$.

Claim M5: For any $\underset{\sim}{\beta} \in \Theta$, $\lambda_{\min}(-\frac{1}{n}H_n(\underset{\sim}{\beta})) \geq \delta > 0,\ \forall n \geq N$.

First note that $\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N) = \underset{\sim}{S}_{N,M}^{(2)}(\hat{\underset{\sim}{\beta}}_{N,M}) = \underset{\sim}{0}$. Together with claim M1 we have:

$$\left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N) - \underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right\| \leq \left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N) - \underset{\sim}{S}_{N,M}^{(2)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right\| + \left\|\frac{1}{N}\left\{\underset{\sim}{S}_{N,M}^{(2)}(\hat{\underset{\sim}{\beta}}_{N,M}) - \underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right\|$$

$$= \left\|\frac{1}{N}\left\{\underset{\sim}{S}_{N,M}^{(2)}(\hat{\underset{\sim}{\beta}}_{N,M}) - \underset{\sim}{S}_{N,M}^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right\| \tag{3.1}$$

$$\leq \max_{\underset{\sim}{\beta}\in\Theta}\left[\frac{1}{N}\left\|\underset{\sim}{S}_{N,M}^{(2)}(\underset{\sim}{\beta}) - \underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})\right\|\right] \xrightarrow{p} 0 \text{ as } N, M \to \infty .$$

By Mean Value Theorem,

$$\left(\hat{\underset{\sim}{\beta}}_{N,M} - \hat{\underset{\sim}{\beta}}_N\right)^T \left[\frac{1}{N}\left\{\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N) - \underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right] = \left(\hat{\underset{\sim}{\beta}}_{N,M} - \hat{\underset{\sim}{\beta}}_N\right)^T \left[\frac{1}{N}\frac{\partial}{\partial \underset{\sim}{\beta}}\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta}_N^*)\right]\left(\hat{\underset{\sim}{\beta}}_{N,M} - \hat{\underset{\sim}{\beta}}_N\right)$$

$= \left(\hat{\underset{\sim}{\beta}}_{N,M} - \hat{\underset{\sim}{\beta}}_N\right)^T \left[\frac{1}{N}H_N(\underset{\sim}{\beta}_N^*)\right]\left(\hat{\underset{\sim}{\beta}}_{N,M} - \hat{\underset{\sim}{\beta}}_N\right)$, for some $\underset{\sim}{\beta}_N^*$ between

$\hat{\underset{\sim}{\beta}}_{N,M}$ *and* $\hat{\underset{\sim}{\beta}}_N$. Taking the norm of both sides we have:

$$\left|\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\left[\frac{1}{N}H_N(\underset{\sim}{\beta}_N^*)\right]\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)\right| \le \left\|\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\right\|\left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N)-\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right\|$$

$\xrightarrow{p} 0$ as $N,M\to\infty$ since the first factor is bounded and the second factor $\xrightarrow{p} 0$ as shown above.

Hence, $\left|\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\left[-\frac{1}{N}H_N(\underset{\sim}{\beta}_N^*)\right]\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)\right|\xrightarrow{p} 0$ as $N,M\to\infty$

By claim M5, $\lambda_{\min}(-\frac{1}{N}H_N(\underset{\sim}{\beta}_N^*))\ge\delta>0$ for some δ.

Now applying claim M4 we conclude $\left(\hat{\underset{\sim}{\beta}}_{M,N}-\hat{\underset{\sim}{\beta}}_N\right)\xrightarrow{p} 0$ as $N,M\to\infty$.

Therefore, Lemma 2 is proved.

Proof of Lemma 3:

Theorem 1 from Hoadley (1971) shows that $\hat{\underset{\sim}{\beta}}_N \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N\to\infty$

under the following five conditions:

C1. Θ is a closed subset of $\Re^p$

C2. $f_i(y_i\,|\,w_i,\underset{\sim}{\beta},\underset{\sim}{P})$ is an upper semicontinuous (u.s.c.) function of $\underset{\sim}{\beta}$, uniformly in i, a.s. [P].

C3. There exists $\rho^*=\rho^*(\underset{\sim}{\beta})>0$ and $r>0$ for which

i) $E[R_i(\underset{\sim}{\beta},\rho)]^2\le K_1,\ 0\le\rho\le\rho^*$;

ii) $E[V_i(r)]^2\le K_2$.

C4. i) $\lim \overline{r}_N(\underset{\sim}{\beta})<0,\ \underset{\sim}{\beta}\ne\underset{\sim}{\beta}_0$;

ii) $\lim \overline{v}_N(r)<0.$

C5. $R_i(\underset{\sim}{\beta},\rho)$ and $V_i(r)$ are measurable functions of Y_i

These five conditions are verified for our model. The proof's detail is given in Appendix C. Hence lemma 3 is proved.

3.4 Asymptotic Normality of $\hat{\underset{\sim}{\beta}}_{N,M}$

Theorem 2: Under assumptions A2, A4, and A5, $\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\underset{\sim}{\beta}_0\right)\xrightarrow{L} N\left(\underset{\sim}{0},\bar{\Gamma}^{-1}(\underset{\sim}{\beta}_0)\right)$ as $N, M \to \infty$.

Proof:

Theorem 2 is proved using the following two Lemmas:

Lemma 4: Under assumption A2, $\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)\xrightarrow{p} 0$ as $N, M \to \infty$.

Lemma 5: Under assumption A4, and A5, $\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_N-\underset{\sim}{\beta}_0\right)\xrightarrow{L} N\left(\underset{\sim}{0},\bar{\Gamma}^{-1}(\underset{\sim}{\beta}_0)\right)$ as $N \to \infty$.

The result follows because $\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\underset{\sim}{\beta}_0\right)=\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)+\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_N-\underset{\sim}{\beta}_0\right)$

Proof of Lemma 4:

Lemma 2 above together with the following claims M1 and M2, which are proved in appendix B, are used to prove lemma 4.

Claim M1: $\max_{\underset{\sim}{\beta}\in\Theta}\left[\frac{1}{N}\left\|\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})-\underset{\sim}{S}_{N,M}^{(2)}(\underset{\sim}{\beta})\right\|\right]\xrightarrow{p} 0$ as $N, M \to \infty$.

Claim M2: $\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial \underset{\sim}{P}}\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})\right\|=\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\underset{\sim}{P}}\frac{\partial}{\underset{\sim}{\beta}}Lnf_i(y_i \mid w_i,\underset{\sim}{\beta},\underset{\sim}{P})\right]\right\|$ is Cauchy in probability and hence is bounded in probability as $N \to \infty$.

$$\left|N\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\left(\frac{1}{N}\left\{\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_N)\right\}\right)\right|\le N\left\|\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\right\|\left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_N)\right\}\right\|$$

$$=\frac{N}{\sqrt{M}}\left\|\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\right\|\bullet\sqrt{M}\left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_N)\right\}\right\|$$

But $\sqrt{M}\left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N)\right\}\right\| \le \sqrt{M}\left\|\frac{1}{N}\left\{\underset{\sim}{S}_N^{(2)}(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_{N,M})\right\}\right\|$ by (3.1)

$$\le \sqrt{M}\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial \underset{\sim}{P}}\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N)\right\|\bullet\left\|\hat{\underset{\sim}{P}}_M-\underset{\sim}{P}\right\| \quad \text{by M1.1 (see Appendix B).}$$

$$=\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial \underset{\sim}{P}}\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N)\right\|\bullet\sqrt{M}\left\|\hat{\underset{\sim}{P}}_M-\underset{\sim}{P}\right\|$$

This is bounded in probability because $\sqrt{M}\left\|\hat{\underset{\sim}{P}}_M-\underset{\sim}{P}\right\|$ is bounded in probability and $\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial \underset{\sim}{P}}\underset{\sim}{S}_N^{(1)}(\hat{\underset{\sim}{\beta}}_N)\right\|$ is also bounded in probability by claim M2.

Hence $\left|N\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\left(\frac{1}{N}\left\{\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_N)\right\}\right)\right|\xrightarrow{p}0$ because

$\frac{N}{\sqrt{M}}$ is bounded by assumption A2 and $\left\|\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)\right\|\xrightarrow{p}0$ by lemma 2.

Moreover, by Mean Value Theorem:

$$\left|N\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\left(\frac{1}{N}\left\{\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_{N,M})-\underset{\sim}{S}_N^1(\hat{\underset{\sim}{\beta}}_N)\right\}\right)\right|=\left|\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)^T\left(\frac{1}{N}\underset{\sim}{H}_N(\underset{\sim}{\beta}_N^*)\right)\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)\right|$$

Where $\underset{\sim}{\beta}_N^*$ is between $\hat{\underset{\sim}{\beta}}_{N,M}$ and $\hat{\underset{\sim}{\beta}}_N$.

Similar argument used in Lemma 2 show that $\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_{N,M}-\hat{\underset{\sim}{\beta}}_N\right)\xrightarrow{p}0$.

Proof of Lemma 5:

Theorem 2 from Hoadley (1971) shows $\sqrt{N}\left(\hat{\underset{\sim}{\beta}}_N-\underset{\sim}{\beta}_0\right)\xrightarrow{L}N\left(\underset{\sim}{0},\bar{\Gamma}^{-1}(\underset{\sim}{\beta}_0)\right)$

under the following nine conditions:

Condition N1: The interior of Θ is an open set of R^p and $\underset{\sim}{\beta}_0$ is an interior point of Θ.

Condition N2: $\hat{\underset{\sim}{\beta}}_N \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N \to \infty$

Condition N3:

i) $\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ exists.

ii) $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ exists.

Condition N4: $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is a continuous function of $\underset{\sim}{\beta}$, uniformly in i, a.s. [P], and is measurable functions of Y_i

Condition N5: $E\left[\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right] = \underset{\sim}{0}; \; i = 1, 2, ...$

Condition N6:

$$\Gamma_i(\underset{\sim}{\beta}) = E\left[\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})^T \mid \underset{\sim}{\beta}\right] = -E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]$$

Condition N7: $\exists \bar{\Gamma}(\underset{\sim}{\beta})$ such that $\lim_{N \to \infty} \bar{\Gamma}_N(\underset{\sim}{\beta}) = \bar{\Gamma}(\underset{\sim}{\beta})$, and $\bar{\Gamma}(\underset{\sim}{\beta})$ is positive definite.

Condition N8: $E\left[\dot{\Phi}_{i,j}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]^3 \le K$

Condition N9: There exists $\varepsilon > 0$ and random variables $B_{i,jl}(Y_i)$ such that

i) $\sup\left\{\left|\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| : \left\|\underset{\sim}{\beta} - \underset{\sim}{\beta}_0\right\| \le \varepsilon\right\} \le B_{i,jl}(Y_i)$.

ii) $E\left[B_{i,jl}(Y_i)\right]^{1+\delta} \le K$, for some $\delta > 0, K > 0$.

These nine conditions are verified in Appendix D for our model. Therefore, Lemma 5 is proved.

CHAPTER 4

EXTENSION AND DISCUSSION

4.1 Model for Including Continuous Variables

Let Z be an error-free variable that can be observed independently of X. Then:

$$P(Y_i \mid W_i, Z_i) = \frac{P(Y_i, W_i, Z_i)}{P(W_i, Z_i)} = \frac{\sum_{j=1}^{3} P(Y_i, W_i, Z_i, X_i = j)}{P(W_i, Z_i)}$$

$$= \frac{\sum_{j=1}^{3} P(Y_i \mid W_i, X_i = j, Z_i) P(W_i, Z_i, X_i = j)}{P(W_i, Z_i)}$$

$$= \frac{\sum_{j=1}^{3} P(Y_i \mid Z_i, X_i = j) P(W_i, Z_i, X_i = j)}{P(W_i, Z_i)}$$ by non-differential error assumption.

$$= \frac{\sum_{j=1}^{3} P(Y_i \mid Z_i, X_i = j) P(X_i = j \mid W_i, Z_i) P(W_i, Z_i)}{P(W_i, Z_i)}$$

$$= \sum_{j=1}^{3} P(Y_i \mid Z_i, X_i = j) P(X_i = j \mid W_i, Z_i)$$

$$= \sum_{j=1}^{3} P(Y_i \mid Z_i, X_i = j) P(X_i = j \mid W_i)$$ by independent assumption (4.1)

Comparing to equation (1.3) in chapter 1, equation (4.1) has similar form except now taking into account information of variable Z.

4.2 Model for Multiple Misclassified Variables

Let W_1 and W_2 be the misclassified version of unobserved variables X_1 and X_2 respectively that are independently observed.

$$P(Y_i \mid W_{1i}, W_{2i}) = \frac{P(Y_i, W_{1i}, W_{2i})}{P(W_{1i}, W_{2i})} = \frac{\sum_{j=1}^{K_1}\sum_{h=1}^{K_2} P(Y_i, W_{1i}, X_{1i} = j, W_{2i}, X_{2i} = l)}{P(W_{1i}, W_{2i})}$$

$$= \frac{\sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i, W_{1i}, X_{1i} = j \mid W_{2i}, X_{2i} = l) P(W_{2i}, X_{2i} = l)}{P(W_{1i}, W_{2i})}$$

$$= \frac{\sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i, W_{1i}, X_{1i} = j \mid X_{2i} = l) P(W_{2i}, X_{2i} = l)}{P(W_{1i}, W_{2i})}$$

$$= \frac{\sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i, W_{1i}, X_{1i} = j \mid X_{2i} = l) P(X_{2i} = l \mid W_{2i}) P(W_{2i})}{P(W_{1i}, W_{2i})}$$

$$= \frac{\sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i, W_{1i}, X_{1i} = j \mid X_{2i} = l) P(X_{2i} = l \mid W_{2i})}{P(W_{1i})}$$ by independent

$$= \frac{\sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i \mid W_{1i}, X_{1i} = j, X_{2i} = l) P(W_{1i}, X_{1i} = j \mid X_{2i} = h) P(X_{2i} = l \mid W_{2i})}{P(W_{1i})}$$

$$= \frac{\sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i \mid X_{1i} = j, X_{2i} = l) P(W_{1i}, X_{1i} = j) P(X_{2i} = l \mid W_{2i})}{P(W_{1i})}$$ by

non-differential error assumption and independent.

$$= \sum_{j=1}^{K_1}\sum_{l=1}^{K_2} P(Y_i \mid X_{1i} = j, X_{2i} = l) P(X_{1i} = j \mid W_{1i}) P(X_{2i} = l \mid W_{2i}) \qquad (4.2)$$

Again, equation (4.2) has similar form as that of (1.3) but now dealing with two variables.

In general,

$$P(Y_i | W_{1i}, W_{2i}, ..., W_{Di}) = \sum_{j=1}^{K_1} \sum_{l=1}^{K_2} ... \sum_{q=1}^{K_l} P(Y_i | X_{1i} = j, X_{2i} = l, ..., X_{Di} = q) P(X_{1i} = j | W_{1i}) P(X_{2i} = l | W_{2i}) ... P(X_{Di} = q | W_{Di})$$

4.3 Future Work

The implementation of this work is only for model of binary response variable and a misclassified discrete covariate. It would be interesting if we could extend the program to model of more than one misclassified discrete covariates, with the inclusion of continuous variables, and model for categorical response variable.

APPENDIX A

SOME USEFUL INEQUALITIES

C_r -Inequality: $(a+b)^r \leq C_r(a^r + b^r)$, where $a>0$, $b>0$, $C_r = \begin{cases} 1 \ if \quad \leq 1 \\ 2^{r-1} \ if \ \ r > 1 \end{cases}$

Markov Inequality: For a random variable X and positive constants $c>0,\ r>0$,

$$P\left[\left|X\right| \geq c\right] \leq \frac{E\left(\left|X\right|^r\right)}{c^r}$$

APPENDIX B

PROOFS OF MAIN RESULT

Claim M1: $\max_{\underset{\sim}{\beta}\in\Theta}\left[\frac{1}{N}\left\|\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})-\underset{\sim}{S}_{N,M}^{(2)}(\underset{\sim}{\beta})\right\|\right]\xrightarrow{p}0$ as $N,M\rightarrow\infty$.

Claim M2: $\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial\underset{\sim}{P}}\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})\right\|=\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\underset{\sim}{P}}\frac{\partial}{\underset{\sim}{\beta}}Lnf_i(y_i\,|\,w_i,\underset{\sim}{\beta},\underset{\sim}{P})\right]\right\|$ is Cauchy in probability and hence is bounded in probability as $N\rightarrow\infty$

Claim M3: $E\left[\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\frac{\partial}{\partial P_l}\left(\frac{\partial}{\partial\beta_j}Lnf_i(y_i\,|\,w_i,\underset{\sim}{\beta},\underset{\sim}{P})\right)\right]\le K<\infty$, uniformly in i;

Claim M4: If $x_n^T A_n x_n\xrightarrow{p}0$ and $\lambda_{\min}(A_n)\ge\delta>0,\ \forall n\ge N$, then $x_n\xrightarrow{p}0$.

Claim M5: For any $\underset{\sim}{\beta}\in\Theta$,

i) $\left\|-\frac{1}{n}H_n(\underset{\sim}{\beta})-\overline{\Gamma}_n(\underset{\sim}{\beta})\right\|\xrightarrow{p}0$ as $n\rightarrow\infty$

ii) $\lambda_{\min}(-\frac{1}{n}H_n(\underset{\sim}{\beta}))\ge\delta>0,\ \forall n\ge N$.

Claim M1: $\max_{\underset{\sim}{\beta}\in\Theta}\left[\frac{1}{N}\left\|\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})-\underset{\sim}{S}_{N,M}^{(2)}(\underset{\sim}{\beta})\right\|\right]\xrightarrow{p}0$ as $N,M\to\infty$.

Proof of M1:

Define $\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P})=\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})$ *and* $\underset{\sim}{G}(\underset{\sim}{\beta},\hat{\underset{\sim}{P}}_M)=\underset{\sim}{S}_{N,M}^{(2)}(\underset{\sim}{\beta})$.

By Mean Value Theorem,

$\frac{1}{N}\left[\underset{\sim}{G}(\underset{\sim}{\beta},\hat{\underset{\sim}{P}}_M)-\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P})\right]=\frac{1}{N}\frac{\partial}{\partial\underset{\sim}{P}}\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P}^*)\left[\hat{\underset{\sim}{P}}_M-\underset{\sim}{P}\right]$, where $\underset{\sim}{P}^*$ is between $\hat{\underset{\sim}{P}}_M$ and $\underset{\sim}{P}$. Then:

$$\frac{1}{N}\left\|\underset{\sim}{G}(\underset{\sim}{\beta},\hat{\underset{\sim}{P}}_M)-\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P})\right\|\le\left\|\frac{1}{N}\frac{\partial}{\partial\underset{\sim}{P}}\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P}^*)\right\|\left\|\hat{\underset{\sim}{P}}_M-\underset{\sim}{P}\right\|$$

$$\le\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial\underset{\sim}{P}}\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P})\right\|\bullet\left\|\hat{\underset{\sim}{P}}_M-\underset{\sim}{P}\right\| \qquad \text{(M1.1)}$$

$\xrightarrow{p}0$ as $N,M\to\infty$ since $\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial\underset{\sim}{P}}\underset{\sim}{G}(\underset{\sim}{\beta},\underset{\sim}{P})\right\|$ is bounded in probability by claim M2 and $\hat{\underset{\sim}{P}}_M\xrightarrow{p}\underset{\sim}{P}$ as $M\to\infty$ by assumption A1.

Claim M2: $\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\frac{\partial}{\partial\underset{\sim}{P}}\underset{\sim}{S}_N^{(1)}(\underset{\sim}{\beta})\right\|=\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\underset{\sim}{P}}\frac{\partial}{\underset{\sim}{\beta}}Lnf_i(y_i\mid w_i,\underset{\sim}{\beta},\underset{\sim}{P})\right]\right\|$ is Cauchy in probability and hence is bounded in probability as $N\to\infty$

Proof of M2:

Let $A_{N,jl} = \frac{1}{N}\sum_{i=1}^{N} \max_{\underset{\sim}{\beta},\underset{\sim}{P}} \frac{\partial}{P_l}\frac{\partial}{\beta_j} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$,

$$B_{i,jl} = \max_{\underset{\sim}{\beta},\underset{\sim}{P}} \frac{\partial}{P_l}\frac{\partial}{\beta_j} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

Then,

$$A_{N+k,jl} - A_{N,jl} = \frac{1}{N+k}\sum_{i=1}^{N+k} B_{i,jl} - \frac{1}{N}\sum_{i=1}^{N} B_{i,jl} = \left(\frac{1}{N+k} - \frac{1}{N}\right)\sum_{i=1}^{N} B_{i,jl} + \frac{1}{N+k}\sum_{i=N+1}^{N+k} B_{i,jl}$$

By Markov Inequality,

$$P\left\{\left|A_{N+k,jl} - A_{N,jl}\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} E\left[A_{N+k,jl} - A_{N,jl}\right]^2$$

$$= \frac{1}{\varepsilon^2} E\left[\left(\frac{1}{N+k} - \frac{1}{N}\right)\sum_{i=1}^{N} B_{i,jl} + \frac{1}{N+k}\sum_{i=N+1}^{N+k} B_{i,jl}\right]^2$$

$$\leq \frac{2}{\varepsilon^2}\left\{\left(\frac{1}{N+k} - \frac{1}{N}\right)^2 E\left[\sum_{i=1}^{N} B_{i,jl}\right]^2 + \left(\frac{1}{N+k}\right)^2 E\left[\sum_{i=N+1}^{N+k} B_{i,jl}\right]^2\right\}$$

By $C_r - Inequality$

$$= \frac{2}{\varepsilon^2}\left\{\frac{k^2}{\left(N(N+k)\right)^2}\sum_{i=1}^{N}\sum_{s=1}^{N} E\left[B_{i,jl}B_{s,jl}\right] + \frac{1}{\left(N+k\right)^2}\sum_{i=N+1}^{N+k}\sum_{s=N+1}^{N+k} E\left[B_{i,jl}B_{s,jl}\right]\right\}$$

Because of independent, $E\left[B_{i,jl}B_{s,jl}\right] = E\left(B_{i,jl}\right)E\left(B_{s,jl}\right)$ and by claim M3,

$$E\left[\max_{\underset{\sim}{\beta},\underset{\sim}{P}} \frac{\partial}{\partial P_l}\left(\frac{\partial}{\partial \beta_j} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right)\right] \leq K < \infty$$, uniformly in i:

$$P\left\{\left|A_{N+k,jl} - A_{N,jl}\right| \geq \varepsilon\right\} \leq \frac{2}{\varepsilon^2}\left\{\frac{k^2}{\left(N(N+k)\right)^2} N^2 K^2 + \frac{1}{\left(N+k\right)^2} k^2 K^2\right\}$$

$\rightarrow 0$ as $N \rightarrow \infty$.

Therefore, $A_{N,jl} = \frac{1}{N}\sum_{i=1}^{N} \max_{\underset{\sim}{\beta},\underset{\sim}{P}} \frac{\partial}{P_l}\frac{\partial}{\beta_j} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is bounded in probability.

Then

$\frac{1}{N}\sum_{i=1}^{N} \max_{\underset{\sim}{\beta},\underset{\sim}{P}} \left\| \frac{\partial}{\underset{\sim}{P}}\frac{\partial}{\underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \right\|$ is bounded in probability. And:

$\max_{\underset{\sim}{\beta},\underset{\sim}{P}} \left\| \frac{1}{N}\sum_{i=1}^{N} \left[\frac{\partial}{\underset{\sim}{P}}\frac{\partial}{\underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right] \right\| \leq \frac{1}{N}\sum_{i=1}^{N} \max_{\underset{\sim}{\beta},\underset{\sim}{P}} \left\| \frac{\partial}{\underset{\sim}{P}}\frac{\partial}{\underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \right\|$ is bounded in probability.

Claim M3: $E\left[\max_{\underset{\sim}{\beta},\underset{\sim}{P}} \frac{\partial}{\partial P_l}\left(\frac{\partial}{\partial \beta_j} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right)\right] \leq K < \infty$, uniformly in i;

Proof of M3:

For our model here, $\frac{\partial}{\partial P_l} \in \left\{ \frac{\partial}{\partial p_{1w_i}}, \frac{\partial}{\partial p_{2w_i}}, \frac{\partial}{\partial p_{3w_i}} \right\}$, $w_i \in \{1,2,3\}$ and

$\frac{\partial}{\partial \beta_j} \in \left\{ \frac{\partial}{\partial \beta_0}, \frac{\partial}{\partial \beta_1}, \frac{\partial}{\partial \beta_2} \right\}$

Case $Y_i = 1$:

From condition N3 of Appendix D:

$$\frac{\partial}{\partial p_{1w_i}} \frac{\partial}{\partial \beta_0} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{1w_i}} \frac{p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2}}{p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}}}$$

$$= \frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \left(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}} \right)}{\left(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}} \right)^2}$$

$$- \frac{\left(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2} \right) \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}} \right)^2}$$

$$< \frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})} \left(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}} \right)}{\left(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}} \right)^2}$$

$$-\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2} \text{ since}$$

$$\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}<\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$$

$$=\frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})}\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\left(1-\frac{1}{1+e^{\beta_0+\beta_1}}\right)+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}\left(1-\frac{1}{1+e^{\beta_0+\beta_2}}\right)+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\left(1-\frac{1}{1+e^{\beta_0}}\right)\right)}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2}$$

$$<\frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})}\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2} \text{ since } \left(1-\frac{1}{1+e^x}\right)=\frac{e^x}{1+e^x}<1$$

$$=\frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}$$

$$<\frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)}=\frac{1}{p_{1w_i}}$$

Similarly, we could show:

$$\frac{\partial}{\partial p_{2w_i}}\frac{\partial}{\partial\beta_0}Lnf_i(y_i\,|\,w_i,\underset{\sim}{\beta},\underset{\sim}{P})<\frac{1}{p_{2w_i}} \text{ and } \frac{\partial}{\partial p_{3w_i}}\frac{\partial}{\partial\beta_0}Lnf_i(y_i\,|\,w_i,\underset{\sim}{\beta},\underset{\sim}{P})<\frac{1}{p_{3w_i}}$$

$$\frac{\partial}{\partial p_{1w_i}}\frac{\partial}{\partial\beta_1}Lnf_i(y_i\,|\,w_i,\underset{\sim}{\beta},\underset{\sim}{P})=\frac{\partial}{\partial p_{1w_i}}\frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}}$$

$$= \frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right) - p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2}$$

$$< \frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2}$$

$$= \frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}$$

$$< \frac{\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)} < \frac{1}{p_{1w_i}}$$

$$\frac{\partial}{\partial p_{2w_i}}\frac{\partial}{\partial \beta_1} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{2w_i}} \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}}$$

$$= -\frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2} < 0$$

$$\frac{\partial}{\partial p_{3w_i}}\frac{\partial}{\partial \beta_1}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{3w_i}}\frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}}$$

$$= -\frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{e^{\beta_0}}{1+e^{\beta_0}}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2} < 0$$

Similarly,

$$\frac{\partial}{\partial p_{1w_i}}\frac{\partial}{\partial \beta_2}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{1w_i}}\frac{p_{1w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}} < \frac{1}{p_{1w_i}}$$

$$\frac{\partial}{\partial p_{2w_i}}\frac{\partial}{\partial \beta_2}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{2w_i}}\frac{p_{1w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}} < 0$$

$$\frac{\partial}{\partial p_{3w_i}}\frac{\partial}{\partial \beta_2}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{3w_i}}\frac{p_{1w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}} < 0$$

Case $Y_i = 0$:

$$\frac{\partial}{\partial p_{1w_i}}\frac{\partial}{\partial \beta_0}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{1w_i}}\frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}$$

$$
= \frac{-\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}
$$

$$
+\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)\frac{1}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}
$$

$$
< \frac{-\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}
$$

$$
+\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})}\right)\frac{1}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2} \text{ since } \frac{e^x}{(1+e^x)^2}<\frac{e^x}{1+e^x}
$$

$$
= \frac{-\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+\frac{1}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}
$$

$$
= \frac{\frac{1}{(1+e^{\beta_0+\beta_1})^2}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}
$$

$$
< \frac{\frac{1}{(1+e^{\beta_0+\beta_1})^2}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}\right)} = \frac{1}{p_{1w_i}\left(1+e^{\beta_0+\beta_1}\right)} < \frac{1}{p_{1w_i}}
$$

Similarly, we could show:

$$\frac{\partial}{\partial p_{2w_i}}\frac{\partial}{\partial \beta_0} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) < \frac{1}{p_{2w_i}} \text{ and } \frac{\partial}{\partial p_{3w_i}}\frac{\partial}{\partial \beta_0} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) < \frac{1}{p_{3w_i}}.$$

$$\frac{\partial}{\partial p_{1w_i}}\frac{\partial}{\partial \beta_1} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{1w_i}} \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}}$$

$$\frac{-\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right) + p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{1}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}$$

$$< \frac{-\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}\right) + p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{1}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2} = 0$$

$$\frac{\partial}{\partial p_{2w_i}}\frac{\partial}{\partial \beta_1} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{2w_i}} \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}}$$

$$= \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{1}{1+e^{\beta_0+\beta_2}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}$$

$$< \frac{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_2}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2} \text{ since } \frac{e^x}{(1+e^x)^2} = \frac{e^x}{(1+e^x)}\frac{1}{1+e^x} < 1$$

$$< \frac{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_2}}}{\left(p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}\right)^2} = \frac{p_{1w_i}}{\left(p_{2w_i}\right)^2\frac{1}{1+e^{\beta_0+\beta_2}}} < \frac{p_{1w_i}}{\left(p_{2w_i}\right)^2}$$

$$\frac{\partial}{\partial p_{3w_i}}\frac{\partial}{\partial \beta_1} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{3w_i}}\frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}}$$

$$= \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\frac{1}{1+e^{\beta_0}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2} < \frac{p_{1w_i}}{\left(p_{3w_i}\right)^2}$$

$$\frac{\partial}{\partial p_{1w_i}}\frac{\partial}{\partial \beta_2} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{1w_i}}\frac{-p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}}$$

$$= \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\frac{1}{1+e^{\beta_0+\beta_1}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2} < \frac{p_{2w_i}}{\left(p_{1w_i}\right)^2}$$

$$\frac{\partial}{\partial p_{2w_i}}\frac{\partial}{\partial \beta_2} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{2w_i}}\frac{-p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}}$$

$$= \frac{-\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right) + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\frac{1}{1+e^{\beta_0+\beta_2}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}$$

$$< 0$$

$$\frac{\partial}{\partial p_{3w_i}}\frac{\partial}{\partial \beta_2}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial p_{3w_i}}\frac{-p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}}$$

$$=\frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\frac{1}{1+e^{\beta_0}}}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2} < \frac{p_{2w_i}}{\left(p_{3w_i}\right)^2}$$

Now let $K = \max\left\{\frac{1}{p_{1w_i}}, \frac{1}{p_{2w_i}}, \frac{1}{p_{3w_i}}, \frac{p_{1w_i}}{\left(p_{2w_i}\right)^2}, \frac{p_{1w_i}}{\left(p_{3w_i}\right)^2}, \frac{p_{2w_i}}{\left(p_{1w_i}\right)^2}, \frac{p_{2w_i}}{\left(p_{3w_i}\right)^2}\right\}$, then

$\max\limits_{\underset{\sim}{\beta},\underset{\sim}{P}}\frac{\partial}{\partial P_l}\left(\frac{\partial}{\partial \beta_j}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right) \le K < \infty$ and

$$E\left[\max_{\underset{\sim}{\beta},\underset{\sim}{P}}\frac{\partial}{\partial P_l}\left(\frac{\partial}{\partial \beta_j}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right)\right] \le K < \infty.$$

Claim M4: If $x_n^T A_n x_n \xrightarrow{p} 0$ and $\lambda_{\min}(A_n) \ge \delta > 0,\ \forall n \ge N$, then $x_n \xrightarrow{p} 0$.

Proof of M4:

Note that $\lambda_{\min}(A_n)\|x_n\|^2 \le x_n^T A_n x_n$.

From hypothesis, $x_n^T A_n x_n \xrightarrow{p} 0$ and $\lambda_{\min}(A_n) \ge \delta > 0,\ \forall n \ge N$ we must have $\|x_n\|^2 \xrightarrow{p} 0$. Therefore, $x_n \xrightarrow{p} 0$.

Claim M5: For any $\{\underset{\sim}{\beta}\} \in \Theta$,

i) $\left\| -\frac{1}{n} H_n(\underset{\sim}{\beta}) - \bar{\Gamma}_n(\underset{\sim}{\beta}) \right\| \xrightarrow{p} 0$ as $n \to \infty$

ii) $\lambda_{\min}(-\frac{1}{n} H_n(\underset{\sim}{\beta})) \geq \delta > 0, \ \forall n \geq N$.

Proof of M5-i:

From condition N9(ii) of Appendix D, $E\left[\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right]^2 \leq K < \infty$ for some K and $\forall i$. Then $E\left[\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right] \leq K$ and the variance:

$$\sigma_i^2 = \operatorname{var}\left(\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right) = E\left[\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right]^2 - \left(E\left[\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right]\right)^2 \leq K$$

Hence $\lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right) \leq \lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n}K\right) = \lim_{n\to\infty}\left(\frac{1}{n^2}nK\right) \to 0$.

By Chebyshev theorem (Serfling, p.27),

$$\left\| \frac{1}{n}\sum_{i=1}^{n}\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - \frac{1}{n}\sum_{i=1}^{n}E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right] \right\| \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Or $\left\| -\frac{1}{n}\sum_{i=1}^{n}\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) + \frac{1}{n}\sum_{i=1}^{n}E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right] \right\| \xrightarrow{p} 0$ as $n \to \infty$.

From definition $-\frac{1}{n}H_n(\underset{\sim}{\beta}_n) = -\frac{1}{n}\sum_{i=1}^{n}\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}_n, \underset{\sim}{P})$ and

$\bar{\Gamma}_n(\underset{\sim}{\beta}) = -\frac{1}{n}\sum_{i=1}^{n}E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right]$.

Therefore, claim M5-i is proved.

Proof of M5-ii:

From the proof of condition N7 in Appendix D, $\bar{\Gamma}_n(\underset{\sim}{\beta})$ is positive definite for $\forall \underset{\sim}{\beta} \in \Theta$.

Being positive definite matrix, all the eigen values of $\bar{\Gamma}_n(\underset{\sim}{\beta})$ are greater than zeros. Let $\lambda_{\min}(\bar{\Gamma}_n(\underset{\sim}{\beta})) > 0$ be the smallest eigen value of $\bar{\Gamma}_n(\underset{\sim}{\beta})$ and

$\alpha = \inf_{\underset{\sim}{\beta} \in \Theta} \lambda_{\min}(\bar{\Gamma}_n(\underset{\sim}{\beta})) > 0$

It follows from M5-i, $\left| \lambda_{\min}(-\frac{1}{n} H_n(\underset{\sim}{\beta})) - \lambda_{\min}(\bar{\Gamma}_n(\underset{\sim}{\beta})) \right| \xrightarrow{p} 0$ as $n \to \infty$.

Then $\exists N$ such that for $n \geq N$, $\left| \lambda_{\min}(-\frac{1}{n} H_n(\underset{\sim}{\beta})) - \lambda_{\min}(\bar{\Gamma}_n(\underset{\sim}{\beta})) \right| < \frac{\alpha}{2}$

Let $\delta = \frac{\alpha}{2}$, then for sufficiently large N, $\lambda_{\min}(-\frac{1}{n} H_n(\underset{\sim}{\beta})) \geq \delta > 0, \ \forall n \geq N$.

APPENDIX C

PROOF OF CONSISTENCY CONDITIONS C1-C5

C1. Θ is a closed subset of $\Re^p$

C2. $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is an upper semicontinuous (u.s.c.) function of $\underset{\sim}{\beta}$, uniformly in i, a.s. [P].

C3. There exists $\rho^* = \rho^*(\underset{\sim}{\beta}) > 0$ and $r > 0$ for which

i) $E[R_i(\underset{\sim}{\beta}, \rho)]^2 \le K_1,\ 0 \le \rho \le \rho^*;$

ii) $E[V_i(r)]^2 \le K_2.$

C4. i) $\lim \bar{r}_N(\underset{\sim}{\beta}) < 0,\ \underset{\sim}{\beta} \ne \underset{\sim}{\beta}_0;$

ii) $\lim \bar{v}_N(r) < 0.$

C5. $R_i(\underset{\sim}{\beta}, \rho)$ and $V_i(r)$ are measurable functions of Y_i

C5.1: $R_i(\underset{\sim}{\beta}, \rho) = \sup\left\{R_i(\underset{\sim}{\beta}^*) : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\} = \max\left\{R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho)\right\}$

C5.2: $V_i(r) = \sup\left\{R_i(\underset{\sim}{\beta}) : \left\|\underset{\sim}{\beta}\right\| > r\right\} = \max\left\{R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, r)\right\}$

Condition C1: Θ is a closed subset of $\Re^p$

Proof: By assumption, Θ is a compact subset of $\Re^p$ and so it is closed and bounded by Heine-Borel Theorem.

Condition C2: $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is an upper semicontinuous (u.s.c.) function of $\underset{\sim}{\beta}$, uniformly in i, a.s. [P].

Proof: We will prove for the case $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ continuous (uniformly in i) which also holds for semicontinuous. From equation (1.4) of chapter 1, the distribution function is:

$$f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$= p_{1w_i}\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)^{1-y_i} + p_{2w_i}\left(\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0+\beta_2}}\right)^{1-y_i} + p_{3w_i}\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0}}\right)^{1-y_i}$$

$$= \begin{cases} p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\dfrac{e^{\beta_0}}{1+e^{\beta_0}} & if \;\; y_i = 1 \\ p_{1w_i}\dfrac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\dfrac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\dfrac{1}{1+e^{\beta_0}} & if \;\; y_i = 0 \end{cases}$$

Case $y_i = 1$: First we note $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is a composition of basic exponential functions which are continuous, then $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is continuous. From the assumption, Θ is a compact subset of $\Re^p$ we can show $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is uniformly continuous in i.

Let $\varepsilon > 0$ be given. For every $\underset{\sim}{\beta} \in \Theta$ there exists $\delta(\underset{\sim}{\beta}) > 0$ such that $\underset{\sim}{\beta}' \in \Theta$ and

$$\left\|\underset{\sim}{\beta} - \underset{\sim}{\beta}'\right\| < \delta(\underset{\sim}{\beta}) \Rightarrow \left| f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - f_i(y_i \mid w_i, \underset{\sim}{\beta}', \underset{\sim}{P}) \right| < \tfrac{1}{2}\varepsilon$$

Then $\Theta \subset \left\{ B(\underset{\sim}{\beta}, \tfrac{1}{2}\delta(\underset{\sim}{\beta})) \mid \underset{\sim}{\beta} \in \Theta \right\}$.

Since Θ is compact, there exist $\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, ..., \underset{\sim}{\beta}_N \in \Theta$ such that $\Theta \subset \bigcup_{k=1}^{N} B(\underset{\sim}{\beta}_k, \tfrac{1}{2}\delta(\underset{\sim}{\beta}_k))$

Let $\delta = \min\left\{\frac{1}{2}\delta(\underset{\sim}{\beta}_1), \frac{1}{2}\delta(\underset{\sim}{\beta}_2), ..., \frac{1}{2}\delta(\underset{\sim}{\beta}_N)\right\}$.

Now if we let $\underset{\sim}{\beta}, \underset{\sim}{\beta}' \in \Theta$ and $\left\|\underset{\sim}{\beta} - \underset{\sim}{\beta}'\right\| < \delta$, since $\underset{\sim}{\beta} \in \Theta$, there exists k such that

$\underset{\sim}{\beta} \in B(\underset{\sim}{\beta}_k, \frac{1}{2}\delta(\underset{\sim}{\beta}_k))$ and $\left\|\underset{\sim}{\beta} - \underset{\sim}{\beta}'\right\| < \delta < \frac{1}{2}\delta(\underset{\sim}{\beta}_k)$. Thus $\underset{\sim}{\beta}' \in B(\underset{\sim}{\beta}_k, \delta(\underset{\sim}{\beta}_k))$

By triangle inequality,

$$\left|f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - f_i(y_i \mid w_i, \underset{\sim}{\beta}', \underset{\sim}{P})\right| < \left|f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - f_i(y_i \mid w_i, \underset{\sim}{\beta}_k, \underset{\sim}{P})\right| + \left|f_i(y_i \mid w_i, \underset{\sim}{\beta}_k, \underset{\sim}{P}) - f_i(y_i \mid w_i, \underset{\sim}{\beta}', \underset{\sim}{P})\right|$$

$$= \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon$$

Hence the uniform continuity in i is proved.

Case $y_i = 0$: The above argument also applied to this case.

Condition C3. There exists $\rho^* = \rho^*(\underset{\sim}{\beta}) > 0$ and $r > 0$ for which

i) $E[R_i(\underset{\sim}{\beta}, \rho)]^2 \leq K_1,\ 0 \leq \rho \leq \rho^*$;

ii) $E[V_i(r)]^2 \leq K_2$.

<u>Proof of C3-i:</u>

Let $\underset{\sim}{\beta}_0 = (\beta_0^0, \beta_1^0, \beta_2^0)$ be the true parameter. By definition:

$$R_i(\underset{\sim}{\beta}) = \begin{cases} Ln\left[\dfrac{f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})}{f_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P})}\right] & if \quad f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) > 0 \\ 0 \ otherwise & \end{cases}$$

Case $y_i = 1$:

$$R_i(\underset{\sim}{\beta}) = Ln\left[\frac{p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\dfrac{e^{\beta_0}}{1+e^{\beta_0}}}{p_{1w_i}\dfrac{e^{\beta_0^0+\beta_1^0}}{1+e^{\beta_0^0+\beta_1^0}} + p_{2w_i}\dfrac{e^{\beta_0^0+\beta_2^0}}{1+e^{\beta_0^0+\beta_2^0}} + p_{3w_i}\dfrac{e^{\beta_0^0}}{1+e^{\beta_0^0}}}\right]$$

$$< Ln\left[\frac{3}{p_{1w_i}\frac{e^{\beta_0^0+\beta_1^0}}{1+e^{\beta_0^0+\beta_1^0}}+p_{2w_i}\frac{e^{\beta_0^0+\beta_2^0}}{1+e^{\beta_0^0+\beta_2^0}}+p_{3w_i}\frac{e^{\beta_0^0}}{1+e^{\beta_0^0}}}\right]$$

$$< Ln\left[\frac{3}{p_{3w_i}\frac{e^{\beta_0^0}}{1+e^{\beta_0^0}}}\right]=Ln\left[\frac{3\left(1+e^{\beta_0^0}\right)}{p_{3w_i}e^{\beta_0^0}}\right]$$

$$< Ln\left[\frac{3e^{\beta_0^0}}{p_{3w_i}e^{\beta_0^0}}\right]=Ln\left[\frac{3}{p_{3w_i}}\right]$$

Case $y_i=0$:

$$R_i(\underset{\sim}{\beta}) = Ln\left[\frac{p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}}{p_{1w_i}\frac{1}{1+e^{\beta_0^0+\beta_1^0}}+p_{2w_i}\frac{1}{1+e^{\beta_0^0+\beta_2^0}}+p_{3w_i}\frac{1}{1+e^{\beta_0^0}}}\right]$$

$$< Ln\left[\frac{3}{p_{1w_i}\frac{1}{1+e^{\beta_0^0+\beta_1^0}}+p_{2w_i}\frac{1}{1+e^{\beta_0^0+\beta_2^0}}+p_{3w_i}\frac{1}{1+e^{\beta_0^0}}}\right]$$

$$< Ln\left[\frac{3}{p_{3w_i}\frac{1}{1+e^{\beta_0^0}}}\right]=Ln\left[\frac{3\left(1+e^{\beta_0^0}\right)}{p_{3w_i}}\right]$$

$$< Ln\left[\frac{3}{p_{3w_i}}\right]$$

Let $C = Ln\left[\frac{3}{p_{3w_i}}\right]$, then $\left|R_i(\underset{\sim}{\beta})\right| <$ C

$$R_i(\underset{\sim}{\beta},\rho) = \sup\left\{R_i(\underset{\sim}{\beta}^*) : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\}$$

$$\le \sup_{\underset{\sim}{\beta}^* \in \Theta}\left\{R_i(\underset{\sim}{\beta}^*)\right\} \text{ since } \left\{\underset{\sim}{\beta}^* : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\} \subseteq \Theta$$

$$\le \sup_{\underset{\sim}{\beta}^* \in \Theta}\left\{\left|R_i(\underset{\sim}{\beta}^*)\right|\right\}$$

$$-R_i(\underset{\sim}{\beta},\rho) = -\sup\left\{R_i(\underset{\sim}{\beta}^*) : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\} = \inf\left\{-R_i(\underset{\sim}{\beta}^*) : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\}$$

$$\le \sup\left\{-R_i(\underset{\sim}{\beta}^*) : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\}$$

$$\le \sup_{\underset{\sim}{\beta}^* \in \Theta}\left\{-R_i(\underset{\sim}{\beta}^*)\right\} \text{ since } \left\{\underset{\sim}{\beta}^* : \left\|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\right\| \le \rho\right\} \subseteq \Theta$$

$$\le \sup_{\underset{\sim}{\beta}^* \in \Theta}\left\{\left|R_i(\underset{\sim}{\beta}^*)\right|\right\}$$

Hence, $\left|R_i(\underset{\sim}{\beta},\rho)\right| \le \sup_{\underset{\sim}{\beta}^* \in \Theta}\left\{\left|R_i(\underset{\sim}{\beta}^*)\right|\right\} \le C$

And $E\left[R_i(\underset{\sim}{\beta},\rho)\right]^2 \le C^2 = K$ for some K.

<u>Proof of C3-ii:</u>

$$V_i(r) = \sup\left\{R_i(\underset{\sim}{\beta}) : \left\|\underset{\sim}{\beta}\right\| > r\right\}$$

$$\le \sup_{\underset{\sim}{\beta}^* \in \Theta}\left\{R_i(\underset{\sim}{\beta}^*)\right\} \text{ since } \left\{\underset{\sim}{\beta}^* : \left\|\underset{\sim}{\beta}^*\right\| > r\right\} \subseteq \Theta$$

$$\leq \sup_{\underset{\sim}{\beta}^* \in \Theta} \left\{ \left| R_i(\underset{\sim}{\beta}^*) \right| \right\}$$

$$-V_i(r) = -\sup\left\{ R_i(\underset{\sim}{\beta}) : \left\| \underset{\sim}{\beta} \right\| > r \right\} = \inf \left\{ -R_i(\underset{\sim}{\beta}) : \left\| \underset{\sim}{\beta} \right\| > r \right\}$$

$$\leq \sup\left\{ -R_i(\underset{\sim}{\beta}) : \left\| \underset{\sim}{\beta} \right\| > r \right\} \leq \sup_{\underset{\sim}{\beta}^* \in \Theta} \left\{ -R_i(\underset{\sim}{\beta}^*) \right\}$$

$$\leq \sup_{\underset{\sim}{\beta}^* \in \Theta} \left\{ \left| R_i(\underset{\sim}{\beta}^*) \right| \right\}$$

Hence, $\left| V_i(r) \right| \leq \sup_{\underset{\sim}{\beta}^* \in \Theta} \left\{ \left| R_i(\underset{\sim}{\beta}^*) \right| \right\} \leq C$

And the result follows from argument in C3-i.

Condition C4:

i) $\lim \overline{r}_N(\underset{\sim}{\beta}) < 0, \ \underset{\sim}{\beta} \neq \underset{\sim}{\beta}_0;$

ii) $\lim \overline{v}_N(r) < 0.$

Proof of C4-i:

By Wald's Lemma 1: $E\left[Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \right] < E\left[Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P}) \right]$

So that $E\left[Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \right] - E\left[Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P}) \right] < 0$

$$E\left[Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P}) \right] < 0$$

$$E\left[Ln \frac{f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})}{f_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P})} \right] < 0$$

Or $E\left[R_i(\underset{\sim}{\beta}) \right] < 0$

Hence, $\bar{r}_N(\underset{\sim}{\beta}) = \dfrac{E\left[\sum_{i=1}^{N} R_i(\underset{\sim}{\beta})\right]}{N} = \dfrac{\sum_{i=1}^{N} E\left[R_i(\underset{\sim}{\beta})\right]}{N} < 0$.

Proof of C4-ii:

Since Θ is compact, it is bounded. Then there exists $r > 0$ such that if $\underset{\sim}{\beta} \in \Theta$ then $\|\underset{\sim}{\beta}\| < r$. Therefore, $\left\{R_i(\underset{\sim}{\beta}) : \|\underset{\sim}{\beta}\| > r\right\} = \varnothing, \forall i$. Hence, $\lim\sup \bar{v}_N(r) = -\infty < 0$

Condition C5: $R_i(\underset{\sim}{\beta}, \rho)$ and $V_i(r)$ are measurable functions of Y_i

Proof:

By definition $R_i(\underset{\sim}{\beta}) = Ln \dfrac{f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})}{f_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P})} = Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P})$

Note that $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is a continuous function of $\underset{\sim}{\beta}$ for fixed (y_i, w_i) and so is a measurable function. Then $R_i(\underset{\sim}{\beta})$ is a measurable function of y_i.

Let $W(\underset{\sim}{\beta}, \rho)$ be a countable set such that $W(\underset{\sim}{\beta}, \rho) \subseteq B(\underset{\sim}{\beta}, \rho) = \left\{\underset{\sim}{\varphi} : \|\underset{\sim}{\varphi} - \underset{\sim}{\beta}\| \le \rho\right\}$

Next we will show $R_i(\underset{\sim}{\beta}, \rho) = \sup\left\{R_i(\underset{\sim}{\beta}^*) : \|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\| \le \rho\right\}$ is y_i-measurable.

By claim C5.1 below, $\left\{y_i : \sup\left\{R_i(\underset{\sim}{\beta}^*) : \|\underset{\sim}{\beta}^* - \underset{\sim}{\beta}\| \le \rho\right\} < \varepsilon\right\}$

$$= \left\{y_i : \max\left\{R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho)\right\} < \varepsilon\right\}$$

$$= \bigcap_{\underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho)} \left\{y_i : R_i(\underset{\sim}{\beta}^*) < \varepsilon\right\}$$

Which is measurable since $W(\underset{\sim}{\beta}, \rho)$ is countable and $R_i(\underset{\sim}{\beta}^*)$ is measurable.

Therefore $R_i(\underset{\sim}{\beta}, \rho)$ is measurable.

To prove $V_i(r)$ is measurable, we let $W(\underset{\sim}{\beta}, r)$ be a countable set such that

$W(\underset{\sim}{\beta}, r) \subseteq \Theta \setminus \left\{ B(\underset{\sim}{\beta}, r) = \left\{ \underset{\sim}{\varphi} : \left\| \underset{\sim}{\varphi} - \underset{\sim}{\beta} \right\| \le r \right\} \right\}$ and follow the same argument above together with claim C5.2 below.

Claim C5.1: $R_i(\underset{\sim}{\beta}, \rho) = \sup\left\{ R_i(\underset{\sim}{\beta}^*) : \left\| \underset{\sim}{\beta}^* - \underset{\sim}{\beta} \right\| \le \rho \right\} = \max\left\{ R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho) \right\}$

Proof:

Since $W(\underset{\sim}{\beta}, \rho) \subseteq B(\underset{\sim}{\beta}, \rho) = \left\{ \underset{\sim}{\varphi} : \left\| \underset{\sim}{\varphi} - \underset{\sim}{\beta} \right\| \le \rho \right\}$, $R_i(\underset{\sim}{\beta}, \rho) \ge \max\left\{ R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho) \right\}$

Note that $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is a continuous function of $\underset{\sim}{\beta}$ for fixed (y_i, w_i), $R_i(\underset{\sim}{\beta}) = Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) - Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}_0, \underset{\sim}{P})$ is continuous at $\underset{\sim}{\beta} \in W(\underset{\sim}{\beta}, \rho)$.

Because $B(\underset{\sim}{\beta}, \rho)$ is compact and $R_i(\underset{\sim}{\beta})$ is continuous, $\exists \underset{\sim}{\beta}^* \in B(\underset{\sim}{\beta}, \rho)$ such that $R_i(\underset{\sim}{\beta}, \rho) = R_i(\underset{\sim}{\beta}^*)$. By the continuity of $R_i(\underset{\sim}{\beta})$, for any $\varepsilon > 0$, there exists $\delta > 0$ and $\underset{\sim}{\beta}^s \in W(\underset{\sim}{\beta}, \rho)$ such that if $\left\| \underset{\sim}{\beta}^* - \underset{\sim}{\beta}^s \right\| < \delta$ then $\left| R_i(\underset{\sim}{\beta}^*) - R_i(\underset{\sim}{\beta}^s) \right| < \varepsilon$.

Then $R_i(\underset{\sim}{\beta}, \rho) = R_i(\underset{\sim}{\beta}^*) < R_i(\underset{\sim}{\beta}^s) + \varepsilon \le \max\left\{ R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho) \right\} + \varepsilon$.

Since ε is arbitrary, $R_i(\underset{\sim}{\beta}, \rho) \le \max\left\{ R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho) \right\}$

Therefore, $R_i(\underset{\sim}{\beta}, \rho) = \max\left\{ R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, \rho) \right\}$.

Claim C5.2: $V_i(r) = \sup\left\{ R_i(\underset{\sim}{\beta}) : \left\| \underset{\sim}{\beta} \right\| > r \right\} = \max\left\{ R_i(\underset{\sim}{\beta}^*) : \underset{\sim}{\beta}^* \in W(\underset{\sim}{\beta}, r) \right\}$

Proof:

Use similar argument in claim C5.1.

APPENDIX D

PROOFS OF ASYMPTOTIC NORMALITY CONDITIONS N1-N9

Condition N1: The interior of Θ is an open set of R^p and $\underset{\sim}{\beta}_0$ is an interior point of Θ.

Condition N2: $\hat{\underset{\sim}{\beta}}_N \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N \to \infty$

Condition N3:

i) $\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ exists.

ii) $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ exists.

Condition N4: $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is a continuous function of $\underset{\sim}{\beta}$, uniformly in i, a.s. [P], and is measurable functions of Y_i

Condition N5: $E\left[\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right] = \underset{\sim}{0};\ i = 1, 2, \ldots$

Condition N6:

$$\Gamma_i(\underset{\sim}{\beta}) = E\left[\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})^T \mid \underset{\sim}{\beta}\right] = -E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]$$

Condition N7: $\exists \bar{\Gamma}(\underset{\sim}{\beta})$ such that $\lim_{N \to \infty} \bar{\Gamma}_N(\underset{\sim}{\beta}) = \bar{\Gamma}(\underset{\sim}{\beta})$, and $\bar{\Gamma}(\underset{\sim}{\beta})$ is positive definite.

Condition N8: $E\left[\dot{\Phi}_{i,j}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]^3 \le K$

Condition N9: There exists $\varepsilon > 0$ and random variables $B_{i,jl}(Y_i)$ such that

i) $\sup\left\{\left|\ddot{\Phi}_{i,jl}(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| : \left\|\underset{\sim}{\beta} - \underset{\sim}{\beta}_0\right\| \le \varepsilon\right\} \le B_{i,jl}(Y_i)$.

ii) $E\left[B_{i,jl}(Y_i)\right]^{1+\delta} \le K$, for some $\delta > 0, K > 0$.

Condition N1: The interior of Θ is an open set of R^p and $\underset{\sim}{\beta}_0$ is an interior point of Θ.

Proof:

By assumption Θ is compact (closed and bounded), then the interior of Θ is an open set. It is also assumed $\underset{\sim}{\beta}_0 \in \Theta$

Condition N2: $\hat{\underset{\sim}{\beta}}_N \xrightarrow{p} \underset{\sim}{\beta}_0$ as $N \to \infty$

Proof:

Condition N2 is proved in Theorem 1.

Condition N3:

i) $\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ exists.

ii) $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ exists.

Proof:

i) From equation (1.4) of chapter 1, $f_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$

$$= p_{1w_i}\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)^{1-y_i} + p_{2w_i}\left(\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0+\beta_2}}\right)^{1-y_i} + p_{3w_i}\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0}}\right)^{1-y_i}$$

$$= \begin{cases} p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \dfrac{e^{\beta_0}}{1+e^{\beta_0}} & if \;\; y_i = 1 \\ p_{1w_i} \dfrac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \dfrac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \dfrac{1}{1+e^{\beta_0}} & if \;\; y_i = 0 \end{cases}$$

Let $A_{w_i} = p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \dfrac{e^{\beta_0}}{1+e^{\beta_0}}$,

$B_{w_i} = p_{1w_i} \dfrac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \dfrac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \dfrac{1}{1+e^{\beta_0}}$, and

$$C_{w_i} = \frac{\partial}{\partial \beta_0} A_{w_i} = p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2}$$

Then, $\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \begin{bmatrix} \frac{d}{d\beta_0} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_1} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_2} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \end{bmatrix} = \begin{bmatrix} \frac{d}{d\beta_0} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_1} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_2} Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \end{bmatrix}$

$$= \begin{cases} \begin{bmatrix} \frac{C_{w_i}}{A_{w_i}} \\ \frac{1}{A_{w_i}} p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \frac{1}{A_{w_i}} p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \ \ Y_i = 1 \\ \begin{bmatrix} \frac{-C_{w_i}}{B_{w_i}} \\ \frac{-1}{B_{w_i}} p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \frac{-1}{B_{w_i}} p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \ \ Y_i = 0 \end{cases}$$

ii) $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{\beta}} \frac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$

$$
= \begin{bmatrix}
\frac{d}{d\beta_0}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_1}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_2}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\
\frac{d}{d\beta_0}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_1}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_2}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\
\frac{d}{d\beta_0}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_1}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_2}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})
\end{bmatrix},
$$

where

Case $Y_i = 1$:

$$
\frac{d}{d\beta_0}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{D_{w_i} A_{w_i} - C_{w_i}^2}{A_{w_i}^2}
$$

$$
\frac{d}{d\beta_1}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_0}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i} - C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2}
$$

$$
\frac{d}{d\beta_2}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_0}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i} - C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2}
$$

$$
\frac{d}{d\beta_1}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i} - (p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2}{A_{w_i}^2}
$$

$$
\frac{d}{d\beta_2}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_1}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2}
$$

$$
\frac{d}{d\beta_2}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i} - (p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2}{A_{w_i}^2}
$$

Case $Y_i = 0$:

$$\frac{d}{d\beta_0}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-(D_{w_i} B_{w_i} + C_{w_i}^2)}{B_{w_i}^2}$$

$$\frac{d}{d\beta_1}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_0}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i} + C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)}{B_{w_i}^2}$$

$$\frac{d}{d\beta_2}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_0}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i} + C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)}{B_{w_i}^2}$$

$$\frac{d}{d\beta_1}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i} + (p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2\right)}{B_{w_i}^2}$$

$$\frac{d}{d\beta_2}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_1}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2}$$

$$\frac{d}{d\beta_2}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i} + (p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2\right)}{B_{w_i}^2}$$

And $D_{w_i} = \frac{\partial}{\partial\beta_0}C_{w_i} = p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3} + p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3} + p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}$

Condition N4: $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is a continuous function of $\underset{\sim}{\beta}$, uniformly in i, a.s. [P], and is measurable functions of Y_i

Proof:

From condition N3,

$$\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial\underset{\sim}{\beta}}\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{\partial}{\partial\underset{\sim}{\beta}}\frac{\partial}{\partial\underset{\sim}{\beta}}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$= \begin{bmatrix} \frac{d}{d\beta_0}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_1}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_2}\frac{d}{d\beta_0}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_0}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_1}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_2}\frac{d}{d\beta_1}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_0}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_1}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) & \frac{d}{d\beta_2}\frac{d}{d\beta_2}\Phi_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \end{bmatrix}$$

Note that each element is a sum, difference, and quotient of continuous function of $\underset{\sim}{\beta}$ and the denominator always greater than zero. Therefore $\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is continuous for each i. The uniformity also holds because Θ is a compact subset of R^p and argument from the proof of Condition C2.

$\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$ is measurable function of Y_i since continuous function is measurable.

Condition N5: $E\left[\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right] = \underset{\sim}{0};\ i = 1, 2, ...$

Proof:

$$E[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})] = \dot{\Phi}_i(Y_i = 1, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) f_i(y_i = 1 \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) + \dot{\Phi}_i(Y_i = 0, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) f_i(y_i = 0 \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$= \begin{bmatrix} \frac{C_{w_i}}{A_{w_i}} \\ \frac{1}{A_{w_i}} p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \frac{1}{A_{w_i}} p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} * A_{w_i} + \begin{bmatrix} \frac{-C_{w_i}}{B_{w_i}} \\ \frac{-1}{B_{w_i}} p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \frac{-1}{B_{w_i}} p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} * B_{w_i}$$

$$
=\begin{bmatrix} C_{w_i} \\ p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix}+\begin{bmatrix} -C_{w_i} \\ -p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ -p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix}=0_{3\times 1}
$$

Condition N6:

$$
\Gamma_i(\underset{\sim}{\beta})=E\left[\dot{\Phi}_i(y_i,w_i,\underset{\sim}{\beta},\underset{\sim}{P})\dot{\Phi}_i(y_i,w_i,\underset{\sim}{\beta},\underset{\sim}{P})^T\mid\underset{\sim}{\beta}\right]=-E\left[\ddot{\Phi}_i(y_i,w_i,\underset{\sim}{\beta},\underset{\sim}{P})\mid\underset{\sim}{\beta}\right]
$$

Proof:

$$
E[\dot{\Phi}_i(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\dot{\Phi}_i(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})^T]=
$$

$$
\dot{\Phi}_i(Y_i=1,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\dot{\Phi}_i(Y_i=1,W_i,\underset{\sim}{\beta},\underset{\sim}{P})^T f_i(y_i=1\mid w_i,\underset{\sim}{\beta},\underset{\sim}{P})+\dot{\Phi}_i(Y_i=0,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\dot{\Phi}_i(Y_i=0,W_i,\underset{\sim}{\beta},\underset{\sim}{P})^T f_i(y_i=0\mid w_i,\underset{\sim}{\beta},\underset{\sim}{P})
$$

$$
=\begin{bmatrix}
\dfrac{C_{w_i}^2}{A_{w_i}^2} & \dfrac{C_{w_i}p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2} & \dfrac{C_{w_i}p_{2w_i}\dfrac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} \\
\dfrac{C_{w_i}p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2} & \dfrac{\left(p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)^2}{A_{w_i}^2} & \dfrac{p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\dfrac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} \\
\dfrac{C_{w_i}p_{2w_i}\dfrac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} & \dfrac{p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\dfrac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} & \dfrac{\left(p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)^2}{A_{w_i}^2}
\end{bmatrix}*A_{w_i}
$$

$$
+\begin{bmatrix}
\frac{C_{w_i}^2}{B_{w_i}^2} & \frac{C_{w_i} p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{B_{w_i}^2} & \frac{C_{w_i} p_{2w_i} \frac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} \\
\frac{C_{w_i} p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{B_{w_i}^2} & \frac{(p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2}{B_{w_i}^2} & \frac{p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} p_{2w_i} \frac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} \\
\frac{C_{w_i} p_{2w_i} \frac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} & \frac{p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} p_{2w_i} \frac{e^{\beta_0+\beta 2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} & \frac{(p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2}{B_{w_i}^2}
\end{bmatrix} * B_{w_i}
$$

$$
=\begin{bmatrix}
\frac{C_{w_i}^2}{A_{w_i}}+\frac{C_{w_i}^2}{B_{w_i}} & (\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\
(\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2 & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\
(\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})(p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2
\end{bmatrix}
$$

Similarly,

$$-E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right] =$$

$$\ddot{\Phi}_i(Y_i=1, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\ddot{\Phi}_i(Y_i=1, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})^T f_i(y_i=1 \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) + \ddot{\Phi}_i(Y_i=0, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\ddot{\Phi}_i(Y_i=0, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})^T f_i(y_i=0 \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$
=\begin{bmatrix}
\frac{D_{w_i}A_{w_i}-C_{w_i}^2}{A_{w_i}^2} & \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i}-C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2} & \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i}-C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} \\
\frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i}-C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2} & \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i}-(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2}{A_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} \\
\frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i}-C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} & \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i}-(p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2}{A_{w_i}^2}
\end{bmatrix}*A_{w_i}
$$

$$
+\begin{bmatrix}
\frac{-(D_{w_i}B_{w_i}+C_{w_i}^2)}{B_{w_i}^2} & \frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i}+C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)}{B_{w_i}^2} & \frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i}+C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)}{B_{w_i}^2} \\
\frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i}+C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)}{B_{w_i}^2} & \frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i}+(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2\right)}{B_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} \\
\frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i}+C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)}{B_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} & \frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i}+(p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2\right)}{B_{w_i}^2}
\end{bmatrix}*B_{w_i}
$$

$$
=\begin{bmatrix}
\frac{C_{w_i}^2}{A_{w_i}}+\frac{C_{w_i}^2}{B_{w_i}} & (\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\
(\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2 & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\
(\frac{C_{w_i}}{A_{w_i}}+\frac{C_{w_i}}{B_{w_i}})p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i}}+\frac{1}{B_{w_i}})(p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2
\end{bmatrix}
$$

Condition N7: $\exists\bar{\Gamma}(\underset{\sim}{\beta})$ such that $\lim_{N\to\infty}\bar{\Gamma}_N(\underset{\sim}{\beta})=\bar{\Gamma}(\underset{\sim}{\beta})$, and $\bar{\Gamma}(\underset{\sim}{\beta})$ is positive definite.

Proof:

We will prove condition N7 in two steps:

i) $\bar{\Gamma}_N(\underset{\sim}{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})$ is positive definite.

ii) $\lim_{N\to\infty}\bar{\Gamma}_N(\underset{\sim}{\beta})$ exists, denoted by $\bar{\Gamma}(\underset{\sim}{\beta})$, $\bar{\Gamma}(\underset{\sim}{\beta})$ is positive definite.

Proof of i:

First we prove $\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})$ positive definite, then conclude $\bar{\Gamma}_N(\underset{\sim}{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})$ is also positive definite.

From condition N6 we can easily show $\Gamma_i(\underset{\sim}{\beta})$ is positive semi-definite for each i since each of its leading sub-matrices has zero determinant.

By condition N5, $E[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})] = 0$ for i = 1, 2, …Then

$$\Gamma_i(\underset{\sim}{\beta}) = E[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})^T] = Var[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})],\ \ i = 1, 2, ...$$

Let $X^T = (x_1, x_2, x_3) \in R^3$. Then

$$X^T\left[\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})\right]X = \sum_{i=1}^{N}\left[X^T\Gamma_i(\underset{\sim}{\beta})X\right] = \sum_{i=1}^{N}\left[X^T Var[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})]X\right] = \sum_{i=1}^{N}Var\left(X^T[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})]\right)$$

Suppose $X^T\left[\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})\right]X = \sum_{i=1}^{N}Var\left(X^T[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})]\right) = 0$

Since $Var\left(X^T[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})]\right) \geq 0$, we have $Var\left(X^T[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})]\right) = 0$ for each $i = 1, 2, ..., N$

Also, by condition N5, $E[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})] = 0$ for i = 1, 2, …, we imply $X^T[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})] = 0$.

Now, $0 = X^T[\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})] = X^T \begin{cases} \begin{bmatrix} \dfrac{C_{w_i}}{A_{w_i}} \\ \dfrac{1}{A_{w_i}} p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \dfrac{1}{A_{w_i}} p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \;\; Y_i = 1 \\ \begin{bmatrix} \dfrac{-C_{w_i}}{B_{w_i}} \\ \dfrac{-1}{B_{w_i}} p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \dfrac{-1}{B_{w_i}} p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \;\; Y_i = 0 \end{cases}$

$$= \begin{cases} X^T \begin{bmatrix} C_{w_i} \\ p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \;\; Y_i = 1 \\ X^T \begin{bmatrix} -C_{w_i} \\ -p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ -p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \;\; Y_i = 0 \end{cases} \quad \text{since } A_{w_i}, B_{w_i} \neq 0$$

or $0 = x_1 C_{w_i} + x_2 p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + x_3 p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}$

With each $w_i = 1, 2, \; or \; 3$; over all N we have:

$$x_1(p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{31}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2})+x_2p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+x_3p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}=0$$
$$x_1(p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{32}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2})+x_2p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+x_3p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}=0$$
$$x_1(p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{33}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2})+x_2p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+x_3p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}=0$$

Now let $A=\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}$, $B=\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta 2})^2}$, and $C=\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}$

For the above system to have $(x_1,x_2,x_3)=\underset{\sim}{0}$ as the only solution in order to assert the positive definite of $\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})$ we need to have the determinant of matrix $H=\begin{bmatrix}(p_{11}A+p_{21}B+p_{31}C) & p_{11}A & p_{21}B\\ (p_{12}A+p_{22}B+p_{32}C) & p_{12}A & p_{22}B\\ (p_{13}A+p_{23}B+p_{33}C) & p_{13}A & p_{23}B\end{bmatrix}$ does not equal zero.

Recall for a matrix $G=\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$, the determinant

$$\det(G)=a_{11}a_{22}a_{33}-a_{11}a_{32}a_{23}-a_{12}a_{21}a_{33}+a_{12}a_{31}a_{23}+a_{13}a_{21}a_{32}-a_{13}a_{31}a_{22}$$

Now,

$\det(H)=$

$$(p_{11}A+p_{21}B+p_{31}C)(p_{12}A)(p_{23}B)-(p_{11}A+p_{21}B+p_{31}C)(p_{13}A)(p_{22}B)-(p_{12}A+p_{22}B+p_{32}C)(p_{11}A)(p_{23}B)$$
$$+\ (p_{13}A+p_{23}B+p_{33}C)(p_{11}A)(p_{22}B)+(p_{12}A+p_{22}B+p_{32}C)(p_{21}B)(p_{13}A)-(p_{13}A+p_{23}B+p_{33}C)(p_{21}B)(p_{12}A)$$

$$=\ p_{11}p_{12}p_{23}A^2B+p_{21}p_{12}p_{23}AB^2+p_{31}p_{12}p_{23}ABC$$
$$-p_{11}p_{13}p_{22}A^2B-p_{21}p_{13}p_{22}AB^2-p_{31}p_{13}p_{22}ABC$$
$$-p_{11}p_{12}p_{23}A^2B-p_{11}p_{23}p_{22}AB^2-p_{11}p_{23}p_{32}ABC$$

$$+p_{11}p_{13}p_{22}A^2B + p_{11}p_{23}p_{22}AB^2 + p_{11}p_{22}p_{33}ABC$$

$$+p_{12}p_{21}p_{13}A^2B + p_{22}p_{21}p_{13}AB^2 + p_{32}p_{21}p_{13}ABC$$

$$-p_{13}p_{21}p_{12}A^2B - p_{23}p_{21}p_{12}AB^2 - p_{33}p_{21}p_{12}ABC$$

$$=\left(p_{31}p_{12}p_{23} - p_{31}p_{13}p_{22} - p_{11}p_{23}p_{32} + p_{11}p_{22}p_{33} + p_{32}p_{21}p_{13} - p_{33}p_{21}p_{12}\right)ABC$$

$$=\left[p_{11}\left(p_{22}p_{33} - p_{23}p_{32}\right) - p_{21}\left(p_{33}p_{12} - p_{32}p_{13}\right) + p_{31}\left(p_{12}p_{23} - p_{13}p_{22}\right)\right]ABC$$

$$=\det(\underset{\sim}{P})ABC$$

Since $ABC > 0$ and $\det(P) \neq 0$ (by assumption A5 $\underset{\sim}{P}$ is nonsingular), then $\det(H) \neq 0$.

Proof of ii:

From condition N6:

$$\Gamma_i(\underset{\sim}{\beta}) = E\left[\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})\dot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P})^T \mid \underset{\sim}{\beta}\right] = -E\left[\ddot{\Phi}_i(y_i, w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]$$

$$=\begin{bmatrix} \frac{C_{w_i}^2}{A_{w_i}} + \frac{C_{w_i}^2}{B_{w_i}} & \left(\frac{C_{w_i}}{A_{w_i}} + \frac{C_{w_i}}{B_{w_i}}\right)p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & \left(\frac{C_{w_i}}{A_{w_i}} + \frac{C_{w_i}}{B_{w_i}}\right)p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ \left(\frac{C_{w_i}}{A_{w_i}} + \frac{C_{w_i}}{B_{w_i}}\right)p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & \left(\frac{1}{A_{w_i}} + \frac{1}{B_{w_i}}\right)\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)^2 & \left(\frac{1}{A_{w_i}} + \frac{1}{B_{w_i}}\right)p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ \left(\frac{C_{w_i}}{A_{w_i}} + \frac{C_{w_i}}{B_{w_i}}\right)p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & \left(\frac{1}{A_{w_i}} + \frac{1}{B_{w_i}}\right)p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & \left(\frac{1}{A_{w_i}} + \frac{1}{B_{w_i}}\right)\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)^2 \end{bmatrix}$$

Where

$$A_{w_i} = p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}},$$

$$B_{w_i} = p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}, \text{ and}$$

$$C_{w_i} = \frac{\partial}{\partial\beta_0}A_{w_i} = p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}$$

So

$\bar{\Gamma}_N(\underset{\sim}{\beta})$

$$=\frac{n_1}{N}\begin{bmatrix} \frac{C_{w_i=1}^2}{A_{w_i=1}}+\frac{C_{w_i=1}^2}{B_{w_i=1}} & (\frac{C_{w_i=1}}{A_{w_i=1}}+\frac{C_{w_i=1}}{B_{w_i=1}})p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{C_{w_i=1}}{A_{w_i=1}}+\frac{C_{w_i=1}}{B_{w_i=1}})p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ (\frac{C_{w_i=1}}{A_{w_i=1}}+\frac{C_{w_i=1}}{B_{w_i=1}})p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{1}{A_{w_i=1}}+\frac{1}{B_{w_i=1}})(p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2 & (\frac{1}{A_{w_i=1}}+\frac{1}{B_{w_i=1}})p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ (\frac{C_{w_i=1}}{A_{w_i=1}}+\frac{C_{w_i=1}}{B_{w_i=1}})p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i=1}}+\frac{1}{B_{w_i=1}})p_{11}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i=1}}+\frac{1}{B_{w_i=1}})(p_{21}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2 \end{bmatrix}$$

$$+\frac{n_2}{N}\begin{bmatrix} \frac{C_{w_i=2}^2}{A_{w_i=2}}+\frac{C_{w_i=2}^2}{B_{w_i=2}} & (\frac{C_{w_i=2}}{A_{w_i=2}}+\frac{C_{w_i=2}}{B_{w_i=2}})p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{C_{w_i=2}}{A_{w_i=2}}+\frac{C_{w_i=2}}{B_{w_i=2}})p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ (\frac{C_{w_i=2}}{A_{w_i=2}}+\frac{C_{w_i=2}}{B_{w_i=2}})p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{1}{A_{w_i=2}}+\frac{1}{B_{w_i=2}})(p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2 & (\frac{1}{A_{w_i=2}}+\frac{1}{B_{w_i=2}})p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ (\frac{C_{w_i=2}}{A_{w_i=2}}+\frac{C_{w_i=2}}{B_{w_i=2}})p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i=2}}+\frac{1}{B_{w_i=2}})p_{12}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i=2}}+\frac{1}{B_{w_i=2}})(p_{22}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2 \end{bmatrix}$$

$$+\frac{n_3}{N}\begin{bmatrix} \frac{C_{w_i=3}^2}{A_{w_i=3}}+\frac{C_{w_i=3}^2}{B_{w_i=3}} & (\frac{C_{w_i=3}}{A_{w_i=3}}+\frac{C_{w_i=3}}{B_{w_i=3}})p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{C_{w_i=3}}{A_{w_i=3}}+\frac{C_{w_i=3}}{B_{w_i=3}})p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ (\frac{C_{w_i=3}}{A_{w_i=3}}+\frac{C_{w_i=3}}{B_{w_i=3}})p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} & (\frac{1}{A_{w_i=3}}+\frac{1}{B_{w_i=3}})(p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2 & (\frac{1}{A_{w_i=3}}+\frac{1}{B_{w_i=3}})p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \\ (\frac{C_{w_i=3}}{A_{w_i=3}}+\frac{C_{w_i=3}}{B_{w_i=3}})p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i=3}}+\frac{1}{B_{w_i=3}})p_{13}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} & (\frac{1}{A_{w_i=3}}+\frac{1}{B_{w_i=3}})(p_{23}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2 \end{bmatrix}$$

$$=\frac{n_1}{N}\Gamma_{w_i=1}(\underset{\sim}{\beta})+\frac{n_2}{N}\Gamma_{w_i=2}(\underset{\sim}{\beta})+\frac{n_3}{N}\Gamma_{w_i=3}(\underset{\sim}{\beta})$$

where $n_1=\sum_{i=1}^{N}I_{(w_i=1)}$, $n_2=\sum_{i=1}^{N}I_{(w_i=2)}$, $n_3=\sum_{i=1}^{N}I_{(w_i=3)}$, $\Gamma_{w_i=j}(\underset{\sim}{\beta})$ *is* $\Gamma_i(\underset{\sim}{\beta})$ *when* $w_i=j$

Since W is a random variable taking on three possible values, let $P(W=1)=C_1$, $P(W=2)=C_2$, and $P(W=3)=C_3$ with $C_i>0$, $C_1+C_2+C_3=1$. By law of large number, $\lim_{N\to\infty}\frac{n_1}{N}=C_1$, $\lim_{N\to\infty}\frac{n_2}{N}=C_2$, and $\lim_{N\to\infty}\frac{n_3}{N}=C_3$.

Thus $\lim_{N\to\infty}\bar{\Gamma}_N(\underset{\sim}{\beta})$ exists and denoted by $\bar{\Gamma}(\underset{\sim}{\beta})$,

$\bar{\Gamma}(\underset{\sim}{\beta})=C_1\Gamma_{w_i=1}(\underset{\sim}{\beta})+C_2\Gamma_{w_i=2}(\underset{\sim}{\beta})+C_3\Gamma_{w_i=3}(\underset{\sim}{\beta})$

Let $C_{\min} = \min\{C_1, C_2, C_3\} > 0$, then $\bar{\Gamma}(\underset{\sim}{\beta}) \geq C_{\min}\left(\Gamma_{w_i=1}(\underset{\sim}{\beta}) + \Gamma_{w_i=2}(\underset{\sim}{\beta}) + \Gamma_{w_i=3}(\underset{\sim}{\beta})\right)$

We will prove $\left(\Gamma_{w_i=1}(\underset{\sim}{\beta}) + \Gamma_{w_i=2}(\underset{\sim}{\beta}) + \Gamma_{w_i=3}(\underset{\sim}{\beta})\right)$ is positive definite and conclude that $\bar{\Gamma}(\underset{\sim}{\beta})$ is also positive definite.

Let $X^T = (x_1, x_2, x_3) \in R^3$, then

$$X^T\left[\Gamma_{w_i=1}(\underset{\sim}{\beta}) + \Gamma_{w_i=2}(\underset{\sim}{\beta}) + \Gamma_{w_i=3}(\underset{\sim}{\beta})\right]X = X^T\Gamma_{w_i=1}(\underset{\sim}{\beta})X + X^T\Gamma_{w_i=2}(\underset{\sim}{\beta})X + X^T\Gamma_{w_i=3}(\underset{\sim}{\beta})X$$
$$= X^T Var[\dot{\Phi}_i(Y_i, W_i = 1, \underset{\sim}{\beta}, \underset{\sim}{P})]X + X^T Var[\dot{\Phi}_i(Y_i, W_i = 2, \underset{\sim}{\beta}, \underset{\sim}{P})]X + X^T Var[\dot{\Phi}_i(Y_i, W_i = 3, \underset{\sim}{\beta}, \underset{\sim}{P})]X$$
$$= Var[X^T\dot{\Phi}_i(Y_i, W_i = 1, \underset{\sim}{\beta}, \underset{\sim}{P})] + Var[X^T\dot{\Phi}_i(Y_i, W_i = 2, \underset{\sim}{\beta}, \underset{\sim}{P})] + Var[X^T\dot{\Phi}_i(Y_i, W_i = 3, \underset{\sim}{\beta}, \underset{\sim}{P})]$$

From here, the same lines of proof for $\sum_{i=1}^{N}\Gamma_i(\underset{\sim}{\beta})$ positive definite in part (i) is used to show $\left(\Gamma_{w_i=1}(\underset{\sim}{\beta}) + \Gamma_{w_i=2}(\underset{\sim}{\beta}) + \Gamma_{w_i=3}(\underset{\sim}{\beta})\right)$ positive definite.

Therefore we conclude $\bar{\Gamma}(\underset{\sim}{\beta})$ positive definite.

Condition N8: $E\left[\dot{\Phi}_{i,j}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]^3 \leq K$

Proof:

From condition N3,

$$\dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \begin{bmatrix} \frac{d}{d\beta_0}\Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_1}\Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_2}\Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \end{bmatrix} = \begin{bmatrix} \frac{d}{d\beta_0}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_1}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \\ \frac{d}{d\beta_2}Lnf_i(y_i \mid w_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \end{bmatrix}$$

$$
= \begin{cases} \begin{bmatrix} \dfrac{C_{w_i}}{A_{w_i}} \\ \dfrac{1}{A_{w_i}} p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \dfrac{1}{A_{w_i}} p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \;\; Y_i = 1 \\ \begin{bmatrix} \dfrac{-C_{w_i}}{B_{w_i}} \\ \dfrac{-1}{B_{w_i}} p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ \dfrac{-1}{B_{w_i}} p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} \end{bmatrix} & if \;\; Y_i = 0 \end{cases}
$$

where, $A_{w_i} = p_{1w_i} \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \dfrac{e^{\beta_0}}{1+e^{\beta_0}}$

$$
B_{w_i} = p_{1w_i} \frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{1}{1+e^{\beta_0}}
$$

$$
C_{w_i} = \frac{\partial}{\partial \beta_0} A_{w_i} = p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2}
$$

Case $Y_i = 1$:

$$
\dot{\Phi}_{i,1}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{C_{w_i}}{A_{w_i}}
$$

$$= \frac{p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2}}{p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}}}$$

$\left|\dot{\Phi}_{i,1}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| < 1$ since both the numerator and denominator are positive and term by term

$$p_{1w_i} \frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} = p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \frac{1}{1+e^{\beta_0+\beta_1}} < p_{1w_i} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$$

$$p_{1w_i} \frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} = p_{1w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} \frac{1}{1+e^{\beta_0+\beta_2}} < p_{1w_i} \frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}$$

$$p_{1w_i} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2} = p_{1w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}} \frac{1}{1+e^{\beta_0}} < p_{1w_i} \frac{e^{\beta_0}}{1+e^{\beta_0}}$$

Let K = 1, then $E\left[\dot{\Phi}_{i,j}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) \mid \underset{\sim}{\beta}\right]^3 \le K$

Condition N9: There exists $\varepsilon > 0$ and random variables $B_{i,jl}(Y_i)$ such that

i) $\sup\left\{\left|\ddot{\Phi}_{i,jl}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| : \left\|\underset{\sim}{\beta} - \underset{\sim}{\beta}_0\right\| \le \varepsilon\right\} \le B_{i,jl}(Y_i)$.

ii) $E\left[B_{i,jl}(Y_i)\right]^{1+\delta} \le K$, for some $\delta > 0, K > 0$.

Proof:

i) Case $Y_i = 1$: From condition N3,

$$\ddot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \tfrac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$= \begin{bmatrix} \frac{D_{w_i}A_{w_i} - C_{w_i}^2}{A_{w_i}^2} & \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i} - C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2} & \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i} - C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} \\ \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i} - C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}}{A_{w_i}^2} & \frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}A_{w_i} - (p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2})^2}{A_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} \\ \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i} - C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{A_{w_i}^2} & \frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}A_{w_i} - (p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2})^2}{A_{w_i}^2} \end{bmatrix}$$

where, $A_{w_i} = p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}$

$$C_{w_i} = \frac{\partial}{\partial\beta_0}A_{w_i} = p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}$$

and $D_{w_i} = p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3} + p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3} + p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}$

First consider,

$$\ddot{\Phi}_{i,11}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_0}\frac{d}{d\beta_0}\Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{D_{w_i}A_{w_i} - C_{w_i}^2}{A_{w_i}^2} = \frac{D_{w_i}}{A_{w_i}} - \frac{C_{w_i}^2}{A_{w_i}^2}$$

$$=\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}+p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}+p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}\right)}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}-\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2}$$

$$\left|\ddot{\Phi}_{i,11}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|\le\left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}+p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}+p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}\right)}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}\right|$$

$$+\left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2}\right|$$

$$\le\left|\frac{p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}\right|$$

$$+\left|\frac{p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}}{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)}\right|$$

$$+\left|\frac{p_{3w_i}\dfrac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}}{\left(p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\dfrac{e^{\beta_0}}{1+e^{\beta_0}}\right)}\right|$$

$$+\left|\frac{\left(p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\dfrac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2}{\left(p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\dfrac{e^{\beta_0}}{1+e^{\beta_0}}\right)^2}\right|$$

$$<\left|\frac{p_{1w_i}\dfrac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}}{p_{1w_i}\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}\right|+\left|\frac{p_{2w_i}\dfrac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}}{p_{2w_i}\dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}}\right|+\left|\frac{p_{3w_i}\dfrac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}}{p_{3w_i}\dfrac{e^{\beta_0}}{1+e^{\beta_0}}}\right|+1$$

$$<\left|\frac{(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^2}\right|+\left|\frac{(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^2}\right|+\left|\frac{(1-e^{\beta_0})}{(1+e^{\beta_0})^2}\right|+1$$

$$<2+2+2+1=7$$

Similarly we have:

$$\left|\ddot{\Phi}_{i,12}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,21}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<4$$

$$\left|\ddot{\Phi}_{i,13}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,31}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<4$$

$$\left|\ddot{\Phi}_{i,22}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,33}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<4$$

$$\left|\ddot{\Phi}_{i,23}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,32}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<4$$

Case $Y_i = 0$: From condition N3,

$$\ddot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \dot{\Phi}_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \tfrac{\partial}{\partial \underset{\sim}{\beta}} \tfrac{\partial}{\partial \underset{\sim}{\beta}} \Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})$$

$$= \begin{bmatrix}
\frac{-(D_{w_i}B_{w_i}+C_{w_i}^2)}{B_{w_i}^2} & \frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i}+C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)}{B_{w_i}^2} & \frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i}+C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)}{B_{w_i}^2} \\
\frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i}+C_{w_i}p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)}{B_{w_i}^2} & \frac{-\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}B_{w_i}+\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}\right)^2\right)}{B_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} \\
\frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i}+C_{w_i}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)}{B_{w_i}^2} & \frac{-p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}}{B_{w_i}^2} & \frac{-\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}B_{w_i}+\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}\right)^2\right)}{B_{w_i}^2}
\end{bmatrix}$$

Where, $B_{w_i} = p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}$

First consider,

$$\ddot{\Phi}_{i,11}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{d}{d\beta_0}\frac{d}{d\beta_0}\Phi_i(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P}) = \frac{-(D_{w_i}B_{w_i}+C_{w_i}^2)}{B_{w_i}^2} = -\left(\frac{D_{w_i}}{B_{w_i}} + \frac{C_{w_i}^2}{B_{w_i}^2}\right)$$

$$= \frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3} + p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3} + p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)} - \frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} + p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2} + p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}$$

$$\left|\ddot{\Phi}_{i,11}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| \le \left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3} + p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3} + p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}} + p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}} + p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}\right|$$

$$+\left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}\right|$$

$$\le\left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}\right|+\left|\frac{\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}\right|$$

$$+\left|\frac{\left(p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}\right|+\left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2}+p_{2w_i}\frac{e^{\beta_0+\beta_2}}{(1+e^{\beta_0+\beta_2})^2}+p_{3w_i}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}+p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}+p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)^2}\right|$$

$$<\left|\frac{\left(p_{1w_i}\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^3}\right)}{\left(p_{1w_i}\frac{1}{1+e^{\beta_0+\beta_1}}\right)}\right|+\left|\frac{\left(p_{2w_i}\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^3}\right)}{\left(p_{2w_i}\frac{1}{1+e^{\beta_0+\beta_2}}\right)}\right|+\left|\frac{\left(p_{3w_i}\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^3}\right)}{\left(p_{3w_i}\frac{1}{1+e^{\beta_0}}\right)}\right|+1$$

$$=\left|\frac{e^{\beta_0+\beta_1}(1-e^{\beta_0+\beta_1})}{(1+e^{\beta_0+\beta_1})^2}\right|+\left|\frac{e^{\beta_0+\beta_2}(1-e^{\beta_0+\beta_2})}{(1+e^{\beta_0+\beta_2})^2}\right|+\left|\frac{e^{\beta_0}(1-e^{\beta_0})}{(1+e^{\beta_0})^2}\right|+1$$

$$<1+1+1+1=4$$

Similarly we have:

$$\left|\ddot{\Phi}_{i,12}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,21}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<2$$

$$\left|\ddot{\Phi}_{i,13}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,31}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<2$$

$$\left|\ddot{\Phi}_{i,22}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|=\left|\ddot{\Phi}_{i,33}(Y_i,W_i,\underset{\sim}{\beta},\underset{\sim}{P})\right|<2$$

$$\left|\ddot{\Phi}_{i,23}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| = \left|\ddot{\Phi}_{i,32}(Y_i, W_i, \underset{\sim}{\beta}, \underset{\sim}{P})\right| < 1$$

Now we define $B_{i,jl}(Y_i) = \begin{cases} 7 \ \ if \ \ Y_i = 1 \\ 4 \ \ if \ \ Y_i = 0 \end{cases}$ and condition N9(i) is satisfied.

ii) Condition N9(ii) is automatically hold since Y_i can take on only two values.

APPENDIX E

SAS PROGRAM FOR MONTE CARLO SIMULATION STUDY

```
/*****************************************************************************************
***
* Filename: Regression With Misclassification - Using Auxiliary Data.SAS
* Function: Perform the Monte Carlo simulation of proposed method for
"Logistic Regresstion
*           with Misclassified Covariate using Auxiliary Data"
* Output:   Estimation of parameter Beta's, standard error, and 95% confidence
interval
* Author:   Nathan N. Dong
* Created:  08/01/2008
* Modified: 04/08/2009
*****************************************************************************************
****/

ODS Listing Close;
ODS HTML;

PROC IML;
*****************************************************************************************
*;
  /***            Module Declaration Block          ***/

  ***  Randomly generate observed value matrix W of X;
  START Generate_W(N, D);
    W = j(1,N,0.);
        Seed = -1;
        m = 20080215;   ** Mean of the Poison random generaor;
        CALL RANPOI(Seed,m, W);  ** System clock is used as seed to
RANPOI;
```

```
        W = MOD(W, D) + 1;
    RETURN (W);
  FINISH Generate_W;

  ***  Generate true value matrix X from W with misclassification matrix P;
  START Generate_X(W, P, N);
    X = j(1,N,0.);
        Cut11 = P[1,1]; Cut21 = P[1,1]+P[2,1];
        Cut12 = P[1,2]; Cut22 = P[1,2]+P[2,2];
        Cut13 = P[1,3]; Cut23 = P[1,3]+P[2,3];

        Val1 = UNIFORM(X);  ** Call Uniform random generator with system
clock as seed;
        Val2 = UNIFORM(X);  ** Call Uniform random generator with system
clock as seed;
        Val3 = UNIFORM(X);  ** Call Uniform random generator with system
clock as seed;
        DO k=1 to N;
          IF W[k] = 1 THEN DO;
                IF Val1[k] < Cut11 THEN X[k] = 1;
                    ELSE IF Val1[k] < Cut21 THEN X[k] = 2;
                    ELSE X[k] = 3;
              END;
      ELSE IF W[k] = 2 THEN DO;
                IF Val2[k] < Cut12 THEN X[k] = 1;
                    ELSE IF Val2[k] < Cut22 THEN X[k] = 2;
                    ELSE X[k] = 3;
              END;
              ELSE IF W[k] = 3 THEN DO;
```

```
                  IF Val3[k] < Cut13 THEN X[k] = 1;
                      ELSE IF Val3[k] < Cut23 THEN X[k] = 2;
                      ELSE X[k] = 3;
                END;
                ELSE PUT 'Error in value range for X.';
          END;
      RETURN (X);
    FINISH Generate_X;



    ***  Generate binary outcome Y given X according to logistic regression
result;
    START Generate_Y(X, B0, B1, B2, N);
      Y = j(1,N,0.);
          pk = 0;
          Val = UNIFORM(Y);  ** Call Uniform random generator with system
clock as seed;
          DO k = 1 TO N;
        pk = exp(B0 + B1*(X[k]=1) + B2*(X[k]=2))/(1 + exp(B0 + B1*(X[k]=1) +
B2*(X[k]=2)));
        IF Val[k] < pk THEN Y[k] = 1;
                ** Else Y[k] = 0 already so we don't need to reassign;
          END;
      RETURN (Y);
    FINISH Generate_Y;

   ***  Estimate Pij from auxiliary data X2 and W2;
   START Estimate_P(W2, X2, D, M);
               P = j(D,D, 0.);
```

```
DO k = 1 to M;
   P[1,1] = P[1,1] + (X2[k]=1)*(W2[k]=1);
   P[2,1] = P[2,1] + (X2[k]=2)*(W2[k]=1);
   P[3,1] = P[3,1] + (X2[k]=3)*(W2[k]=1);

   P[1,2] = P[1,2] + (X2[k]=1)*(W2[k]=2);
   P[2,2] = P[2,2] + (X2[k]=2)*(W2[k]=2);
   P[3,2] = P[3,2] + (X2[k]=3)*(W2[k]=2);

   P[1,3] = P[1,3] + (X2[k]=1)*(W2[k]=3);
   P[2,3] = P[2,3] + (X2[k]=2)*(W2[k]=3);
   P[3,3] = P[3,3] + (X2[k]=3)*(W2[k]=3);
END;
Sum1 = P[1,1] + P[2,1] + P[3,1];
Sum2 = P[1,2] + P[2,2] + P[3,2];
Sum3 = P[1,3] + P[2,3] + P[3,3];
P[1,1] = P[1,1]/Sum1;
P[2,1] = P[2,1]/Sum1;
P[3,1] = P[3,1]/Sum1;

P[1,2] = P[1,2]/Sum2;
P[2,2] = P[2,2]/Sum2;
P[3,2] = P[3,2]/Sum2;

P[1,3] = P[1,3]/Sum3;
P[2,3] = P[2,3]/Sum3;
P[3,3] = P[3,3]/Sum3;
```

```
                RETURN (P);
  FINISH Estimate_P;


   ***  Evaluate the log likelihood function given current values of Beta  ;
   START log_likelihood(Beta) global(Y, W, Y2, X2, P);
     /* Beta[1] = Beta0, Beta[2] = Beta1  Beta[3] = Beta2 */
    D = ncol(P);   *** Dimension of missclassified matrix P;
          N = ncol(W);   **  Sample size of primary data;
    M = ncol(X2);  **  Sample size of auxiliary data;

          eta1 = exp(Beta[1] + Beta[2])/(1 + exp(Beta[1] + Beta[2]));
          eta2 = exp(Beta[1] + Beta[3])/(1 + exp(Beta[1] + Beta[3]));
          eta0 = exp(Beta[1])/(1 + exp(Beta[1]));
    a1 = eta1*P[1,1] + eta2*P[2,1] + eta0*P[3,1];
          a2 = eta1*P[1,2] + eta2*P[2,2] + eta0*P[3,2];
          a3 = eta1*P[1,3] + eta2*P[2,3] + eta0*P[3,3];

          ***   Do loop below will determine n1 (# of W'k=1), n11 (# of Y'k=1 given
W'k=1)
                                          n2 (# of W'k=2), n12 (# of Y'k=1 given W'k=2)
                                          n3 (# of W'k=3), n13 (# of Y'k=1 given W'k=3)   ;
          n1 = 0; n2 = 0; n3 = 0;
          n11 = 0; n12 = 0; n13 = 0;
          DO k = 1 TO N;
       n1 = n1 + (W[k]=1);
       n11 = n11 + (W[k]=1)*(Y[k]=1);
                n2 = n2 + (W[k]=2);
       n12 = n12 + (W[k]=2)*(Y[k]=1);
```

```
                n3 = n3 + (W[k]=3);
      n13 = n13 + (W[k]=3)*(Y[k]=1);
        END;   *** End do loop;
        m1 = 0; m2 = 0; m3 = 0;
        m11 = 0; m12 = 0; m13 = 0;
        DO k = 1 to M;
          m1 = m1 + (X2[k]=1);
      m11 = m11 + (X2[k]=1)*(Y2[k]=1);
                m2 = m2 + (X2[k]=2);
      m12 = m12 + (X2[k]=2)*(Y2[k]=1);
                m3 = m3 + (X2[k]=3);
      m13 = m13 + (X2[k]=3)*(Y2[k]=1);
        END;   *** End do loop;
        *print n1 n2 n3;
        ***   Calculate the log likelihood function according to the work-out
formula  ;
        f_log = n11*log(a1) + (n1-n11)*log(1-a1) + n12*log(a2) + (n2-n12)*log(1-
a2) +
         n13*log(a3) + (n3-n13)*log(1-a3) ;
        *f_log = n11*log(a1) + (n1-n11)*log(1-a1) + n12*log(a2) + (n2-
n12)*log(1-a2) +
         n13*log(a3) + (n3-n13)*log(1-a3) +
         m11*log(eta1) + (m1-m11)*log(1-eta1) + m12*log(eta2) + (m2-
m12)*log(1-eta2) +
         m13*log(eta0) + (m3-m13)*log(1-eta0) ;
        return(f_log);
  FINISH log_likelihood;
```

```
*****************************************************************************************
*;

   ***  Evaluate the gradient (first derivative) of log likelihood function
                                     given current values of Beta  ;
   START Glog_likelihood(Beta) global(Y, W, Y2, X2, P);
     /* Beta[1] = Beta0, Beta[2] = Beta1  Beta[3] = Beta2 */
    g = j(1,3, 0.);
    D = ncol(P);   *** Dimension of missclassified matrix P;
          N = ncol(W);
    M = ncol(X2);

          eta1 = exp(Beta[1] + Beta[2])/(1 + exp(Beta[1] + Beta[2]));
          eta2 = exp(Beta[1] + Beta[3])/(1 + exp(Beta[1] + Beta[3]));
          eta0 = exp(Beta[1])/(1 + exp(Beta[1]));
    a1 = eta1*P[1,1] + eta2*P[2,1] + eta0*P[3,1];
          a2 = eta1*P[1,2] + eta2*P[2,2] + eta0*P[3,2];
          a3 = eta1*P[1,3] + eta2*P[2,3] + eta0*P[3,3];

         e1 = exp(Beta[1] + Beta[2])/(1 + exp(Beta[1] + Beta[2]))##2;
         e2 = exp(Beta[1] + Beta[3])/(1 + exp(Beta[1] + Beta[3]))##2;
         e3 = exp(Beta[1])/(1 + exp(Beta[1]))##2;

          ***   Do loop below will determine n1 (# of W'k=1), n11 (# of Y'k=1 given
W'k=1)
                                     n2 (# of W'k=2), n12 (# of Y'k=1 given W'k=2)
                                     n3 (# of W'k=3), n13 (# of Y'k=1 given W'k=3)   ;
```

```
n1 = 0; n2 = 0; n3 = 0;
n11 = 0; n12 = 0; n13 = 0;
DO k = 1 TO N;
n1 = n1 + (W[k]=1);
n11 = n11 + (W[k]=1)*(Y[k]=1);
n2 = n2 + (W[k]=2);
n12 = n12 + (W[k]=2)*(Y[k]=1);
n3 = n3 + (W[k]=3);
n13 = n13 + (W[k]=3)*(Y[k]=1);
END;   *** End do loop;
m1 = 0; m2 = 0; m3 = 0;
m11 = 0; m12 = 0; m13 = 0;
DO k = 1 to M;
m1 = m1 + (X2[k]=1);
m11 = m11 + (X2[k]=1)*(Y2[k]=1);
m2 = m2 + (X2[k]=2);
m12 = m12 + (X2[k]=2)*(Y2[k]=1);
m3 = m3 + (X2[k]=3);
m13 = m13 + (X2[k]=3)*(Y2[k]=1);
END;   *** End do loop;

***   Calculate the gradident of log likelihood function according to the
work-out formula  ;
g[1] = (n11/a1 - (n1-n11)/(1-a1))*(e1*P[1,1] + e2*P[2,1] + e3*P[3,1]) +
(n12/a2 - (n2-n12)/(1-a2))*(e1*P[1,2] + e2*P[2,2] +
e3*P[3,2]) +
(n13/a3 - (n3-n13)/(1-a3))*(e1*P[1,3] + e2*P[2,3] +
e3*P[3,3]) ;
g[2] = (n11/a1 - (n1-n11)/(1-a1))*e1*P[1,1] +
```

```
                    (n12/a2 - (n2-n12)/(1-a2))*e1*P[1,2] +
                    (n13/a3 - (n3-n13)/(1-a3))*e1*P[1,3] ;
          g[3] = (n11/a1 - (n1-n11)/(1-a1))*e2*P[2,1] +
                    (n12/a2 - (n2-n12)/(1-a2))*e2*P[2,2] +
                    (n13/a3 - (n3-n13)/(1-a3))*e2*P[2,3] ;
          *g[1] = (n11/a1 - (n1-n11)/(1-a1))*(e1*P[1,1] + e2*P[2,1] + e3*P[3,1]) +
                    (n12/a2 - (n2-n12)/(1-a2))*(e1*P[1,2] + e2*P[2,2] +
e3*P[3,2]) +
                    (n13/a3 - (n3-n13)/(1-a3))*(e1*P[1,3] + e2*P[2,3] +
e3*P[3,3]) +
          (m11/eta1 - (m1-m11)/(1-eta1))*e1 +
                    (m12/eta2 - (m2-m12)/(1-eta2))*e2 +
                    (m13/eta0 - (m3-m13)/(1-eta0))*e3 ;
          *g[2] = (n11/a1 - (n1-n11)/(1-a1))*e1*P[1,1] +
                    (n12/a2 - (n2-n12)/(1-a2))*e1*P[1,2] +
                    (n13/a3 - (n3-n13)/(1-a3))*e1*P[1,3] +
          (m11/eta1 - (m1-m11)/(1-eta1))*e1 ;
          *g[3] = (n11/a1 - (n1-n11)/(1-a1))*e2*P[2,1] +
                    (n12/a2 - (n2-n12)/(1-a2))*e2*P[2,2] +
                    (n13/a3 - (n3-n13)/(1-a3))*e2*P[2,3] +
                    (m12/eta2 - (m2-m12)/(1-eta2))*e2 ;

          return(g);
  FINISH Glog_likelihood;


*************************************************************************************
*;
```

```
****************************************************************************************
*;
  ***   Initialization and Main Calling Block          ;

  R = 500;                          ** Number of replications used to average the
estimate parameters;
  N = 500;                 ** Sample size for main data;
  M = 100;                          ** Sample size for auxiliary data;
  D = 3;                            ** Number of category of X and W;
  X1 = j(1,N,0.);          ** True value X;
  W1 = j(1,N,0.);          ** Misclassified version of X;
  Y1 = j(1,N,0.);          ** Binary outcome Yi;
  X2 = j(1,M,0.);          ** True value X2 for auxiliary data;
  W2 = j(1,M,0.);          ** Misclassified version of X1 for auxiliary data;
  Y2 = j(1,M,0.);          ** Binary outcome Y2 for auxiliary data;
  Beta_Opt = j(R,D,0.);      ** Hold the estimate parameters of R replications;
  Std_Err = j(R,D,0.);     ** Hold the estimated standard error of Beta_Opt;

  ***  Pij = Pr(X=i|W=j), this matrix is used to generate X from W of the
simulation data;
  P_Temp = {0.9    0.05  0.05,
          0.05   0.9   0.05,
          0.05   0.05  0.9};
  *P_Temp = {0.80   0.15  0.00,
           0.20   0.70  0.20,
```

```
            0.00   0.15  0.80};   ** Matrix taken from Kuchenhoff paper which is not
work using his method ;
                                              ** The Misclassification SIMEX (MC-
SIMEX);
  *P_Temp = {0.50   0.35  0.15,
            0.30   0.40  0.25,
            0.20   0.25  0.60};   ** Example of badly misclassified matrix;

  Beta0 = 0.0;    **  True intercept;
  *Beta1 = 0.3;
  *Beta2 = -0.5;    **  True parameter in model logit(Y) = Beta0 + Beta1*(Xi=1) +
Beta2(Xi=2) ;
  Beta1 = 1;
  Beta2 = -1.5;

  Init_Beta = {0.5 0.5 0.5};
  optn = {1 2 3};
  const = {1.e-6 1.e-6 1.e-6, . . .};

  DO k = 1 to R;      * do loop through R replication;
    *** Three function calls below to generate auxiliary data;
    W2 = Generate_W(M, D);
    X2 = Generate_X(W2, P_Temp, M);
    Y2 = Generate_Y(X2, Beta0, Beta1, Beta2, M);
          P = Estimate_P(W2, X2, D, M);   ** Estimate Pij from auxiliary data X2
and W2;

    *** Three function calls below to generate data for main study;
    W1 = Generate_W(N, D);
```

```
   X1 = Generate_X(W1, P, N);
   Y1 = Generate_Y(X1, Beta0, Beta1, Beta2, N);

      W = W1;              ** This is primary data;
      Y = Y1;

      /*  Call the Nonlinear Optimization by Trust Region Method   */
   *call nlptr(rc,xres,"log_likelihood",Init_Beta,optn,const,,,,"Glog_likelihood");
   call nlptr(rc,xres,"log_likelihood",Init_Beta,optn) grd = "Glog_likelihood";
   Beta_Opt[k,1:D] = xres;
      print k xres;

  ** Call NLPFDD - Approximates derivatives by finite diferences method;
  ** Input: log likelihood function, estimated parameter, and optional Gradient
of log likelihood;
  ** Output: f = fun(Beta_Hat), g = grad(Beta_Hat), and h =
Hessian(Beta_Hat);
  call nlpfdd(f, g, h, "log_likelihood", Beta_Opt[k,1:D], , "Glog_likelihood");
  h_inverse = inv(h);
  Std_Err[k,1:D] = sqrt(abs(vecdiag(h_inverse)))`;

   if k=R then do;
    print k xres;
          print P;
   end;
 END;             * End do loop;


 Beta_Hat = Beta_Opt[:,]`;             * Mean of each column;
```

```
  stderr = Std_Err[:,]`;

  ** Calculate 95% confidence interval of the estimated Beta_Hat;
  prob = 0.05;   *** quantile of normal distribution;
  noqua = probit(1.0 - prob/2);
  Beta_Low = Beta_Hat - noqua*stderr;
  Beta_Up = Beta_Hat + noqua*stderr;
  Bias = j(3,1,0);
  Bias[1,1] = Beta_Hat[1,1] - Beta0;
  Bias[2,1] = Beta_Hat[2,1] - Beta1;
  Bias[3,1] = Beta_Hat[3,1] - Beta2;

  print "Estimate Parameter, 95% Normal Confidence Interval, Standard Error";
  print Beta_Hat Bias stderr Beta_Low Beta_Up;

*************************************************************************************
*;
ODS HTML Close;
ODS Listing;
QUIT;
RUN;
```

REFERENCES

Aigner, D. (1973). Regression with a binary independent variable subject to errors of observations. Journal of Econometrics, 1, pp. 49-60.

Apostol, T. M. (1975). Mathematical Analysis, Second Edition. Addison-Wesley Publishing Company.

Bain, L. J. and Engelhardt, M. (1992). Introduction to Probability and Mathematical Statistics, Second Edition. Duxbury Press.

Bollinger, C. (1996). Bounding mean regressions when a binary regressor is mismeasured. Journal of Econometrics, 73, pp. 387-399.

Carroll, Raymond, Ruppert, David etc. (2006). Measurement Error in Nonlinear Models: A Modern Perspective. Chapman & Hall/CRC

Cheng, Chi-Lun and Van Ness, John (1999). Statistical Regression with Measurement Error. Kendall's Library of Statistics 6.

Fuller, Wayne (1987). Measurement Error Models. John Wiley & Sons, Inc.

Gustafson, Paul (2005). Measurement Error and Misclassification in Statistics and Epidemiology. Chapman & Hall/CRC.

Hoadley, Bruce (1971). Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case. The Annals of Mathematical Statistics Vol. 42, No. 6, 1977-1991.

Hu, Yingyao (2008). Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: a General Solution. Journal of Econometrics, 144, p. 27-61.

Jones, Frank (1993). Lebesgue Integration on Euclidean Space. Jones and Bartlett Publishers.

Kuchenhoff, Helmut etc. (2006). A General Method for Dealing with Misclassification in Regression: The Misclassification Simex. Biometrics 62, 85-96.

Kutner, M., Nachtsheim, C., Neter, J., and Li W. (2005). Applied Linear Statistical Models, Fifth Edition. McGraw Hill.

Mahajan, Aprajit (2006). Identification and Estimation of Regression Models with Misclassification. Econometrica, Vol. 74, 631-665.

Molinari, Francesca (2008). Partial Identification of Probability Distribution of Misclassified Data. Journal of Econometrics, 144, p. 81-117.

Nguyen, Vinh (2003). Regression Model with Explanatory Variable Subject to Asymmetric Measurement Errors, Ph.D. Thesis. The University of Texas at Arlington.

Royden, H. L. (1968). Real Analysis, Second Edition. The MacMillan Compary.

Schneeweiss, Hans and Augustin, Thomas (2006). Some Recent Advances in Measurement Error Models and Methods. Allgemeines Statistisches Archiv 90, 183-197.

Serfling, R. J. (2002). Approximation Theorems of Mathematical Statistics. John Wiley & Sons.

Swartz, T., Haitovsky, Y., Vexler, A, and Yang, T. (2004). Bayesian Identifiability and Misclassification in Multinomial Data. Canadian Journal of Statistics, 32, 285-302.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20 595-601.

BIOGRAPHICAL INFORMATION

Nathan Dong received his Bachelor in Computer Science in 2001, Master of Science in Mathematics in 2004, and Doctor of Philosophy in Mathematics in August 2009. All three degrees were from the University of Texas at Arlington. He is also a certified Base and Advanced Programmer for SAS 9 from the SAS Institute.