Simulation-based estimation in regression models with categorical response variable and mismeasured covariates

by

Rojiar Haddadian

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Statistics
University of Manitoba
Winnipeg

Abstract

A common problem in regression analysis is that some covariates are measured with errors. In this dissertation we present simulation-based approach to estimation in two popular regression models with a categorical response variable and classical measurement errors in covariates. The first model is the regression model with a binary response variable. The second one is the proportional odds regression with an ordinal response variable.

In both regression models we consider method of moments estimators for therein unknown parameters that are defined via minimizing respective objective functions. The later functions involve multiple integrals and make obtaining of such estimators unfeasible. To overcome this computational difficulty, we propose Simulation-Based Estimators (SBE). This method does not require parametric assumptions for the distributions of the unobserved covariates and error components. We prove consistency and asymptotic normality of the proposed SBE's under some regularity conditions. We also examine the performance of the SBE's in finite-sample situations through simulation studies and two real data sets: the data set from the AIDS Clinical Trial Group (ACTG175) study for our logistic and probit regression models and one from the Adult Literacy and Life Skills (ALL) Survey for our regression model with the ordinal response variable and mismeasured covariates.

Keywords and phrases: Cumulative logit model, Instrumental variables, Measurement error, Method of moments, Ordinal response, Probit model, Proportional odds model, Simulation-based estimation.

Acknowledgment

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production.

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Liqun Wang, who has provided me all support and guidance to do my Ph.D at the Department of Statistics and has shared his knowledge and experience generously to overcome the difficulties I faced in my research. I would also heartily thank my co-advisor Dr. Yuliya Martsynyuk for her constant supervision as well as the time she dedicated to our regular meetings. Without her precious support and patience it would not be possible to conduct this research.

Besides my advisor and co-advisor, I would like to thank the rest of my thesis committee, Dr. Xikui Wang and Dr. Mamoud Torabi from University of Manitoba, and Dr. Wenqing He from University of Western Ontario, for their time, valuable suggestions and encouragement to improve this research.

I am also thankful to every member of the Department of Statistics at the University of Manitoba for assisting me in many different ways over the past six years.

Most importantly, this work would not have been possible without the love and patience of my family. I would like to express my heart-felt gratitude to my parents and to my brothers for supporting me spiritually through my life and writing this thesis.

Many friends have contributed to the great cheerfulness of my life and have helped me stay prudent through past years. I greatly value their friendship and I deeply appreciate their belief in me.

Last but not least, I would like to acknowledge the financial support provided during my Ph.D. studies through the NSERC Canada Individual Discovery grants of my advisors and funds from the Department of Statistics, the Faculty of Science and the Faculty of Graduate Studies of the University of Manitoba. In particular, I appreciate receiving the University of Manitoba Graduate Fellowship and the Manitoba Graduate Scholarship.

Dedication

This thesis is dedicated to my beloved parents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Measurement error models

Regression analyses often contain covariates that can not be directly or precisely measured and indirect or mismeasured covariates are used instead. For example, blood pressure and cholesterol level in cardiovascular disease research, or CD4 cell counts in studies of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) cannot be accurately measured because of instruments limitation or individual biological variation. It is well-known in the literature that ignoring measurement error will result in biased estimates and thus unreliable or misleading conclusions (cf., Stefanski and Buzas (1995)). In this section, we introduce three widely used measurement error models, (cf. Carroll et al. (2006)).

1.1.1 Classical measurement error

We suppose that $\mathbf{X}$ is unobserved and instead we observe

$$\mathbf{W} = \mathbf{X} + \delta, \tag{1.1}$$

where δ is a measurement error that is independent identically distributed (i.i.d.) with a zero mean and finite covariance matrix and is independent of $\mathbf{X}$. Some of the components of δ may have zero variances so that the corresponding covariates (components of $\mathbf{X}$) are precisely measured.

1.1.2 Berkson measurement error

The Berkson measurement error model is given by

$$\mathbf{X} = \mathbf{W} + \delta, \tag{1.2}$$

where δ is a measurement error that is independent of $\mathbf{W}$, and is independent identically distributed (i.i.d.) with zero mean and a positive-definite variance matrix.

1.1.3 Multiplicative measurement error

The multiplicative measurement error model has the form

$$\mathbf{W} = \mathbf{X}\delta \tag{1.3}$$

where δ is a measurement error that is independent of $\mathbf{X}$, and is independent and identically distributed (i.i.d.) with zero mean and a positive-definite variance matrix.

1.2 Instrumental Variable

In this thesis, we consider the classical measurement error model (1.1). It is well-known in the literature that the error model 1.1 is not identifiable, if the observed data only consist of $\mathbf{W}$ and the error distribution is unknown, Fuller (1987). To overcome this non-identifiability without additional data sources, one needs to make a distributional assumption on the error δ, and the corresponding parameters of this distribution need to be known or estimated by external data. The distributional assumption may be restrictive, and usually sensitivity analysis is required to assess the bias on estimation if the assumption is violated. In analysis with measurement error models, additional data sources are often needed.

Sometimes a second measurement of X, is available from another measurement method. Following the literature, this variable is called an instrumental variable and we denote it by $\mathbf{Z}$. An instrumental variable is defined as a random variable that is correlated with $\mathbf{X}$ but uncorrelated with measurement error and other random error terms in the model (Fuller (1987); Carroll et al. (2006))

1.3 Literature review of existing methods on generalized linear model with measurement error

Regression models with categorical response variables and missmeasured covariates are widely used in many scientific disciplines such as medicine and health, engineering, business and economics, and social sciences. For these models the ordinary least squares (OLS) method is biased and no longer provides the best unbiased estimator. The categorical dependent variable regression models are not linear and this nonlinearity results in difficulties of estimating the therein unknown parameters. These models belong to the family of generalized linear models (GLM).

Nakamura (1990), Stefanski and Carroll (1991), Buzas and Stefanski (1996), and Ma and Tsiatis (2006) studied approximation methods of estimation such as corrected score functions, deconvolution-based score, and conditional score function for GLM with mismeasured covariates. However, most of these approaches are applicable under the normality assumption of measurement errors and provide approximately consistent estimators.

The instrumental variable method has been widely used in GLM, for example in Buzas and Stefanski (1996), where the unobserved predictor is treated as a vector of fixed constants. They studied conditional score function estimation under the assumption that measurement error and instrumental variables are both conditionally and normally distributed.

Method of moments estimation for nonlinear models with Berkson type measurement error was investigated by Wang (2003) and Wang (2004), where a so-called Second-Order Least Squares Estimator (SLSE) based on the first two conditional moments of the response variables given the observed covariates was proposed. Under some regularity conditions, the SLSE was shown to be consistent and asymptotically normally distributed in these papers.

Later, Wang (2007) extended the SLSE method to nonlinear mixed effects models with homoscedastic errors, and Wang and Leblanc (2008) compared this estimator to the OLS estimator in general nonlinear models. Furthermore, Wang and Hsiao (1995) used the instrumental variable approach for general nonlinear measurement error models. They considered method of moments estimators for the therein unknown parameters that are defined via minimizing a certain objective function. The later involves multiple integrals that makes obtaining of such estimators unfeasible. To overcome this computational difficulty, Wang and Hsiao (2011) proposed and studied simulation-based estimators. Recently, Abarin and Wang (2012) used a similar simulation-based estimation method for GLM which allows very general heteroscedastic regression errors. Their approach does not require parametric assumptions for the distributions of the unobserved covariates and measurement errors that are difficult to check in practice.

Although the method of moments estimation for GLM with mismeasured covariates have been extensively studied in the literature, most works are concerned with continuous response variables only, and works focusing on categorical response variables, specially on ordinal variables, are very limited.

1.4 Outline of the thesis

The focus of this thesis is to extend the SLSE approach of Wang (2003) further to regression models with the categorical response variable and mismeasured covariates, via using instrumental variables. Since the closed forms of conditional moments are not available for these models, we propose Simulation-Based Estimators (SBE) for consistent estimation of the parameters of the model. This method does not require normality assumption on covariates and measurement errors, and can be considered as a generalization of Buzas and Stefanski (1996) with much more flexible assumptions.

Estimation in logistic regression models with mismeasured covariates and non-normal random errors have being considered before, but most of the authors rely on restrictive conditions to achieve consistent estimation. Furthermore, most of the methods in the literature are limited to the case where either validation or replicate data are available, which are restrictive conditions in many real data experiments. In the first part of this thesis, Chapter 2, we propose and study SBE's on the basis of SLSE approach for unknown parameters of a regression model with a binary response variable and classical measurement error in covariates, via using instrumental variables. Regression models for binary outcomes are the foundation for studying more complex models, such as ordinal, nominal, and count models. Consistency and asymptotic normality for the proposed SBE's are derived under some regularity conditions. Furthermore, the performance of the SBE's is illustrated through simulation studies and a real data set from the AIDS Clinical Trial Group (ACTG175) study.

Estimation in regression models with ordinal response variables have been considered before and a popular approach to the regression analysis of these variables is to treat the ordinal responses as discretized continuous responses (cf., e.g., Agresti (2003)). But there are limited papers about this model with mismeasured covariates. In the second part of this thesis, Chapter 3, we extend the results obtained in Chapter 2 for the regression model with binary responses to allow for ordinal responses. In other words, we study SBE's for regression models with an ordinal response variable and classical measurement error in covariates, via using instrumental variables. We derive consistency and asymptotic normality for the SBE's under some regularity conditions. Furthermore, we use simulations to illustrate the performance of the SBE's in finite sample situation. The real data set from the Adult Literacy and Life Skills (ALL) Survey is also applied in this regard.

Finally, conclusions and a discussion are contained in Chapter 4. We briefly summarize our overall findings and outline possible extensions for future work. The proofs of our theorems stated in Chapters 2 and 3 are given in Appendix.

Chapter 2

Regression models with binary response variable and mismeasured covariates

Consider a binary response random variable (r.v.) Y taking values one and zero with probabilities

$$P(Y = 1|\mathbf{X} = \mathbf{x}) = p(\alpha + \beta'\mathbf{x}) \tag{2.1}$$

and $P(Y = 0|\mathbf{X} = \mathbf{x}) = 1 - p(\alpha + \beta'\mathbf{x})$ that depend on some covariate $\mathbf{X} \in \mathbf{R}^p$, where $\beta \in \mathbf{R}^p$ and $\alpha \in \mathbf{R}$ are unknown parameters. For example, for the logistic model $p(\alpha + \beta'\mathbf{x}) = [\mathbf{1} + \exp(-\alpha - \beta'\mathbf{x})]^{-\mathbf{1}}$, and for the probit model $p(\alpha + \beta'\mathbf{x}) = \mathbf{\Phi}(\alpha + \beta'\mathbf{x})$, where Φ stands for the cumulative distribution function of the standard normal distribution.

Real data analyses of (2.1) often involve covariates that are not observed directly or are measured with error. We suppose that $\mathbf{X}$ is unobserved and instead we observe

$$\mathbf{W} = \mathbf{X} + \delta, \tag{2.2}$$

where δ is a measurement error that has a zero mean and finite covariance matrix. Some of the components of δ may have zero variances so that the corresponding covariates (components of $\mathbf{X}$) are precisely measured.

It is well-known in the literature that in the presence of measurement errors, supplementary information is required for model identification that allow consistent estimation of unknown parameters. One practical and useful information can be obtained through instrumental variables (Wang and Hsiao (2011), Xu et al. (2015)). Following the literature, we assume that an instrumental (vector) variable $\mathbf{Z} \in \mathbf{R}^q$, $q \geq p$ is available that is related to $\mathbf{X}$ through

$$\mathbf{X} = \mathbf{H}\mathbf{Z} + \mathbf{U}, \tag{2.3}$$

where $\mathbf{H}$ is a $p \times q$ matrix of unknown parameters with rank p, $\mathbf{U}$ is a random vector independent of $\mathbf{Z}$ with $E(\mathbf{U}) = \mathbf{0}$ and a probability density function (pdf) $f_{\mathbf{U}}(\mathbf{u}; \theta)$ with an unknown vector of parameters $\theta \in \mathbf{\Theta} \subset \mathbf{R}^k$. Further, we assume that $E(\delta|\mathbf{X}, \mathbf{Z}) = \mathbf{0}$. In the measurement error literature, it is common to assume that the measurement error is non-differential, so that given the true covariate $\mathbf{X}$, the conditional distribution of Y does not depend on the random vectors $\mathbf{W}$ and $\mathbf{Z}$. Here, we make a weaker assumption: $E(Y|\mathbf{X}, \mathbf{W}, \mathbf{Z}) = E(Y|\mathbf{X})$.

Note that there is no assumption concerning the functional forms of the distributions of $\mathbf{X}$ and δ and in this sense our model is semiparametric. Our interest is to estimate the vector of unknown parameters $\gamma = (\alpha, \beta', \theta')' \in \mathbf{\Gamma}$, where the parameter space $\mathbf{\Gamma} = \mathbf{A} \times \mathbf{B} \times \mathbf{\Theta}$ is compact in $\mathbf{R}^{p+k+1}$. We denote the true parameter value by $\gamma_{\mathbf{0}} \in \mathbf{\Gamma}$.

There are two main approaches to statistical estimation and inference in logistic regression models with mismeasured covariates. The first approach is the likelihood based one, which has been taken by several authors. For example, Stefanski and Carroll (1985) introduced a biased-adjusted estimator and a functional maximum likelihood estimator appropriate for normally distributed measurement errors. Stefanski and Buzas (1995) described two approaches to instrumental variable estimation in binary regression models. Their methods are fairly general, but usually result in only approximately consistent estimators. Huang and Wang (2001) proposed parametric and nonparametric correction estimation procedures in the presence of classical measurement error. Their parametric procedure requires the error distribution to be known, whereas their nonparametric procedure relaxes this distributional assumption requirement, but requires additional replicate data or instrumental variables to be available. Rabe-Hesketh et al. (2003) considered logistic models with possibly non-normal exposures measured with error. It should be noted that despite the discreteness of nonparametric maximum likelihood estimation, their method does not assume a truly discrete distribution.

The second approach is moments based, which does not require parametric assumptions for the distribution of the unobserved covariates and measurement errors. Although the method of moments estimation for GLM with mismeasured covariates have been extensively studied in the literature, most works are concerned with continuous response variables only, and works focusing on binary or categorical response variables are very limited. Bollen et al. (2008) studied an ordinal probit regression model with mismeasured covariates. They proposed a consistent instrumental variable estimator under the normality assumption on the unobserved covariates and measurement error. Recently, Xu et al. (2015) proposed an estimator based on the efficient score function combined with instrumental variable, which is fairly efficient under semiparametric setup.

In this chapter, we first consider method of moments estimators that are defined via the SLSE approach of Wang (2003) and Wang (2004). Since the closed forms of conditional moments are not available for the regression model with a binary response variable, we construct SBE's for γ in Section 2.1. This does not require normality assumption on $\mathbf{X}$, δ and U and can be considered as a generalization of Buzas and Stefanski (1996) with much more flexible assumptions. Consistency and asymptotic normality of our proposed SBE's under some regularity assumptions are proved in Section 2.2. In Section 2.3, we illustrate the performance of the proposed estimators in finite sample situation. In Section 2.4, a real data application is given for the data set from the AIDS Clinical Trial Group (ACTG175) study.

2.1 Construction of simulation-based estimators

In this section, we propose simulation-based consistent estimators for the vector $\gamma = (\alpha, \beta', \theta')'$ of the parameters in (2.1) - (2.3) based on a random sample $(Y_j, \mathbf{W}_j, \mathbf{Z}_j), j = 1, 2, ..., n$.

First, from combining (2.2) and (2.3) and using our model assumptions,

$$\begin{aligned} E(\mathbf{W}|\mathbf{Z}) &= E(\mathbf{X} + \delta|\mathbf{Z}) \\ &= E(\mathbf{HZ} + \mathbf{U} + \delta|\mathbf{Z}) \\ &= E(\mathbf{HZ}|\mathbf{Z}) + \mathbf{E}(\mathbf{U}|\mathbf{Z}) + \mathbf{E}(\delta|\mathbf{Z}) \\ &= \mathbf{HZ} \end{aligned} \tag{2.4}$$

It follows that an unknown matrix $\mathbf{H}$ can be consistently estimated by the least squares estimator

$$\hat{\mathbf{H}} = \left(\sum_{j=1}^{n} \mathbf{W}_j \mathbf{Z}_j' \right) \left(\sum_{j=1}^{n} \mathbf{Z}_j \mathbf{Z}_j' \right)^{-1}. \tag{2.5}$$

Further, by our model assumptions and the law of iterative expectation, we have

$$\begin{aligned} E(Y|\mathbf{Z}) &= E\left[E\left(Y|\mathbf{X},\mathbf{W},\mathbf{Z}\right)|\mathbf{Z}\right] \\ &= E\left[E\left(Y|\mathbf{X}\right)|\mathbf{Z}\right] \\ &= E\left[p(\alpha+\beta'\mathbf{X})|\mathbf{Z}\right] \\ &= E\left[p(\alpha+\beta'(\mathbf{HZ}+\mathbf{U}))|\mathbf{Z}\right] \\ &= \int p(\alpha+\beta'(\mathbf{HZ}+\mathbf{U}))\mathbf{f}_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du} \end{aligned} \tag{2.6}$$

and, similarly,

$$\begin{aligned} E(Y\mathbf{W}|\mathbf{Z}) &= E\left[\mathbf{W}E\left(Y|\mathbf{X},\mathbf{W},\mathbf{Z}\right)|\mathbf{Z}\right] \\ &= E\left[\mathbf{W}E\left(Y|\mathbf{X}\right)|\mathbf{Z}\right] \\ &= E\left[(\mathbf{X}+\delta)p(\alpha+\beta'\mathbf{X})|\mathbf{Z}\right] \\ &= E\left[\mathbf{X}p(\alpha+\beta'\mathbf{X})|\mathbf{Z}\right] + E\left[\delta p(\alpha+\beta'\mathbf{X})|\mathbf{Z}\right] \\ &= E\left[(\mathbf{HZ}+\mathbf{u})p(\alpha+\beta'(\mathbf{HZ}+\mathbf{U}))|\mathbf{Z}\right] + E\left[E(\delta p(\alpha+\beta'\mathbf{X})|\mathbf{X},\mathbf{Z})|\mathbf{Z}\right] \\ &= E\left[(\mathbf{HZ}+\mathbf{u})p(\alpha+\beta'(\mathbf{HZ}+\mathbf{U}))|\mathbf{Z}\right] + E\left[p(\alpha+\beta'\mathbf{X})\mathbf{E}(\delta|\mathbf{X},\mathbf{Z})|\mathbf{Z}\right] \\ &= E\left[(\mathbf{HZ}+\mathbf{u})p(\alpha+\beta'(\mathbf{HZ}+\mathbf{U}))|\mathbf{Z}\right] \\ &= \int (\mathbf{HZ}+\mathbf{u})p(\alpha+\beta'(\mathbf{HZ}+\mathbf{U}))\mathbf{f}_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du}. \end{aligned} \tag{2.7}$$

All the integrals in this work are taken over the space $\mathbf{R}^p$. Equation (2.6) and (2.7) can be used to derive method of moments estimators for $\gamma = (\alpha, \beta', \theta')'$ if the integrals therein have closed forms. However, the latter are difficult to find in general. In the following, we propose an estimator for γ based on a simulation approximation of the multiple integrals in (2.6) and (2.7), adapting the approach in Wang and Hsiao (1995), Wang and Hsiao (2011), and Abarin and Wang (2012).

To simplify notations, we denote $\mathbf{T} = (\mathbf{1} \quad \mathbf{W}')'$ and $\tilde{\mathbf{x}}' = (\mathbf{1} \quad \mathbf{x}')$ and write (2.6) and (2.7) together as

$$m(\mathbf{Hz}, \gamma) = E(Y\mathbf{T}|\mathbf{Z}) = \int \tilde{\mathbf{x}} p(\alpha + \beta'\mathbf{x}) \mathbf{f_U}(\mathbf{x} - \mathbf{HZ}; \theta)\mathbf{dx}. \tag{2.8}$$

Then we approximate $m(\mathbf{HZ}, \gamma)$ as follows. We start with choosing a known density $l(\mathbf{x})$ and generate a random sample $\{\mathbf{x}_{js}, s = 1, 2, \ldots, 2S, j = 1, 2, \ldots, n\}$ from $l(\mathbf{x})$. Then we approximate $m(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma)$ by the following two Monte-Carlo simulators:

$$m_1(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma) = \frac{1}{S} \sum_{s=1}^{S} \frac{\tilde{\mathbf{x}}_{js} p(\alpha + \beta'\mathbf{x}_{js})}{l(\mathbf{x}_{js})} f_U(\mathbf{x}_{js} - \hat{\mathbf{H}}\mathbf{Z}_j; \theta) \tag{2.9}$$

and

$$m_2(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma) = \frac{1}{S} \sum_{s=S+1}^{2S} \frac{\tilde{\mathbf{x}}_{js} p(\alpha + \beta'\mathbf{x}_{js})}{l(\mathbf{x}_{js})} f_U(\mathbf{x}_{js} - \hat{\mathbf{H}}\mathbf{Z}_j; \theta), \tag{2.10}$$

where $\tilde{\mathbf{x}}_{js} = (1 \quad \mathbf{x}'_{js})$. Finally, we define the simulation-based estimator (SBE) $\hat{\gamma}_{n,S}$ for γ as follows:

$$\hat{\gamma}_{n,S} = \arg\min_{\gamma\in\Gamma} Q_{n,S}(\gamma, \hat{h}) = \arg\min_{\gamma\in\Gamma} \sum_{j=1}^{n} \rho'_{j,1}(\gamma, \hat{h}) {\Omega_j}^{-1} \rho_{j,2}(\gamma, \hat{h}), \tag{2.11}$$

where

$$\hat{h} = vec\hat{\mathbf{H}} \tag{2.12}$$

and denotes the vector consisting of the columns of $\hat{\mathbf{H}}$,

$$\rho_{j,1}(\gamma, \hat{h}) = Y_j\mathbf{T}_j - m_1(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma),$$

(2.13)

$$\rho_{j,2}(\gamma, \hat{h}) = Y_j\mathbf{T}_j - m_2(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma),$$

and

$$\Omega_j = E(\rho_{j,1}(\gamma_0, h_0)\rho'_{j,2}(\gamma_0, h_0)|\mathbf{Z}_j). \tag{2.14}$$

In (2.14), Ω_j is so-called optimal weight matrix (cf. Abarin and Wang (2006)) and h_0 is the true value of h.

Note that

$$E[m_1(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma)|\mathfrak{W}, \mathfrak{Z}] = E[m_2(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma)|\mathfrak{W}, \mathfrak{Z}] = m(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma),$$

where $\mathfrak{W} = \{\mathbf{W_1}, \mathbf{W_2}, \ldots, \mathbf{W_n}\}$ and $\mathfrak{Z} = \{\mathbf{Z_1}, \mathbf{Z_2}, \ldots, \mathbf{Z_n}\}$. Therefore $m_1(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma)$ and $m_2(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma)$ are unbiased simulators for $m(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma)$. Using the two different sets of independent simulated points $\{\mathbf{x}_{js}, s = 1, 2, \ldots, S, j = 1, 2, \ldots, n\}$ and $\{\mathbf{x}_{js}, s = S+1, \ldots, 2S, j = 1, 2, \ldots, n\}$ from $l(\mathbf{x})$ leads to $Q_{n,S}(\gamma, \hat{h})$ being an unbiased estimator of

$$Q_n(\gamma, \hat{h}) = \sum_{j=1}^{n} \rho'_j(\gamma, \hat{h})\Omega_j{}^{-1}\rho_j(\gamma, \hat{h}), \tag{2.15}$$

where

$$\rho_j(\gamma, \hat{h}) = Y_j \mathbf{T}_j - m(\hat{\mathbf{H}}\mathbf{Z}_j, \gamma).$$

Indeed, since $\rho_{j,1}(\gamma, \hat{h})$ and $\rho_{j,2}(\gamma, \hat{h})$ are conditionally independent given $(Y_j, \mathfrak{W}, \mathfrak{Z})$, we have

$$\begin{aligned} E & \left[\rho'_{j,1}(\gamma, \hat{h}) \Omega_j{}^{-1} \rho_{j,2}(\gamma, \hat{h})\right] \\ = & E\left[E\left[\rho'_{j,1}(\gamma, \hat{h}) | Y_j, \mathfrak{W}, \mathfrak{Z}\right] \Omega_j{}^{-1} E\left[\rho_{j,2}(\gamma, \hat{h}) | Y_j, \mathfrak{W}, \mathfrak{Z}\right]\right] \\ = & E\left[\rho'_j(\gamma, \hat{h}) \Omega_j{}^{-1} \rho_j(\gamma, \hat{h})\right]. \end{aligned}$$

Remark 2.1. Note that in practice, Ω_j of (2.14) is a function of unknown parameters and therefore needs to be estimated. This can be done using the following two-stage procedure. First, minimize $Q_{n,S}(\gamma, \hat{h})$ as in (2.11) with the 2×2 identity matrix I_2 to obtain the first-stage estimator of γ, denoted by $\tilde{\gamma}_{n,s}$. Secondly, estimate Ω_j in (2.14) by $\tilde{\Omega} = \frac{1}{n} \sum_{j=1}^n \rho_{j,1}(\tilde{\gamma}_{n,S}, \hat{h}) \rho'_{j,2}(\tilde{\gamma}_{n,S}, \hat{h})$ or alternatively by a nonparametric estimator, and then minimize $Q_{n,S}(\gamma, \hat{h})$ again with $\tilde{\Omega}^{-1}$ to obtain the second-stage estimator, $\tilde{\tilde{\gamma}}_{n,s}$. Since $\tilde{\Omega}$ is consistent for Ω_j, the asymptotic covariance of the second-stage estimator is given by $E\left[\frac{\partial \rho'(\gamma_0, \hat{h})}{\partial \gamma} \Omega_j^{-1} \frac{\partial \rho'(\gamma_0, \hat{h})}{\partial \gamma}\right]^{-1}$. Consequently the second-stage estimator is asymptotically more efficient than the first-stage estimator.

For the choice of $l(\mathbf{x})$, in theory, it has no impact on the asymptotic efficiency of the estimator, as long as it has sufficiently large support. However, the choice of $l(\mathbf{x})$ will affect the finite sample variances of the simulated moments.

2.2 Asymptotic properties of simulation-based estimators

In Theorem 2.1 we will prove strong consistency of $\hat{\gamma}_{n,s}$ of (2.11), via authorizing the uniform convergence of $\frac{Q_{n,S}(\gamma,\hat{h})}{n}$ to a non-stochastic function $Q(\gamma,\hat{h})$ which has a unique minimizer $\gamma_0 \in \Gamma$. To do so, we will need the following regularity conditions.

Assumption 2.1. Assume that $E\|\Omega_j^{-1}\|(\|W_j\|+1)^2 < \infty$.

Assumption 2.2. The pdf $f_{\mathbf{U}}(\mathbf{u},\theta)$ of $\mathbf{U}$ is continuous in $\theta \in \Theta$ and continuously differentiable with respect to (w.r.t.) u for each $\theta \in \Theta$. Furthermore $f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)$ and $\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \mathbf{u}}$ for each $\theta \in \Theta$ are uniformly bounded by functions $\eta_1(\mathbf{x},\mathbf{z})$ and $\eta_2(\mathbf{x},\mathbf{z})$, respectively, which satisfy $E\|\Omega_j^{-1}\|\left(\int \eta_1(\mathbf{x},\mathbf{Z_j})(\|\mathbf{x}\|+1)d\mathbf{x}\right)^2 < \infty$ and $E\|\Omega_j^{-1}\|\left(\|\mathbf{Z_j}\|\int \eta_2(\mathbf{x},\mathbf{Z_j})(\|\mathbf{x}\|+1)d\mathbf{x}\right)^2 < \infty$.

Assumption 2.3. $E\left[(\rho_j(\gamma,h_0)-\rho_j(\gamma_0,h_0))'\Omega_j^{-1}(\rho_j(\gamma,h_0)-\rho_j(\gamma_0,h_0))\right] = 0$ if and only if $\gamma = \gamma_0$.

These assumptions are common in the literature on GLM. In particular Assumptions 2.1 and 2.2 contain moment conditions to assure that $\frac{Q_{n,S}(\gamma,\hat{h})}{n}$ uniformly converges to $Q(\gamma,h_0) = E\left[\rho_j'(\gamma,h_0)\Omega_j{}^{-1}\rho_j(\gamma,h_0)\right]$. Assumption 2.3 is an identification condition to ensure that $Q(\gamma,h_0)$ has a unique minimum at the true parameter value $\gamma_0 \in \Gamma_0$.

Theorem 2.1. Suppose that the support of $l(\mathbf{x})$ is $\mathbf{R}^p$. Then under Assumptions 2.1 - 2.3, $\hat{\gamma}_{n,s} \stackrel{a.s.}{\longrightarrow} \gamma_0$, as $n \longrightarrow \infty$.

To derive asymptotic normality for $\hat{\gamma}_{n,s}$ (cf. Theorem 2.2) , we assume additional regularity conditions as follows.

Assumption 2.4. There exists an open subset $\Gamma_0 \subset \Gamma$ in which functions $\mathbf{x} f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)$, $\mathbf{x}\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial\theta}$, $\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial\theta}$, and $\frac{\partial^2 f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial\theta^2}$ are uniformly bounded by a function $k(\mathbf{x},\mathbf{z})$ that satisfies $E\|\Omega_j^{-1}\| \left(\int k(\mathbf{x},\mathbf{Z_j})(\|\mathbf{x}\|+1)d\mathbf{x}\right)^2 < \infty$.

Assumption 2.5. There exist an open subset $\Gamma_0 \subset \Gamma$ in which functions $\mathbf{x}\frac{\partial \mathbf{f}_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \mathbf{u}'}$ and $\frac{\partial^2 \mathbf{f}_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial\theta\partial u'}$ are uniformly bounded by a function $r(\mathbf{x},\mathbf{z})$ that satisfies $E\|\Omega_j^{-1}\| \left(\|\mathbf{Z_j}\| \int r(\mathbf{x},\mathbf{Z_j})(\|\mathbf{x}\|+1)d\mathbf{x}\right)^2 < \infty$.

Assumption 2.6. The matrix

$$\mathbf{K} = E\left[\frac{\partial \rho_j'(\gamma_0,\hat{h})}{\partial\gamma}\Omega_j^{-1}\frac{\partial \rho_j(\gamma_0,\hat{h})}{\partial\gamma'}\right] \tag{2.16}$$

is nonsingular.

Theorem 2.2. Under Assumptions 2.1 - 2.6, $\sqrt{n}(\hat{\gamma}_{n,S}-\gamma_0) \xrightarrow{d} N(0, \mathbf{K^{-1}D}\tau_{\mathbf{S}}\mathbf{D'K^{-1}})$, as $n \longrightarrow \infty$, where matrices $\mathbf{D}$ and $\tau_{\mathbf{S}}$ are given by

$$\mathbf{D} = \left(\mathbf{I}_{p+k+1}, E\left[\frac{\partial \rho_1'(\gamma_0, h_0)}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_1(\gamma_0, h_0)}{\partial h'}\right](EZ_1Z_1' \otimes \mathbf{I}_q)^{-1}\right) \tag{2.17}$$

$$\tau_{\mathbf{S,11}} = \frac{1}{2}E\left[\frac{\partial \rho_{1,1}'(\gamma_0, h_0)}{\partial \gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0, h_0)\rho_{1,2}'(\gamma_0, h_0)\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0, h_0)}{\partial \gamma'}\right] + \frac{1}{2}\mathbf{K}$$

$$\tau_{\mathbf{S,21}} = 0 \tag{2.18}$$

$\tau_{\mathbf{S,22}} = E\left[(Z_1 \otimes (W_1 - H_0Z_1))(Z_1 \otimes (W_1 - H_0Z_1))'\right]$, and $\otimes$ stands for the Kronecker product (Magnus and Neudecker (1988), p.30).

Remark 2.2 When closed forms of the conditional moments in (2.6) and (2.7) exist, the SBE $\hat{\gamma}_{n,S}$ becomes the method of moments estimator (MME) $\hat{\gamma}$ of Abarin and Wang (2012) defined as $\hat{\gamma}_n = \arg\min Q_n(\gamma, \hat{h})$. They proved that $\hat{\gamma}_n$ is consistent and asymptotically normally distributed. In other words their MME has the following properties:

1. Under Assumptions 2.1 - 2.3, $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, as $n \longrightarrow \infty$.

2. Under Assumptions 2.1 - 2.6, $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \mathbf{K^{-1}D}\tau\mathbf{D'K^{-1}})$, as $n \longrightarrow \infty$, where

$$\tau = E\left[\frac{\partial \rho_j'(\gamma_0, h_0)}{\partial \gamma}\Omega_j^{-1}\rho_j(\gamma_0, h_0)\rho_j'(\gamma_0, h_0)\Omega_j^{-1}\frac{\partial \rho_j(\gamma_0, h_0)}{\partial \gamma'}\right]. \tag{2.19}$$

2.3 Monte Carlo simulation studies for simulation-based estimators

In this section we present simulation studies to demonstrate how the proposed Simulation-Based Estimator (SBE) for γ can be calculated in the logistic and probit regression models (2.1)-(2.3) with a binary response variable and mismeasured covariates, and to investigate the performance of these estimators for finite sample sizes.

In all the simulation studies, to compute the SBE's, we first choose the density of $l(\mathbf{x})$ such that the support of $l(\mathbf{x})$ covers the support of $p(\alpha + \beta'\mathbf{x}_{js})f_U(\mathbf{x}_{js} - \nu; \theta)$ for all $\nu \in R^{\mathbf{q}}$ and $\gamma \in \Gamma$, and generate independent points $\mathbf{x}_{js}$, $s = 1, 2, \ldots, 2S, j = 1, 2, \ldots, n$ to calculate the simulation moment $m_1(\hat{\mathbf{H}}\mathbf{Z}_j; \gamma)$ of (2.9) and $m_2(\hat{\mathbf{H}}\mathbf{Z}_j; \gamma)$ of (2.10). Secondly, we calculate $m_1(\hat{\mathbf{H}}\mathbf{Z}_j; \gamma)$ and $m_2(\hat{\mathbf{H}}\mathbf{Z}_j; \gamma)$. The two-step SBE $\hat{\gamma}_{n,s}$ is calculated based on Remark 2.1 by minimizing $Q_{n,S}(\gamma, \hat{h})$, using the 2×2 identity matrix I_2 to obtain the first-stage Simulation-Based Estimator (SBE1), and then estimate optimal weight matrix. Finally minimize $Q_{n,S}(\gamma, \hat{h})$ again with $\hat{\Omega}_j^{-1}$ to obtain the second-stage simulation based estimator (SBE2). To eliminate potential nonlinear numerical optimization errors and difficulties, the true parameter values were used as starting values for the first-stage simulation based estimator.

Along with the SBE's, naive estimator for the logistic and probit regression models, and a nonparametric correction estimator (NPE) based on Huang and Wang (2001) for the logistic models for the purpose of comparison were investigated. The NPE of Huang and Wang (2001) for the logistic model (2.1)-(2.3) is based on the following estimating score

$$\sum_{i=1}^{n} \begin{pmatrix} 1 \\ \mathbf{z}_i \end{pmatrix} [y_i + (y_i - 1)\exp(\alpha + \beta' \mathbf{w_i})] = 0.$$

In Naive estimation (Naive), we ignore both the measurement errors δ in (2.2) and U in equation (2.3), and use the least squares method to estimate (α, β'). All computations are done in R, and the naive estimators are reweighed least squares (WLS) estimators, obtained by 'glm' function. To determine how well all these estimators perform, we present their mean, median, and root mean squared errors (RMSE).

2.3.1 Example 2.1

In our first simulation Example 2.1, the data $(y_j, w_j, z_j), j = 1, \ldots, n$ is generated using a standard normal instrumental variables Z_j and the measurement errors δ_j from the normal distribution with mean of zero and standard deviation of 0.5. To simulate X_j via (2.3), we assume that H from (2.3) the identity matrix and $U \sim N(0, \theta^2)$ with $\theta = 0.8$. We generate the response variables Y_j from a binary distribution with parameter $p(\alpha + \beta X_j)$ with $\alpha = 0.5$ and $\beta = 0.3$. The density of $l(x)$ as in (2.9) and (2.10) is $N(0, 4)$ and we generate independent points $\mathbf{x}_{js}$, $s = 1, 2, \ldots, 2S, j = 1, 2, \ldots, n$ using S = 2,000 and 5,000. $N = 1,000$ Monte Carlo replications is carried out and, in each replication, $n = 400$ sample points (y_j, w_j, z_j) are generated.

Table 2.1 contains our simulation results for the logistic model with parameter $p(\alpha + \beta X_j) = (1 + \exp(-(\alpha + \beta X_j)))^{-1}$ and Table 2.2 summarizes our simulation results for the probit model with parameter $p(\alpha + \beta X_j) = \Phi(\alpha + \beta X_j)$. Mean, median and RMSE of the estimators are reported. Although the SBE's show their asymptotic properties, but they converge slower for θ. Note that in most of our simulation studies, mean and median of the estimators are appropriately close to each other, but when covariate measurement error δ becomes substantial, outliers will appear in some of the estimates.

As we can see from Table 2.1, for the regression coefficient α, SBE2 for $S = 2000$ and $S = 5000$, and Naive show much smaller bias than SBE1 for $S = 2000$ and $S = 5000$, and NPE. For the regression coefficient β, SBE2 for $S = 2000$ and $S = 5000$ show much smaller bias than the other estimators. The RMSE of the estimators show that SBE2 for $S = 5000$ performs better than the other estimators.

As we can see from Table 2.2 for the regression coefficients α and β, SBE2 shows much smaller bias than SBE1 and Naive estimator. The RMSE of the estimators of α show that SBE2 and Naive perform equally well. SBE2 shows smaller RMSE than SBE1 and Naive estimator for β.

As seen from both Tables 2.1 and 2.2, when we increase the the size of S for SBE's then RMSE decrease, as we expected. We can also see, SBE2's are less biased and have smaller RMSE compared to SBE1's. In addition, based on Remark 2.1 the second-stage estimator is asymptotically more efficient than the first-stage estimator. Therefore in the next simulations we only compare SBE2, NPE, and the naive estimator.

Table 2.1: Simulation results of logistic model in Example 2.1

		α=0.5	β=0.3	θ=0.8
		$S = 2000$		
SBE1	mean	0.5408	0.2964	0.9830
	median	0.5145	0.2969	0.7826
	RMSE	0.2406	0.1549	0.9524
SBE2	mean	0.5070	0.2994	0.8253
	median	0.5050	0.3000	0.7883
	RMSE	0.1157	0.1272	0.6432
		$S = 5000$		
SBE1	mean	0.5216	0.2979	1.0042
	median	0.5151	0.2987	0.8071
	RMSE	0.2147	0.1038	0.9874
SBE2	mean	0.5005	0.2991	0.8150
	median	0.5023	0.3002	0.8095
	RMSE	0.1007	0.0847	0.5642
NPE	mean	0.4849	0.3120	–
	median	0.4874	0.3047	–
	RMSE	0.1083	0.1235	–
Naive	mean	0.5005	0.2640	–
	median	0.5014	0.2646	–
	RMSE	0.1040	0.0877	–

Table 2.2: Simulation results of probit model in Example 2.1

		α=0.5	β=0.3	θ=0.8
			$S = 2000$	
SBE1	mean	0.5641	0.2915	0.9941
	median	0.5291	0.2937	0.8417
	RMSE	0.3716	0.1472	0.9943
SBE2	mean	0.5228	0.2952	0.9201
	median	0.5113	0.2963	0.8191
	RMSE	0.2057	0.1131	0.7575
			$S = 5000$	
SBE1	mean	0.5117	0.2967	0.8853
	median	0.5085	0.2960	0.8250
	RMSE	0.1122	0.1109	0.6223
SBE2	mean	0.5039	0.3015	0.8054
	median	0.5040	0.3005	0.8127
	RMSE	0.0765	0.0848	0.4118
Naive	mean	0.6796	0.0874	–
	median	0.6802	0.0875	–
	RMSE	0.1810	0.2132	–

2.3.2 Example 2.2

In our second simulation Example 2.2, the data $(y_j, \mathbf{w}_j, \mathbf{z}_j), j = 1, \ldots, n$ is generated from the model (2.1) - (2.3), using instrumental variables

$$\mathbf{Z} = (z_1, z_2)' \sim N\left(\mathbf{0}, diag\left(1, 0.25\right)\right)$$

and the measurement errors

$$\delta = (\delta_1, \delta_2)' \sim N(\mathbf{0}, diag(var(\delta_1), var(\delta_2))),$$

with different values of the variances which are reported in Tables 2.3 - 2.6. To simulate $\mathbf{X_j}$ via (2.3), we assume that H is a known diagonal matrix equal to $H = diag(1, 2)$, and

$$\mathbf{U} = (u_1, u_2)' \sim N\left(\mathbf{0}, diag(\theta_1, \theta_2)\right),$$

with $\theta_1 = 1$ and $\theta_2 = 1$. For the SBE, the density of $l(\mathbf{x})$ as in (2.9) and (2.10) is $N\left(\mathbf{0}, diag(4, 4)\right)$, and we generated independent points $\mathbf{x}_{js}$, $s = 1, 2, \ldots, 2S, j = 1, 2, \ldots, n$ using S = 2,000. We generate the response variables Y_j from a binary distribution with parameter $p(\alpha + \beta'\mathbf{x})$, with $\alpha = 0.4$, and $\beta = (\beta_1, \beta_2)' = (0.2, 0.3)'$. $N = 1000$ Monte Carlo replications is carried out, and for the purpose of comparison in each replication $n = 200, 400$, and 800 sample points $(y_j, \mathbf{w}_j, \mathbf{z}_j)$ are generated.

Our simulation results for the logistic model with parameter $p(\alpha + \beta'\mathbf{x_j}) = (1 + \exp(-(\alpha + \beta'\mathbf{x}_j)))^{-1}$ are summarized in Tables 2.3, 2.4, and 2.5 for three different levels of measurement errors. Everywhere, naive estimators show smaller bias than SBE2 and NPE in estimating the regression coefficient α. For estimating the regression coefficients of mismeasured covariates, β_1 and β_2, SBE2 shows smaller bias than NPE and Naive estimators. Clearly, SBE2 estimators of θ_1 and θ_2 are a little biased, but its bias decreases as the sample size increases.

As we can see from Table 2.3, Naive estimator shows smaller RMSE than SBE2 and NPE for the small measurement errors. Our SBE2 and NPE for β_1 and β_2 perform equally well based on RMSE.

Tables 2.4 and 2.5 represent the results for moderate and large measurement errors, respectively. In these cases for small sample sizes, RMSE of our SBE2 are very close to the Naive estimator, and they both are much better than NPE. But for large sample sizes, SBE2 performs much better than the other two estimators. The results also show that when the sample size increases, the RMSE of each estimator decreases, as we expected.

As we can see from Tables 2.3-2.5, Naive estimator shows smaller RMSE that SBE2 for the small sample size of $n = 200$ in all three different levels of measurement errors. Table 2.6 represent the results of SBE2 using $S = 2000$ and $S = 5000$, and Naive estimator for different levels of measurement errors. We can see from Table 2.6, even for the small sample sizes when the size of generated points from $l(\mathbf{x})$ increase from 2000 to 5000, then the RMSE of SBE2 is much smaller than Naive estimator. Therefore, for the small and moderate sample sizes to decrease the RMSE of SBE's it is better to generated large number of independent points $\mathbf{x}_{js}$ from $l(\mathbf{x})$.

Table 2.3: Simulation results of logistic model in Example 2.2, small measurement error

			var(δ_1)	=1	var(δ_2)	=1	
		n	α	β_1	β_2	θ_1	θ_2
	True		0.4	0.2	0.3	1.0	1.0
		200	0.4587	0.1920	0.2959	1.2482	1.2022
	mean	400	0.4402	0.2020	0.2930	1.1821	1.1590
		800	0.4261	0.1934	0.2920	1.1529	1.1048
		200	0.4573	0.1710	0.2864	0.9956	1.0297
SBE2	median	400	0.4450	0.1921	0.2931	0.9819	1.0295
		800	0.4272	0.1932	0.2963	1.0413	1.0110
		200	0.1937	0.1741	0.1841	1.1084	0.9919
	RMSE	400	0.1321	0.1235	0.1401	0.9466	0.8220
		800	0.0891	0.0897	0.1055	0.7988	0.6919
		200	0.2032	0.2317	0.3529	–	–
	mean	400	0.2921	0.2226	0.3177	–	–
		800	0.3172	0.2079	0.3123	–	–
		200	0.2757	0.2079	0.3064	–	–
NPE	median	400	0.3102	0.2135	0.3035	–	–
		800	0.3214	0.2048	0.3050	–	–
		200	0.5266	0.2647	0.3128	–	–
	RMSE	400	0.1989	0.1392	0.1429	–	–
		800	0.1233	0.0876	0.0966	–	–
		200	0.3997	0.1344	0.1998	–	–
	mean	400	0.3977	0.1336	0.1962	–	–
		800	0.3956	0.1319	0.1962	–	–
		200	0.4013	0.1310	0.1983	–	–
Naive	median	400	0.3990	0.1336	0.1959	–	–
		800	0.3971	0.1308	0.1970	–	–
		200	0.1513	0.1091	0.1331	–	–
	RMSE	400	0.1046	0.0901	0.1194	–	–
		800	0.0725	0.0801	0.1120	–	–

Table 2.4: Simulation results of logistic model in Example 2.2, moderate measurement error

			var(δ_1)	=2	var(δ_2)	=2	
		n	α	β_1	β_2	θ_1	θ_2
	True		0.4	0.2	0.3	1.0	1.0
SBE2	mean	200	0.4544	0.1991	0.2989	1.2651	1.2280
		400	0.4352	0.2034	0.2915	1.1745	1.1782
		800	0.4221	0.1936	0.2932	1.1488	1.0971
	median	200	0.4558	0.1811	0.2801	1.0211	1.0541
		400	0.4393	0.1954	0.2922	0.9943	1.0485
		800	0.4233	0.1930	0.2979	1.0176	1.0285
	RMSE	200	0.1873	0.1777	0.1916	1.1453	1.0412
		400	0.1246	0.1232	0.1343	0.9636	0.8851
		800	0.0872	0.0877	0.1017	0.8336	0.7165
NPE	mean	200	-0.0185	0.2480	0.3770	–	–
		400	0.1761	0.2326	0.3262	–	–
		800	0.2306	0.2110	0.3184	–	–
	median	200	0.1895	0.2056	0.3084	–	–
		400	0.2302	0.2131	0.3004	–	–
		800	0.2548	0.2055	0.3085	–	–
	RMSE	200	1.2121	0.3755	0.4182	–	–
		400	0.3579	0.1689	0.1623	–	–
		800	0.2249	0.0966	0.1119	–	–
Naive	mean	200	0.3960	0.1010	0.1498	–	–
		400	0.3940	0.1006	0.1474	–	–
		800	0.3918	0.0988	0.1467	–	–
	median	200	0.3977	0.0972	0.1440	–	–
		400	0.3949	0.1007	0.1460	–	–
		800	0.3921	0.0979	0.1470	–	–
	RMSE	200	0.1506	0.1246	0.1680	–	–
		400	0.1043	0.1125	0.1611	–	–
		800	0.0724	0.1076	0.1574	–	–

Table 2.5: Simulation results of logistic model in Example 2.2, large measurement error

			var(δ_1)	=4	var(δ_2)	=4	
		n	α	β_1	β_2	θ_1	θ_2
	True		0.4	0.2	0.3	1.0	1.0
SBE2	mean	200	0.4378	0.2039	0.3086	1.3230	1.2419
		400	0.4306	0.2000	0.2944	1.2431	1.1728
		800	0.4200	0.1970	0.2961	1.1606	1.0885
	median	200	0.4308	0.1928	0.2986	1.0360	1.0567
		400	0.4330	0.1903	0.2895	1.0533	1.0344
		800	0.4222	0.1949	0.2970	1.0166	1.0345
	RMSE	200	0.1751	0.1757	0.1841	1.2031	1.1762
		400	0.1197	0.1234	0.1291	1.0341	0.9264
		800	0.0844	0.0921	0.0974	0.8867	0.7592
NPE	mean	200	-0.3121	0.2356	0.3661	–	–
		400	-0.1067	0.2447	0.3354	–	–
		800	0.0041	0.2200	0.3334	–	–
	median	200	0.0146	0.2007	0.3003	–	–
		400	0.0585	0.2120	0.2929	–	–
		800	0.0928	0.2081	0.3076	–	–
	RMSE	200	1.5982	0.3687	0.4077	–	–
		400	0.8179	0.2070	0.2198	–	–
		800	0.5924	0.1281	0.1564	–	–
Naive	mean	200	0.3921	0.0662	0.0978	–	–
		400	0.3902	0.0660	0.0964	–	–
		800	0.3879	0.0644	0.0954	–	–
	median	200	0.3945	0.0641	0.0962	–	–
		400	0.3931	0.0665	0.0948	–	–
		800	0.3885	0.0632	0.0955	–	–
	RMSE	200	0.1502	0.1472	0.2111	–	–
		400	0.1039	0.1405	0.2078	–	–
		800	0.0725	0.1387	0.2066	–	–

Table 2.6: Compare SBE2 with Naive estimator for the small sample size $n = 200$

			α	β_1	β_2	θ_1	θ_2
		True	0.4	0.2	0.3	1.0	1.0
	mean	$S = 2000$	0.4587	0.1920	0.2959	1.2482	1.2022
		$S = 5000$	0.4022	0.2020	0.3066	1.2554	1.1142
		Naive	0.3997	0.1344	0.1998	–	–
var(δ_1)=1 var(δ_2)=1	median	$S = 2000$	0.4573	0.1710	0.2864	0.9956	1.0297
		$S = 5000$	0.4009	0.1984	0.3083	0.9886	0.9915
		Naive	0.4013	0.1310	0.1983	–	–
	RMSE	$S = 2000$	0.1937	0.1741	0.1841	1.1084	0.9919
		$S = 5000$	0.1077	0.1011	0.1303	1.0545	0.9365
		Naive	0.1513	0.1091	0.1331	–	–
	mean	$S = 2000$	0.4544	0.1991	0.2989	1.2651	1.2280
		$S = 5000$	0.4292	0.2019	0.3110	1.2884	1.0441
		Naive	0.3960	0.1010	0.1498	–	–
var(δ_1)=2 var(δ_2)=2	median	$S = 2000$	0.4558	0.1811	0.2801	1.0211	1.0541
		$S = 5000$	0.4236	0.1974	0.3101	1.0121	0.9765
		Naive	0.3977	0.0972	0.1440	–	–
	RMSE	$S = 2000$	0.1873	0.1777	0.1916	1.1453	1.0412
		$S = 5000$	0.1306	0.1240	0.1356	1.0398	0.9379
		Naive	0.1506	0.1246	0.1680	–	–
	mean	$S = 2000$	0.4378	0.2039	0.3086	1.3230	1.2419
		$S = 5000$	0.4293	0.2012	0.3138	1.3573	1.0477
		Naive	0.3921	0.0662	0.0978	–	–
var(δ_1)=4 var(δ_2)=4	median	$S = 2000$	0.4308	0.1928	0.2986	1.0360	1.0567
		$S = 5000$	0.4213	0.1925	0.3104	1.0436	0.9400
		Naive	0.3945	0.0641	0.0962	–	–
	RMSE	$S = 2000$	0.1751	0.1757	0.1841	1.2031	1.1762
		$S = 5000$	0.1314	0.1297	0.1548	1.1726	1.0234
		Naive	0.1502	0.1472	0.2111	–	–

Our simulation results for the probit model with parameter $p(\alpha + \beta'\mathbf{x_j}) = \Phi(\alpha + \beta'\mathbf{x}_j)$ are summarized in Tables 2.7, 2.8, and 2.9 for three different levels of measurement errors. Everywhere, SBE2 show much smaller bias than Naive estimators in estimating the regression coefficients α, β_1, and β_2. Clearly, SBE2 estimators of θ_1 and θ_2 are a little biased, but its bias decreases as the sample size increases.

As we can see from Tables 2.7 - 2.9, for all three different levels of measurement errors SBE2 shows much smaller RMSE than Naive estimator. But for large sample sizes, SBE2 performs much better than the other two estimators. The results also show that when the sample size increases, the RMSE of each estimator decreases, as we expected.

Table 2.7: Simulation results of probit model in Example 2.2, small measurement error

			var(δ_1)	=1	var(δ_2)	=1	
		n	α	β_1	β_2	θ_1	θ_2
	True		0.4	0.2	0.3	1.0	1.0
		200	0.4376	0.1982	0.3009	1.2012	1.1436
	mean	400	0.4210	0.2007	0.2999	1.1016	1.0919
		800	0.4091	0.1960	0.2992	1.0748	1.0180
		200	0.4305	0.1924	0.2999	1.0184	1.0389
SBE2	median	400	0.4205	0.2024	0.3025	0.9839	1.0261
		800	0.4092	0.1986	0.3026	0.9919	1.0034
		200	0.1336	0.1208	0.1373	0.9835	0.8515
	RMSE	400	0.0867	0.0884	0.1002	0.6285	0.8211
		800	0.0578	0.0656	0.0693	0.6472	0.4406
		200	0.6403	0.0450	0.0666	–	–
	mean	400	0.6398	0.0447	0.0666	–	–
		800	0.6393	0.0446	0.0662	–	–
		200	0.6415	0.0455	0.0660	–	–
Naive	median	400	0.6401	0.0449	0.0667	–	–
		800	0.6392	0.0448	0.0661	–	–
		200	0.2426	0.2341	0.1561	–	–
	RMSE	400	0.2409	0.1558	0.2337	–	–
		800	0.2398	0.1556	0.2340	–	–

Table 2.8: Simulation results of probit model in Example 2.2, moderate measurement error

			var(δ_1)	=2	var(δ_2)	=2	
		n	α	β_1	β_2	θ_1	θ_2
	True		0.4	0.2	0.3	1.0	1.0
		200	0.4365	0.1999	0.3100	1.2386	1.1056
	mean	400	0.4206	0.2007	0.3017	1.1199	1.0860
		800	0.4083	0.1964	0.2999	1.0717	0.9968
		200	0.4280	0.1929	0.2988	1.0253	1.0123
SBE2	median	400	0.4206	0.1978	0.3021	1.0045	1.0240
		800	0.4093	0.1987	0.3026	0.9959	1.0011
		200	0.1497	0.1214	0.1923	1.0608	0.8732
	RMSE	400	0.0855	0.0891	0.1009	0.6656	0.8668
		800	0.0567	0.0639	0.0670	0.6675	0.4771
		200	0.6404	0.0342	0.0504	–	–
	mean	400	0.6398	0.0339	0.0505	–	–
		800	0.6393	0.0338	0.0500	–	–
		200	0.6415	0.0346	0.0502	–	–
Naive	median	400	0.6399	0.0342	0.0505	–	–
		800	0.6393	0.0337	0.0498	–	–
		200	0.2427	0.1666	0.2501	–	–
	RMSE	400	0.2409	0.1665	0.2497	–	–
		800	0.2398	0.1664	0.2501	–	–

Table 2.9: Simulation results of probit model in Example 2.2, large measurement error

			var(δ_1)	=4	var(δ_2)	=4	
		n	α	β_1	β_2	θ_1	θ_2
	True		0.4	0.2	0.3	1.0	1.0
		200	0.4369	0.2003	0.3078	1.3094	1.1066
	mean	400	0.4182	0.2021	0.3054	1.1326	1.0740
		800	0.4079	0.1989	0.3026	1.0591	0.9810
		200	0.4289	0.1938	0.3011	1.0754	1.0043
SBE2	median	400	0.4154	0.1990	0.3005	1.0340	1.0378
		800	0.4075	0.1988	0.3018	1.0081	0.9933
		200	0.1498	0.1548	0.1545	1.1572	0.9309
	RMSE	400	0.0835	0.0848	0.0975	0.8812	0.6930
		800	0.0566	0.0623	0.0693	0.7115	0.5338
		200	0.6404	0.0226	0.0331	–	–
	mean	400	0.6398	0.0224	0.0334	–	–
		800	0.6393	0.0222	0.0329	–	–
		200	0.6414	0.0231	0.0331	–	–
Naive	median	400	0.6398	0.0224	0.0331	–	–
		800	0.6396	0.0223	0.0327	–	–
		200	0.2428	0.1778	0.2672	–	–
	RMSE	400	0.2410	0.1778	0.2667	–	–
		800	0.2399	0.1779	0.2672	–	–

2.3.3 Example 2.3

In this simulation Example 2.3, we consider the regression models with binary response variable given by the following equations:

$$
\begin{aligned}
&P(Y_i = 1|X_i = x_i, V_i = v_i) \\
&\quad = p\left(x_i\left(\beta_1 + \beta_2 v_{1i} + \beta_3 v_{2i} + \beta_4 v_{3i}\right) + \alpha_1 + \alpha_2 v_{1i} + \alpha_3 v_{2i} + \alpha_4 v_{3i}\right),
\end{aligned} \tag{2.20}
$$

and

$$
W_i = X_i + \delta \tag{2.21}
$$

$$
X_i = h_1 + h_2 z_i + U_i. \tag{2.22}
$$

The observable covariates v_{1i}, v_{2i} and v_{3i} in (2.20) are all dummy variables, where $v_{1i} = v_{2i} = v_{3i} = 0$ indicates that the ith individual receives treatment 1 , and $v_{1i} = 1$, $v_{2i} = 1$ and $v_{3i} = 1$ mean that the ith individual receives treatment 2, 3 or 4 respectively. For the ith individual, at most one of v_{1i}, v_{2i}, v_{3i} is 1 and the chances of receiving each of four treatments are equal.

The data is generated using Z from a standard normal distribution, U from a normal distribution with mean of zero and standard deviation of $\theta = 0.2$, and δ from a normal distribution with mean of zero and standard deviation of 0.4. The parameter values for this model are $h_1 = 1$, $h_2 = 1$, $\beta_1 = 0.3$, $\beta_2 = -0.5$, $\beta_3 = 0.6$, $\beta_4 = -0.4$, $\alpha_1 = -0.5$, $\alpha_2 = 1$, $\alpha_3 = -1$, $\alpha_4 = 0.5$, and $\theta = 0.2$. $N = 1,000$ Monte Carlo replications is carried out and, in each replication, $n = 500$ sample points (y_j, w_j, z_j) is generated. The least squares estimator of (2.5) result to consistently estimate h_1 and h_2. For the SBE2 we choose the density of $l(x)$ to be $N(1, 2.25)$ and we generated independent points $\mathbf{x}_{js}$, $s = 1, 2, \ldots, 2S, j = 1, 2, \ldots, n$ using S = 2,000. We generated the y_i observations from the model

Our simulation results for the logistic model with

$$P(Y_i = 1|X_i = x_i, V_i = v_i) =$$
$$[1 + \exp[-x_i(\beta_1 + \beta_2 v_{1i} + \beta_3 v_{2i} + \beta_4 v_{3i}) - \alpha_1 - \alpha_2 v_{1i} - \alpha_3 v_{2i} - \alpha_4 v_{3i}]]^{-1},$$

are summarized in Table 2.10, and results for the probit model with

$$P(Y_i = 1|X_i = x_i, V_i = v_i) =$$
$$\Phi(x_i(\beta_1 + \beta_2 v_{1i} + \beta_3 v_{2i} + \beta_4 v_{3i}) + \alpha_1 + \alpha_2 v_{1i} + \alpha_3 v_{2i} + \alpha_4 v_{3i}),$$

are summarized in Table 2.11.

As seen from Tables 2.10 and 2.11, all estimators are seen to be a little biased. Note that this model has 11 unknown parameters, which is a relatively large number, but the inference performance of our SBE2 is still satisfactory

As seen from Table 2.10, Naive estimator shows the largest bias and RMSE for all the parameters in the logistic model. For the coefficientsα_1, α_4,β_2, β_4, β_1, SBE2 shows smaller bias than NPE. SBE2 shows larger bias than NPE for the regression coefficients β_3, α_2, and α_3. NPE reports slightly smaller RMSE than our proposed SBE2 for all the parameters.

As seen from Table 2.11, Naive estimator shows the larger bias and RMSE compared to our SBE2 for all the parameters in the probit model.

Table 2.10: Simulation results of logistic model in Example 2.3

		h_1	h_2	β_1	β_2	β_3	β_4	α_1	α_2	α_3	α_4	θ
	True	1	1	0.3	-0.5	0.6	-0.4	-0.5	1.0	-1.0	0.5	0.2
SBE2	mean	1.0001	0.9996	0.3105	-0.5174	0.6267	-0.4055	-0.5069	1.0178	-1.0711	0.4894	0.1924
	median	1.0000	0.9991	0.3063	-0.5016	0.6027	-0.4045	-0.5134	0.9963	-1.0330	0.4946	0.1952
	RMSE	0.0094	0.0219	0.1091	0.1905	0.1885	0.1557	0.1531	0.2277	0.2567	0.2165	0.0804
NPE	mean	–	–	0.2816	-0.4745	0.5976	-0.3718	-0.5362	1.0058	-0.9836	0.5264	–
	median	–	–	0.2907	-0.4773	0.6084	-0.3725	-0.5240	1.0164	-0.9733	0.5309	–
	RMSE	–	–	0.1077	0.1714	0.1047	0.1711	0.1356	0.1794	0.0928	0.1705	–
Naive	mean	–	–	0.2667	-0.4465	0.5231	-0.3458	-0.4612	0.9464	-0.9324	0.4345	–
	median	–	–	0.2668	-0.4382	0.5169	-0.3390	-0.4518	0.9324	-0.9070	0.4242	–
	RMSE	–	–	0.1734	0.2421	0.2875	0.2464	0.2498	0.3724	0.4260	0.3503	–

Table 2.11: Simulation results of probit model in Example 2.3

		h_1	h_2	β_1	β_2	β_3	β_4	α_1	α_2	α_3	α_4	θ
	True	1	1	0.3	-0.5	0.6	-0.4	-0.5	1.0	-1.0	0.5	0.2
SBE2	mean	1.0001	0.9993	0.3149	-0.5293	0.6149	-0.4118	-0.5191	1.0367	-1.0518	0.4994	0.1683
	median	1.0003	0.9994	0.3072	-0.5330	0.5925	-0.4101	-0.5229	1.0267	-0.9994	0.4919	0.1395
	RMSE	0.0096	0.0218	0.1382	0.1932	0.2887	0.1975	0.1979	0.2972	0.4081	0.2755	0.1282
Naive	mean	–	–	0.4356	-0.7265	0.8497	-0.5697	-0.7575	1.5361	-1.5137	0.7225	–
	median	–	–	0.4250	-0.7246	0.8284	-0.5578	-0.7426	1.5044	-1.5176	0.7038	–
	RMSE	–	–	0.2291	0.3361	0.4180	0.3072	0.3765	0.6635	0.7221	0.4359	–

2.4 Real data example

In this section, we analyze the performance of our SBE2 for γ on the data set from the AIDS Clinical Trial Group (ACTG175) study. This double blind study evaluated four treatments which are assigned randomly to HIV infected adults whose screening CD4 cell counts were between 200 and 500 per cubic millimetre. Treatment 1 is zidovudin alone (ZDV), which is considered as the reference treatment. Treatment 2, 3, and 4 are zidovudin plus didanosine (ZDV+ddI), zidovudin plus zalcitabine (ZDV+ddc), and didanosine (ddI) alone, respectively. See Hammer et al. (1996), Huang and Wang (2000), and Huang and Wang (2001) for more detailed descriptions of the data set. In our model we considered 1036 patients who had never taken any antiretroviral therapy from a total of 2467 HIV infected volunteers participated in this study. We are interested in comparing the effect of these treatments in terms of whether a patient has ≥ 50 percent decline in the CD4 cell count, development of AIDS, or death from HIV related disease (Xu et al. (2015)).We use model (2.20), where v_{mi}, $m = 1, 2, 3$, has the same meaning as in the previous simulation study. In this model, X is log(CD4 cell count) at baseline and within 3 weeks of randomization. Since X could not be measured precisely, we use the average of these two available variables as an observed variable, W. Furthermore, for the instrumental variable Z, we used a screening log(CD4 count). Based on Figure 2.1, the relation between W and X can be described by a linear model. Thus, (2.21) and (2.22) can describe the relation between W and X, and Z and X.

The consistent estimate for h_1 and h_2 by the least squares estimator are 0.2278 and 0.9122, respectively. For the SBE2 we choose the density of $l(x)$ to be $Beta(6, 2)$.

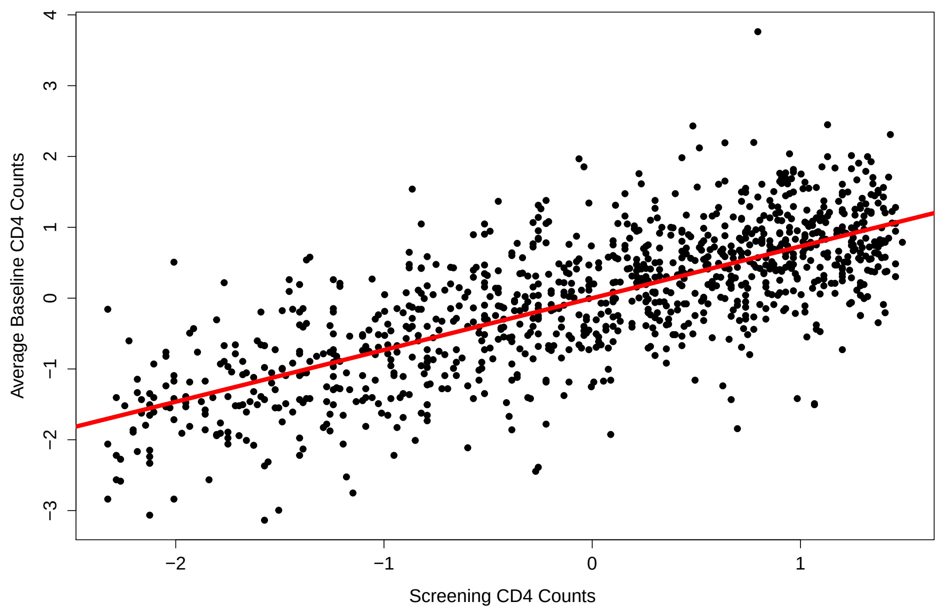

Figure 2.1: Plot of the covariate averaged baseline CD4 count versus the instrumental variable CD4 count.

We also estimated empirical standard errors of regression coefficients, via bootstrap method using 100 independent bootstrap samples each consisting of 200 data values drawing with replacement from 1036 patients who had never taken any antiretroviral therapy.

The consecutive estimates and their empirical standard errors (Emp SE) for regression coefficients in (2.20) and θ are reported in Table 2.12. As we can see from Table 2.12, probit-SBE2 model provides smaller empirical standard errors compare to the logistic-SBE2 and Naive estimator, for the estimated coefficients of CD4 cell count of all 4 treatments in the model. We can also see that the estimated coefficients of CD4 cell count for all 4 treatments are significantly different from zero. Therefore, for this data set, the probit model would fit better and provides more reliable results compare to the logistic model.

Table 2.12: SBE2 estimates and the empirical standard error for the ACTG 175 data

	β_1	β_2	β_3	β_4	
logistic-SBE2	-2.3022	0.1728	0.2637	0.0510	
Emp SE	0.1453	0.0381	0.1472	0.0138	
probit-SBE2	-1.3508	0.1181	-0.1666	0.1177	
Emp SE	0.0908	0.0274	0.0044	0.0124	
Naive	-2.7430	1.1382	-2.6873	-1.0563	
Emp SE	0.2305	0.4305	0.4725	0.3574	
	α_1	α_2	α_3	α_4	θ
logistic-SBE2	-2.2238	-0.5975	-0.4460	-0.3109	0.0442
Emp SE	0.1407	0.0315	0.1154	0.0360	0.0017
probit-SBE2	-2.0242	-0.8643	-1.1350	-0.5357	0.4463
Emp SE	0.1253	0.0546	0.0100	0.0301	0.0085
Naive	5.6344	-3.5799	5.8057	2.3578	–
Emp SE	0.5909	1.0962	1.1693	0.9042	–

The corresponding relations between the baseline log CD4 counts and the estimated linear function of X for the logistic and porbit regression models under the four treatments plotted in Figure 2.2 and 2.3, respectively.

As we can see from the plotted lines in Figures 2.2 and 2.3, treatment 1 has a negative slope. This means that the reference treatment is more adequate for patients with larger baseline CD4 counts, whose condition is less severe. Since in both Figures 2.2 and 2.3 treatments 2 and 4 show positive slopes, they seems to be more adequate for patients with smaller baseline CD4 counts.

Furthermore both Figures 2.2 and 2.3 show that the lines for treatment 1 and the other three treatments crossed around $x = -0.7$. This point of x is related to the baseline CD4 level of 288. Because the probability of having more than 50% drop of CD4 count is pretty small for treatment 1 compared with other treatments, then the

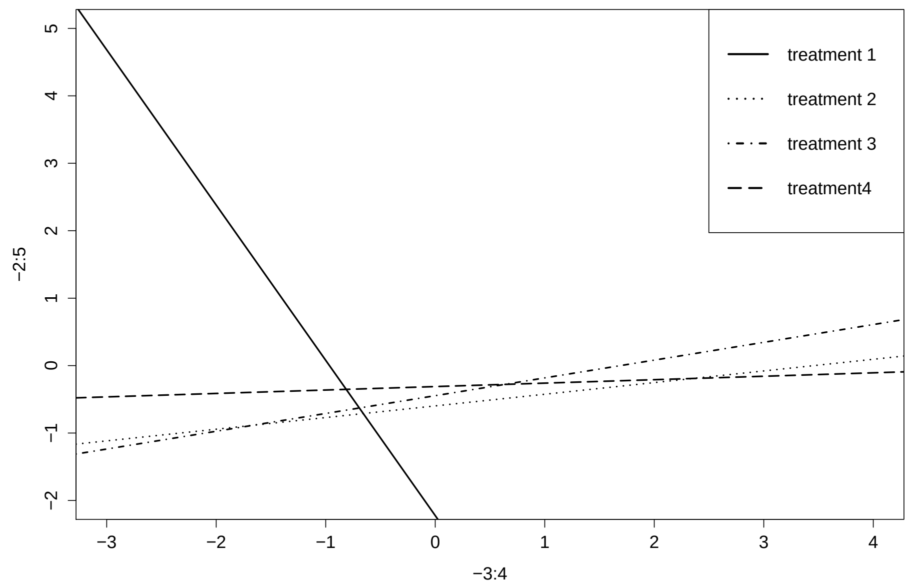

Figure 2.2: Plot of the linear function of x inside the logit link in four treatments, where x is the baseline CD4 counts in the logarithm scale.

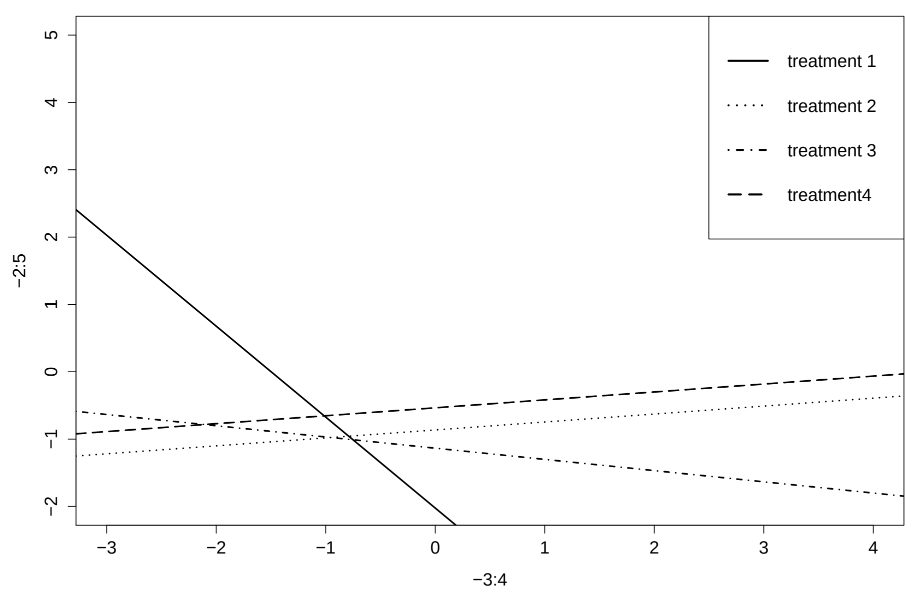

Figure 2.3: Plot of the linear function of x inside the probit link in four treatments, where x is the baseline CD4 counts in the logarithm scale.

standard treatment is better than new treatments for patients with a baseline CD4 counts larger than 288. On the contrary, we may use new treatments for patients with baseline CD4 counts less than 288.

2.5 Conclusions and Discussion

Although, estimation of logistic and probit models with mismeasured covariates and non-normal random errors have being considered before but most of the authors rely on restrictive conditions to achieve consistent estimation. Furthermore, most of the methods in the literature are limited to the case where either validation or replicate data are available, which are restrictive in many real data experiments.

In this chapter, we extended SLSE approach of Wang (2003) further to regression models with binary response variable and mismeasured covariates. Since the closed forms of conditional moments are not available for these models, we proposed Simulation- Based Estimators (SBE) for consistent estimation of the parameters of the model. We used the similar method of Abarin and Wang (2012) to propose and study SBE's for a regression model with binary response variable and classical measurement error in covariates, via using instrumental variables. This method does not require parametric assumptions for the distributions of the unobserved covariates and error components, and can be considered as a generalization of Buzas and Stefanski (1996) with much more flexible assumptions.

We proved consistency and asymptotic normality of the proposed SBE's under some regularity conditions. Furthermore, we used simulations to illustrate the performance of the SBE's in finite sample situation.Along with the SBE's, naive estimator for the logistic and probit regression models, and a nonparametric correction estimator (NPE) based on Huang and Wang (2001) for the logistic models for the purpose of comparison were investigated. Finally, the real data set from the AIDS Clinical Trial Group (ACTG175) study is also applied in this regard.

Chapter 3

Regression models with ordinal response variable and mismeasured covariates

Assume that the ordinal response Y has an underlying continuous response Y*, as in Anderson and Phillips (1981). Suppose that $-\infty = \alpha_0^* < \alpha_1^* < \ldots < \alpha_J^* = \infty$ are cut points of the continuous scale, such that

$$Y = j \qquad if \qquad \alpha_{j-1}^* < Y^* \leq \alpha_j^*, \tag{3.1}$$

for $j = 1, 2, \ldots, J$. Consider the following model,

$$Y^* = \beta_0 + \beta^{*'}\mathbf{X} + \epsilon. \tag{3.2}$$

where $\beta_0 \in \mathbf{R}$, $\beta^* \in \mathbf{R}^p$, and the distribution function of ϵ has mean zero and variance σ^2. The cumulative probability of an observation falling in category j or below for $j = 1, 2, \ldots, J$ is:

$$\begin{aligned}\pi_{ij} = P(Y_i \leq j|\mathbf{x_i}) &= P\left(Y_i^* \leq \alpha_j^*|\mathbf{x_i}\right) \\ &= P\left(Z_i \leq \frac{\alpha_j^* - \beta_0 - \beta^{*'}\mathbf{x_i}}{\sigma}\right) \\ &= G\left(\frac{\alpha_j^* - \beta_0 - \beta^{*'}\mathbf{x_i}}{\sigma}\right) \qquad (3.3)\end{aligned}$$

where $Z_i = \frac{Y_i^* - \beta_0 - \beta^{*'}\mathbf{x_i}}{\sigma}$, and G is a standard form of the Cumulative distribution function (CDF) of ϵ with mean and variance equal to 0 and 1, respectively.

Since the absolute location and scale parameters of the latent variable Y^*, β_0 and σ respectively, are not identifiable from ordinal observations, an identifiable model is

$$\pi_{ij} = P(Y_i \leq j|\mathbf{x_i}) = G(\alpha_j - \beta'\mathbf{x_i}), \qquad j = 1, \ldots, J, \qquad (3.4)$$

that depend on some covariate $\mathbf{x_i} \in \mathbf{R}^p$, where $\beta \in \mathbf{R}^p$ and $\alpha_j \in \mathbf{R}$, $j = 1, \ldots, J-1$ are unknown with the following identifiable parameter functions:

$$\alpha_j = \frac{\alpha_j^* - \beta_0}{\sigma} \quad and \quad \beta = \frac{\beta^*}{\sigma} \qquad (3.5)$$

In this chapter we study the cumulative logit model $\pi_{ij} = [1 + \exp(-(\alpha_j - \beta'\mathbf{x}_i))]^{-1}$ with an ordinal response variable that was primarily proposed by Walker and Duncan (1967) and later called the proportional odds by McCullagh (1980), and the probit link function $\pi_{ij} = \Phi(\alpha_j - \beta'\mathbf{x}_i)$.

Note that the definition of the cut points and regression parameters that we use here are identical to restrictions on β_0 and σ, usually $\beta_0 = 0$ and $\sigma = 1$, in the literatures to get around the identifiability problem. If observations really arise from a continuous latent variable, β_0 and σ are real unknown parameters and it makes little sense to restrict them to take certain values.

Furthermore, in (3.4) the cut point parameters α_j, are usually nuisance parameters of little interest and the relation between $\mathbf{X}_i$ and Y_i is independent of j. In other words, the regression coefficient vector, β, does not depend on j. This assumption of identical log-odds ratios across the J-cut points referred to as a proportional odds model, McCullagh (1980). Specifically, the model implies that odds ratios for describing effects of explanatory variables on the response variable are the same for each of the possible ways of collapsing a J-category response to a binary variable. For the proportional odds model the regression coefficients β will remain unchanged when the categories of the response variable Y are deleted or collapsed, although the intercept parameters α will be affected. The collapsibility property of the proportional odds model let us to model an ordinal outcome Y which may be continuous. Furthermore, the proportional odds model results in a reversal of the sign of the regression parameters when the codes for the response Y are reversed (i.e. $Y = 1$ recoded as $Y = J$, $Y = 2$ recoded as $Y = J - 1$, and so on). More detailed review of these properties can be found in Greenland (1994) and Agresti (2003).

There are two main approaches for analyzing ordinal observed variables with latent variables, (Monari et al. (2009)). First approach which is the most popular one is the Underlying Variable Approach (UVA). The UVA models assume that the continuous variables that generated the observed variables are normally distributed.

This approach is used in structural modelling. An overview of those type of models can be found in Muthen (1984) and Joreskog (1990). Second approach is the Item Response Theory (IRT) explained by, van der Linden and Hambleton (1997) and Bartholomew and Knott (1999). In the IRT models, the observed variables are considered as they are. The unit of analysis is the entire response pattern of a subject, so no loss of information occurs. Generalized Linear Latent Variable Model (GLLVM) for fitting models with different types of observed variables are discussed in Moustaki and Knott (2000) and Moustaki (2000).

Many studies showed that the latter approach is desirable in terms of certainty of estimates and model fit e.g., Joreskog and Moustaki (2001), Huber et al. (2004), and Cagnone et al. (2004). This is due to the fact that the IRT model is an approach based on full information whereas UVA approach is based on limited information estimation methods. However, the IRT models are much more computationally complicated than UVA models, especially when the number of latent variables increases. Huber et al. (2004) and Schilling and Bock (2005) proposed solutions to computational problems for IRT models.

Some of the early literature with such models used weighted least squares for model fitting (e.g. Williams and Grizzle (1972)). Walker and Duncan (1967) and McCullagh (1980) used Fisher scoring algorithms to propose a maximum likelihood(ML) estimation. With the ML approach, one can base significance tests and confidence intervals for the model parameters β on likelihood-ratio, score, or Wald statistics.

Real data analyses of regression models often involve covariates that are not observed directly or are measured with error. In particular, we suppose that $\mathbf{X}$ of

(3.2) is unobserved and instead we observe

$$\mathbf{W} = \mathbf{X} + \delta, \tag{3.6}$$

where δ is a measurement error that has a zero mean and a finite covariance matrix. Some of the components of δ may have zero variances so that the corresponding covariates (component of $\mathbf{X}$) are precisely measured.

Following the literature we assume that an instrumental variable $\mathbf{Z} \in \mathbf{R}^q$, $q \geq p$ is available that is related with $\mathbf{X}$ through

$$\mathbf{X} = \mathbf{H}\mathbf{Z} + \mathbf{U}, \tag{3.7}$$

where $\mathbf{H}$ is a $p \times q$ matrix of unknown parameters with rank p, $\mathbf{U}$ is a random vector independent of $\mathbf{Z}$ with $E(\mathbf{U}) = \mathbf{0}$ and a probability density function (pdf) $f_{\mathbf{U}}(\mathbf{u}; \theta)$ with an unknown vector of parameters $\theta \in \mathbf{\Theta} \subset \mathbf{R}^k$. Further, we assume that $E(\delta|\mathbf{X}, \mathbf{Z}) = \mathbf{0}$, and the conditional moments satisfy $E(Y^r|\mathbf{X}, \mathbf{W}, \mathbf{Z}) = E(Y^r|\mathbf{X}), \quad r = 1, 2$.

Note that there is no assumption concerning the functional forms of the distributions of $\mathbf{X}$, δ and $\mathbf{U}$, and in this sense our model (3.1)-(3.7) is semiparametric. Our interest is to estimate the vector of unknown parameters $\gamma = (\alpha_1, \ldots, \alpha_{J-1}, \beta', \theta')' \in \mathbf{\Gamma}$, where the parameter space $\mathbf{\Gamma} = \mathbf{A} \times \mathbf{B} \times \mathbf{\Theta}$ is compact in $\mathbf{R}^{p+k+J-1}$. We denote the true parameter value by $\gamma_0 \in \mathbf{\Gamma}$.

In this chapter, our focus is on methods of moments estimation that uses instrumental variables for the cumulative logit or probit model (3.1)-(3.7) with an ordinal response variable and classical measurement error in covariates. Since the closed forms of conditional moments are not available for these models, we propose the SBE for consistent estimation of the coefficients α_j and β in (3.4) and the parameter θ in the pdf $f_{\mathbf{U}}(\mathbf{u}, \theta)$ of $\mathbf{U}$ as in (3.7).

In section 3.1, we construct SBE for the vector of the parameters in (3.1)-(3.7). Consistency and asymptotic normality of our SBE under some regularity assumptions are proved in Section 3.2. In Section 3.3 , we illustrate the performance of the SBE in finite sample situations. Furthermore, a real data application for the data set from the Adult Literacy and Life Skills Survey (ALL) is given in Section 3.4.

3.1 Construction of simulation-based estimators

In this section, we construct consistent estimators for the vector $\gamma = (\alpha_1, \ldots, \alpha_{J-1}, \beta', \theta')'$ of the parameters in (3.1)-(3.7) based on a random sample $(Y_i, \mathbf{W}_i, \mathbf{Z}_i), i = 1, 2, ..., n$. First, from combining (3.6) and (3.7) and using our model assumptions,

$$\begin{aligned} E(\mathbf{W}|\mathbf{Z}) &= E(\mathbf{X} + \delta|\mathbf{Z}) \\ &= E(\mathbf{HZ} + \mathbf{U} + \delta|\mathbf{Z}) \\ &= E(\mathbf{HZ}|\mathbf{Z}) + \mathbf{E}(\mathbf{U}|\mathbf{Z}) + \mathbf{E}(\delta|\mathbf{Z}) \\ &= \mathbf{HZ}. \end{aligned} \tag{3.8}$$

It follows that an unknown matrix $\mathbf{H}$ can be consistently estimated by the least squares estimator

$$\hat{\mathbf{H}} = \left(\sum_{i=1}^{n} \mathbf{W}_i \mathbf{Z}_i'\right) \left(\sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i'\right)^{-1}. \tag{3.9}$$

Further, by the model assumptions and the law of iterative expectation we have

$$
\begin{aligned}
E(Y|\mathbf{Z}) &= E\left[E\left(Y|\mathbf{X},\mathbf{W},\mathbf{Z}\right)|\mathbf{Z}\right] \\
&= E\left[E\left(Y|\mathbf{X}\right)|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} jP(Y=j|\mathbf{X}=\mathbf{x})|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} j\left[P(Y\leq j|\mathbf{X}=\mathbf{x})-P(Y\leq j-1|\mathbf{X}=\mathbf{x})\right]|\mathbf{Z}\right]
\end{aligned}
$$

Therefore by (3.4) ,

$$
\begin{aligned}
E(Y|\mathbf{Z}) &= E\left[\sum_{j=1}^{J} j\left[G\left(\alpha_j-\beta'\mathbf{x}\right)-G\left(\alpha_{j-1}-\beta'\mathbf{x}\right)\right]|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} j\left[G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)-G\left(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u})\right)\right]|\mathbf{Z}\right] \\
&= \int\left[\sum_{j=1}^{J} j\left[G\left(\alpha_j-\beta'\left(\mathbf{HZ}+\mathbf{u}\right)\right)-G\left(\alpha_{j-1}-\beta'\left(\mathbf{HZ}+\mathbf{u}\right)\right)\right]\right] f_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du} \\
&= \int\left[-\sum_{j=1}^{J-1} G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)+J\right] f_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du} \qquad (3.10)
\end{aligned}
$$

and, similarly,

$$\begin{aligned}
E(Y^2|\mathbf{Z}) &= E\left[E\left(Y^2|\mathbf{X},\mathbf{W},\mathbf{Z}\right)|\mathbf{Z}\right] \\
&= E\left[E\left(Y^2|\mathbf{X}\right)|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} j^2 P(Y=j|\mathbf{X}=\mathbf{x})|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} j^2\left[P(Y\le j|\mathbf{X}=\mathbf{x})-P(Y\le j-1|\mathbf{X}=\mathbf{x})\right]|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} j^2\left[G\left(\alpha_j-\beta'\mathbf{x}\right)-G\left(\alpha_{j-1}-\beta'\mathbf{x}\right)\right]|\mathbf{Z}\right] \\
&= E\left[\sum_{j=1}^{J} j^2\left[G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)-G(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u}))\right]|\mathbf{Z}\right] \\
&= \int\left[\sum_{j=1}^{J} j^2\left[G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)-G(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u}))\right]\right] f_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du} \\
&= \int\left[-\sum_{j=1}^{J-1}(2j+1)G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)+J^2\right] f_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du}. \qquad (3.11)
\end{aligned}$$

Furthermore,

$$
\begin{aligned}
& E(Y\mathbf{W}|\mathbf{Z}) \\
= \;& E\left[\mathbf{W}E\left(Y|\mathbf{X},\mathbf{W},\mathbf{Z}\right)|\mathbf{Z}\right] \\
= \;& E\left[\mathbf{W}E\left(Y|\mathbf{X}\right)|\mathbf{Z}\right] \\
= \;& E\left[(\mathbf{X}+\delta)\sum_{j=1}^{J} jP(Y=j|\mathbf{X}=\mathbf{x})|\mathbf{Z}\right] \\
= \;& E\left[(\mathbf{HZ}+\mathbf{u})\left[\sum_{j=1}^{J} j\left[G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)-G(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u}))\right]\right]|\mathbf{Z}\right] \\
+ \;& E\left[\left[\sum_{j=1}^{J} j\left[G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)-G(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u}))\right]\right]E(\delta|\mathbf{X},\mathbf{Z})|\mathbf{Z}\right] \\
= \;& E\left[(\mathbf{HZ}+\mathbf{u})\left[\sum_{j=1}^{J} j\left[G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)-G(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u}))\right]\right]|\mathbf{Z}\right] \\
= \;& \int(\mathbf{HZ}+\mathbf{u})\left[\sum_{j=1}^{J} j\left[G(\alpha_j-\beta'\left(\mathbf{HZ}+\mathbf{u}\right))-G(\alpha_{j-1}-\beta'(\mathbf{HZ}+\mathbf{u}))\right]\right]f_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du}. \\
= \;& \int(\mathbf{HZ}+\mathbf{u})\left[-\sum_{j=1}^{J-1} G\left(\alpha_j-\beta'(\mathbf{HZ}+\mathbf{u})\right)+J\right]f_{\mathbf{U}}(\mathbf{u},\theta)\mathbf{du}. \qquad (3.12)
\end{aligned}
$$

All the integrals in this chapter are taken over the space $\mathbf{R}^p$. Equations (3.10), (3.11), and (3.12) can be used to derive method of moments estimators for $\gamma=(\alpha_1,\ldots,\alpha_{J-1},\beta',\theta')'$ if the integrals therein have closed forms. However, the latter are difficult to find in general. In the following, we propose an estimator for γ based on a simulation approximation of the multiple integrals in (3.10), (3.11), and

(3.12).

To simplify notations, we denote $\mathbf{T} = \begin{pmatrix} \mathbf{1} & Y & \mathbf{W}' \end{pmatrix}$ and $\tilde{\mathbf{x}}' = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{x}' \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \end{pmatrix}$ and write (3.10), (3.11), and (3.12) together as

$$m(\mathbf{HZ}, \gamma) = E(Y\mathbf{T}|\mathbf{Z}) = \int \tilde{\mathbf{x}}\mathbf{r}(\beta'\mathbf{x})\mathbf{f_U}(\mathbf{x} - \mathbf{HZ}; \theta)\mathbf{dx}, \tag{3.13}$$

where

$$\mathbf{r}(\beta'\mathbf{x}) = (r_1(\beta'\mathbf{x}), r_2(\beta'\mathbf{x}))', \tag{3.14}$$

with

$$r_1(\beta'\mathbf{x}) = E(Y|\mathbf{X}) = -\sum_{j=1}^{J-1} G(\alpha_j - \beta'\mathbf{x}) + J$$

and

$$r_2(\beta'\mathbf{x}) = E(Y^2|\mathbf{X}) = -\sum_{j=1}^{J-1}(2j+1)G(\alpha_j - \beta'\mathbf{x}) + J^2.$$

Then we approximate $m(\mathbf{HZ}, \gamma)$ as follows. We start with choosing a known density $l(\mathbf{x})$ and generate a random sample $\{\mathbf{x}_{js}, s = 1, 2, \ldots, 2S, j = 1, 2, \ldots, n\}$ from $l(\mathbf{x})$. Then we approximate $m(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma)$ by the following two Monte-Carlo simulators:

$$m_1(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma) = \frac{1}{S}\sum_{s=1}^{S} \tilde{\mathbf{x}}_{is}\frac{r(\beta'\mathbf{x}_{is})f_U(\mathbf{x}_{is} - \hat{\mathbf{H}}\mathbf{Z}_i; \theta)}{l(\mathbf{x}_{is})} \tag{3.15}$$

and

$$m_2(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma) = \frac{1}{S}\sum_{s=s+1}^{2S} \tilde{\mathbf{x}}_{is}\frac{r(\beta'\mathbf{x}_{is})f_U(\mathbf{x}_{is} - \hat{\mathbf{H}}\mathbf{Z}_i; \theta)}{l(\mathbf{x}_{is})}, \tag{3.16}$$

where $\tilde{\mathbf{x}}'_{is} = \begin{pmatrix} 1 & 0 & \mathbf{x}'_{is} \\ 0 & 1 & 0 \end{pmatrix}$. Finally, we define SBE $\hat{\gamma}_{n,S}$ for γ as follows:

$$\hat{\gamma}_{n,S} = \arg\min_{\gamma\in\Gamma} Q_{n,S}(\gamma, \hat{h}) = \arg\min_{\gamma\in\Gamma} \sum_{i=1}^{n} \rho'_{i,1}(\gamma, \hat{h}) {\Omega_i}^{-1} \rho_{i,2}(\gamma, \hat{h}), \tag{3.17}$$

where

$$\hat{h} = vec\hat{\mathbf{H}}, \tag{3.18}$$

$$\begin{aligned} \rho_{i,1}(\gamma, \hat{h}) &= Y_i \mathbf{T}_i - m_1(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma), \\ \rho_{i,2}(\gamma, \hat{h}) &= Y_i \mathbf{T}_i - m_2(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma), \end{aligned} \tag{3.19}$$

and

$$\Omega_i = E(\rho_{i,1}(\gamma_0, h_0)\rho'_{i,2}(\gamma_0, h_0)|\mathbf{Z}_i). \tag{3.20}$$

where in (3.20), Ω_i is so-called optimal weight matrix (Abarin and Wang (2006)) and h_0 is the true value of h.

Note that

$$E[m_1(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma)|\mathfrak{W}, \mathfrak{Z}] = E[m_2(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma)|\mathfrak{W}, \mathfrak{Z}] = m(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma),$$

where $\mathfrak{W} = \{\mathbf{W_1}, \mathbf{W_2}, \ldots, \mathbf{W_n}\}$ and $\mathfrak{Z} = \{\mathbf{Z_1}, \mathbf{Z_2}, \ldots, \mathbf{Z_n}\}$. Therefore $m_1(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma)$ and $m_2(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma)$ are unbiased simulators for $m(\hat{\mathbf{H}}\mathbf{Z}_i, \gamma)$. Using the two different sets of independent simulated points $\{\mathbf{x}_{is}, s = 1, 2, \ldots, S, i = 1, 2, \ldots, n\}$ and $\{\mathbf{x}_{is}, s =$

$S+1,\ldots,2S, i = 1,2,\ldots,n\}$ from $l(\mathbf{x})$ leads to $Q_{n,S}(\gamma,\hat{h})$ being an unbiased estimator of

$$Q_n(\gamma,\hat{h}) = \sum_{i=1}^{n} \rho_i'(\gamma,\hat{h}){\Omega_i}^{-1}\rho_i(\gamma,\hat{h}), \tag{3.21}$$

where

$$\rho_i(\gamma,\hat{h}) = Y_i\mathbf{T}_i - m(\hat{\mathbf{H}}\mathbf{Z}_i,\gamma).$$

Indeed, since $\rho_{i,1}(\gamma,\hat{h})$ and $\rho_{i,2}(\gamma,\hat{h})$ are conditionally independent given $(Y_i,\mathfrak{W},\mathfrak{Z})$, we have

$$\begin{aligned} E & \left[\rho_{i,1}'(\gamma,\hat{h}){\Omega_i}^{-1}\rho_{i,2}(\gamma,\hat{h})\right] \\ = & \ E\left[E\left[\rho_{i,1}'(\gamma,\hat{h})|Y_i,\mathfrak{W},\mathfrak{Z}\right]{\Omega_i}^{-1}E\left[\rho_{i,2}(\gamma,\hat{h})|Y_i,\mathfrak{W},\mathfrak{Z}\right]\right] \\ = & \ E\left[\rho_i'(\gamma,\hat{h}){\Omega_i}^{-1}\rho_i(\gamma,\hat{h})\right]. \end{aligned}$$

Remark 3.1. Note that in practice, Ω_i of (3.20) is a function of unknown parameters and therefore needs to be estimated. This can be done using the following two-stage procedure. First, minimize $Q_{n,S}(\gamma,\hat{h})$ as in (3.17) with the 3×3 identity matrix I_3 to obtain the first-stage estimator of γ, denoted by $\tilde{\gamma}_{n,s}$. Secondly, estimate Ω_i in (3.20) by $\tilde{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\rho_{i,1}(\tilde{\gamma}_{n,S},\hat{h})\rho_{i,2}'(\tilde{\gamma}_{n,S},\hat{h})$ or alternatively by a nonparametric estimator, and then minimize $Q_{n,S}(\gamma,\hat{h})$ again with $\tilde{\Omega}^{-1}$ to obtain the second-stage estimator, $\tilde{\tilde{\gamma}}_{n,s}$. Since $\tilde{\Omega}$ is consistent for Ω_i, the asymptotic covariance of the second-stage estimator is given by $E\left[\frac{\partial\rho'(\gamma_0,\hat{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho'(\gamma_0,\hat{h})}{\partial\gamma}\right]^{-1}$. Consequently the second-stage estimator is asymptotically more efficient than the first-stage estimator.

For the choice of $l(\mathbf{x})$, in theory, it has no impact on the asymptotic efficiency of the estimator, as long as it has sufficiently large support. However, the choice of $l(\mathbf{x})$ will affect the finite sample variances of the simulated moments.

3.2 Asymptotic properties of simulation-based estimators

In Theorem 3.1 we will prove strong consistency of $\hat{\gamma}_{n,S}$ of (3.17), via authorizing the uniform convergence of $\frac{Q_{n,S}(\gamma,\hat{h})}{n}$ to a non-stochastic function $Q(\gamma,\hat{h})$ which has a unique minimizer $\gamma_0 \in \Gamma$. To do so, we will need the following regularity conditions.

Assumption 3.1. Assume that $E\|\Omega_i^{-1}\| \left(\|Y_i\| + \|Y_iW_i\| + \|Y_i^2\|\right)^2 < \infty$.

Assumption 3.2. The pdf $f_{\mathbf{U}}(\mathbf{u};\theta)$ of $\mathbf{U}$ is continuous in $\theta \in \Theta$ and continuously differentiable w.r.t. $\mathbf{u}$ for each $\theta \in \Theta$. Furthermore $f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)$ and $\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \mathbf{u}}$ for each $\theta \in \Theta$ are uniformly bounded by functions $\eta_1(\mathbf{x},\mathbf{z})$ and $\eta_2(\mathbf{x},\mathbf{z})$, respectively, which satisfy $E\|\Omega_i^{-1}\| \left(\int \eta_1(\mathbf{x},\mathbf{Z_i})(\|\mathbf{x}\|+1)d\mathbf{x}\right)^2 < \infty$ and $E\|\Omega_i^{-1}\| \left(\|\mathbf{Z_i}\| \int \eta_2(\mathbf{x},\mathbf{Z_i})(\|\mathbf{x}\|+1)d\mathbf{x}\right)^2 < \infty$

Assumption 3.3. $E\left[(\rho_i(\gamma,h_0)-\rho_i(\gamma_0,h_0))'\Omega^{-1}(\rho_i(\gamma,h_0)-\rho_i(\gamma_0,h_0))\right] = 0$ if and only if $\gamma = \gamma_0$.

These assumptions are common in the literature on GLM. In particular, Assumptions 3.1 and 3.2 contain moment conditions to assure that $\frac{Q_{n,S}(\gamma,\hat{h})}{n}$ uniformly converges to $Q(\gamma, h_0) = E\left[\rho_i'(\gamma, h_0)\Omega_i{}^{-1}\rho_i(\gamma, h_0)\right]$. Assumption 3.3 is an identification condition to ensure that $Q(\gamma, h_0)$ has a unique minimum at the true parameter value $\gamma_0 \in \Gamma_0$.

Theorem 3.1. Suppose that the support of $l(\mathbf{x})$ is $\mathbf{R}^p$. Then under Assumptions 3.1-3.3, $\hat{\gamma}_{n,s} \xrightarrow{a.s.} \gamma_0$, as $n \longrightarrow \infty$.

To derive the asymptotic normality for $\hat{\gamma}_{n,s}$ (cf. Theorem 3.2), we assume additional regularity conditions as follows.

Assumption 3.4. There exists an open subset $\Gamma_0 \subset \Gamma$ in which functions $\mathbf{x} f_{\mathbf{U}}(\mathbf{x} - \mathbf{Hz}, \theta)$, $\mathbf{x}\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \theta}$, $\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \theta}$, and $\frac{\partial^2 f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \theta^2}$ are uniformly bounded by a function $k(\mathbf{x}, \mathbf{z})$ that satisfies $E\|\Omega_i^{-1}\|\left(\int k(\mathbf{x}, \mathbf{Z_i})(\|\mathbf{x}\| + 1)d\mathbf{x}\right)^2 < \infty$.

Assumption 3.5. There exist an open subset $\Gamma_0 \subset \Gamma$ in which functions $x\frac{\partial f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial u'}$ and $\frac{\partial^2 f_{\mathbf{U}}(\mathbf{x}-\mathbf{Hz},\theta)}{\partial \theta \partial u'}$ are uniformly bounded by a function $r(\mathbf{x}, \mathbf{z})$ that satisfies $E\|\Omega_i^{-1}\|\left(\|\mathbf{Z_i}\| \int r(\mathbf{x}, \mathbf{Z_i})(\|\mathbf{x}\| + 1)d\mathbf{x}\right)^2 < \infty$.

Assumption 3.6. The matrix

$$\mathbf{K} = E\left[\frac{\partial \rho_i'(\gamma_0, \hat{h})}{\partial \gamma}\Omega_i^{-1}\frac{\partial \rho_i(\gamma_0, \hat{h})}{\partial \gamma'}\right] \tag{3.22}$$

is nonsingular.

Theorem 3.2. Under Assumptions 3.1 - 3.6, $\sqrt{n}(\hat{\gamma}_{n,S}-\gamma_0) \xrightarrow{d} N(0, \mathbf{K^{-1}D}\tau_{\mathbf{S}}\mathbf{D'K^{-1}})$, as $n \longrightarrow \infty$, where matrices $\mathbf{D}$ and $\tau_{\mathbf{S}}$ are given by

$$
\begin{aligned}
\mathbf{D} &= \left(\mathbf{I}_{p+k+J-1}, E\left[\frac{\partial \rho_1'(\gamma_0, h_0)}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_1(\gamma_0, h_0)}{\partial h'}\right](EZ_1Z_1' \otimes \mathbf{I}_q)^{-1}\right), \qquad (3.23)\\
\tau_{\mathbf{S},11} &= \frac{1}{2}E\left[\frac{\partial \rho_{1,1}'(\gamma_0, h_0)}{\partial \gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0, h_0)\rho_{1,2}'(\gamma_0, h_0)\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0, h_0)}{\partial \gamma'}\right] + \frac{1}{2}\mathbf{K},\\
\tau_{\mathbf{S},21} &= 0, \qquad (3.24)
\end{aligned}
$$

and $\tau_{\mathbf{S},22} = E\left[(Z_1 \otimes (W_1 - \mathbf{H}_0 Z_1))(Z_1 \otimes (W_1 - \mathbf{H}_0 Z_1))'\right]$.

Remark 3.2. When closed forms of the conditional moments in (3.10), (3.11), and (3.12) exist, the SBE $\hat{\gamma}_{n,S}$ becomes the method of moments estimator (MME) $\hat{\gamma}_n$ of Abarin and Wang (2012) defined as, $\hat{\gamma}_n = \arg\min Q_n(\gamma, \hat{h})$. Their MME is consistent and asymptotically normally distributed. In other words their MME has the following properties:

1. Under Assumptions 3.1-3.3, $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, as $n \longrightarrow \infty$.

2. Under Assumptions 3.1-3.6, $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \mathbf{K^{-1}D}\tau\mathbf{D'K^{-1}})$, as $n \longrightarrow \infty$, where

$$\tau = E\left[\frac{\partial \rho_i'(\gamma_0, h_0)}{\partial \gamma}\Omega_i^{-1}\rho_i(\gamma_0, h_0)\rho_i'(\gamma_0, h_0)\Omega_i^{-1}\frac{\partial \rho_i(\gamma_0, h_0)}{\partial \gamma'}\right]. \qquad (3.25)$$

3.3 Monte Carlo simulation studies

In this section we present simulation studies to demonstrate how the proposed SBE for γ can be calculated in the regression models (3.1)-(3.7), and to investigate the performance of these estimators for finite sample sizes.

In all the simulation studies, to compute the SBE's, we first choose the density of $l(\mathbf{x})$ such that the support of $l(\mathbf{x})$ covers the support of $r(\beta'\mathbf{x}_{is})f_U(\mathbf{x}_{is}-\nu;\theta)$ for all $\nu \in R^{\mathbf{q}}$ and $\gamma \in \Gamma$, and generate independent points $\mathbf{x}_{is}$, $s = 1, 2, \ldots, 2S, i = 1, 2, \ldots, n$ using S = 2,000 or 5,000 to calculate the simulation moments $m_1(\hat{\mathbf{H}}\mathbf{Z}_i;\gamma)$ of (3.15) and $m_2(\hat{\mathbf{H}}\mathbf{Z}_i;\gamma)$ of (3.16). Secondly, we calculate $m_1(\hat{\mathbf{H}}\mathbf{Z}_i;\gamma)$ and $m_2(\hat{\mathbf{H}}\mathbf{Z}_i;\gamma)$. The two-step SBE $\hat{\gamma}_{n,s}$ is calculated based on Remark 3.1 by minimizing $Q_{n,S}(\gamma,\hat{h})$, using the identity matrix I_3 to obtain the first-stage simulation based estimator (SBE1), and then estimate optimal weight matrix. Finally minimize $Q_{n,S}(\gamma,\hat{h})$ again with $\hat{\Omega}_i^{-1}$ to obtain the second-stage simulation based estimator (SBE2).

In this section we present simulation studies on models with ordinal response variable to demonstrate how the proposed SBE estimator can be calculated and its performance in finite sample sizes. To eliminate potential nonlinear numerical optimization problems, the true parameter values were used as starting values for the first-stage simulation based estimator, SBE1.

Along with the SBE's, naive estimator which ignores both the measurement error δ in (3.6) and equation (3.7) were investigated. All computations are done in R, and the naive estimators for cumulative link models are obtained with 'clm' from package 'ordinal'. To determine how well all these estimators perform, we present their mean, median, and RMSE.

3.3.1 Example 3.1

In our first simulation Example 3.1, the data $(y_i, w_i, z_i), i = 1, \ldots, n$ is generated using a Z from a standard normal distribution and the measurement errors δ from the normal distribution with mean of zero and standard deviation of 0.5. To simulate X_i via (3.7), we assumed that H from (3.7) is the identity matrix, and $U \sim N(0, \theta^2)$ with $\theta = 0.2$. We generated the response variable Y_i from $G^{-1} = \alpha_j - \beta' x$ where $j = 1, \ldots, J$, $J = 4$, and $\alpha_1 = -1.8$, $\alpha_2 = 0$, $\alpha_3 = 1.8$, and $\beta = 2$. The density of $l(x)$ as in (3.15) and (3.16) is $N(0, 2.25)$. $N = 1000$ Monte Carlo replications is carried out and, in each replication, $n = 400$ sample points (y_i, w_i, z_i) have been generated.

Table 3.1 summarizes the estimation results for this model with the logit link function, $G^{-1} = logit\,[P(Y \leq j|x)] = \alpha_j - \beta' x$ and Table 3.2 summarizes the estimation results for this model with the probit link function, $G^{-1} = \Phi^{-1}\,[P(Y \leq j|x)] = \alpha_j - \beta' x$. Although the SBE1 and SBE2 show their asymptotic properties, but they converge slower for the cut point parameters α_js, which are usually nuisance parameters of little interest. While in general compared to a mean, a median has the advantage of being more robust to outliers and invariant under a one-to-one transformation of parameters. These two are seen to be most reasonably close to each other in Tables 3.1 and 3.2. However, this will not be the case when covariate measurement error becomes substantial and outliers appear in some of the estimations.

Table 3.1: Simulation results of the logit link function in Example 3.1

		α_1	α_2	α_3	β	θ
		-1.8	0	1.8	2	0.2
SBE1	mean	-1.8835	0.0153	1.8872	2.0745	0.1551
	median	-1.8437	0.0097	1.8462	2.0487	0.1794
	RMSE	0.3248	0.3472	0.3268	0.2031	0.1712
SBE2	mean	-1.8692	-0.0028	1.8551	2.0769	0.2099
	median	-1.8638	0.0100	1.8334	2.0575	0.2045
	RMSE	0.2535	0.2986	0.2482	0.2201	0.1400
Naive	mean	-1.5920	-0.0035	1.5823	1.4177	–
	median	-1.5876	-0.0029	1.5755	1.4154	–
	RMSE	0.2533	0.1213	0.2624	0.5932	–

Table 3.2: Simulation results of the probit link function in Example 3.1

		α_1	α_2	α_3	β	θ
		-1.8	0	1.8	2	0.2
SBE1	mean	-2.1763	-0.0041	2.1543	2.2781	0.2165
	median	-1.9473	-0.0033	1.9041	2.2168	0.1931
	RMSE	0.7422	0.1802	0.6528	0.7681	0.1314
SBE2	mean	-2.0181	-0.0028	2.0044	2.2386	0.2034
	median	-1.8799	-0.0026	1.8731	2.0692	0.1868
	RMSE	0.7150	0.1717	0.6352	0.4353	0.1286
Naive	mean	-2.3405	-0.0039	2.3367	2.0903	–
	median	-2.3361	-0.0047	2.3301	2.0856	–
	RMSE	0.5661	0.1353	0.1663	0.5701	–

As seen from the Table 3.1 and 3.2, SBE2's are less biased and have smaller RMSE as compared to SBE1's. In addition, based on Remark 3.1 the second-stage estimator is asymptotically more efficient than the first-stage estimator. Therefore in the next simulations we only compare SBE2, and the Naive estimator.

Furthermore, for a small ratio of the variance of measurement error to the variance of X, RMSE of the SBE2 is smaller than Naive estimator for the regression coefficient of mismeasured covariate β. We can see from Table 3.1, for the logit link function, RMSE of the SBE2 and Naive estimators perform equally well for the cut point parameters α_j's. We can see from Table 3.2, for the probit link function, RMSE of the Naive estimators are smaller than SBE2 for the cut point parameters α_j's. The cut point parameters $\alpha_j' s$, are usually nuisance parameters of little interest, but if we are interested to estimate these cut points we need to generate more points from $l(\mathbf{x})$ to decrease the RMSE of the SBE2's.

3.3.2 Example 3.2

In our second simulation Example 3.2 the data $(y_i, w_{1i}, w_{2i}, z_{1i}, z_{2i}), i = 1, \ldots, n$ is generated from the model (3.1)-(3.7), with the instrumental variables

$$\mathbf{Z} = (z_1, z_2)' \sim N\left(\mathbf{0}, diag(1,1)\right)$$

and the measurement errors

$$\delta = (\delta_1, \delta_2)' \sim N\left(\mathbf{0}, diag(1,1)\right).$$

To simulate $\mathbf{X_i}$ via (3.9), we assume that $\mathbf{H} = diag(1,1)$, and

$$\mathbf{U} = (u_1, u_2)' \sim N\left(\mathbf{0}, diag(\theta_1^2, \theta_2^2)\right),$$

Table 3.3: Simulation results of the logit link function in Example 3.2

		α_1	α_2	α_3	β_1	β_2	θ_1	θ_2
	True	-4.1	0	4.1	3	4	0.7	0.8
SBE2	mean	-3.9435	0.0541	3.9376	2.8941	3.8623	0.5979	0.7084
	median	-3.3716	0.0888	3.4091	2.4955	3.3272	0.6442	0.7326
	RMSE	1.9344	0.6919	1.7364	1.3475	1.7965	0.2038	0.2869
Naive	mean	-1.6439	-0.0019	1.6372	0.7134	0.9915	–	–
	median	-1.6430	-0.00006	1.6326	0.7141	0.9907	–	–
	RMSE	2.4595	0.1162	2.4664	2.2876	3.0093	–	–

with $\theta_1 = 0.7$ and $\theta_2 = 0.8$. The density of $l(\mathbf{x})$ as in (3.15) and (3.16) is $N(\mathbf{0}, diag(2.25, 2.25))$. We generate the response variables Y_i from $G^{-1} = \alpha_j - \beta'\mathbf{x}$ where $j = 1, \ldots, J$, $J = 4$, $\alpha_1 = -4.1$, $\alpha_2 = 0$, $\alpha_3 = 4.1$, and $\beta = (\beta_1, \beta_2)' = (3, 4)'$. In each of the $N = 1000$ Monte Carlo replications, $n = 500$ sample points $(y_i, \mathbf{w}_i, \mathbf{z}_i)$ are generated.

Our estimation results for the cumulative logit link function

$$G^{-1} = logit\,[P(Y \leq j|\mathbf{x})] = \alpha_j - \beta'\mathbf{x}$$

and probit link function

$$G^{-1} = \Phi^{-1}\,[P(Y \leq j|\mathbf{x})] = \alpha_j - \beta'\mathbf{x}$$

are summarized in Table 3.3 and 3.4, respectively.

Clearly, SBE2's are a little biased, but much less so than the Naive estimators for almost all the regression coefficients. We can also see that RMSE of SBE2 are

Table 3.4: Simulation results of the probit link function in Example 3.2

		α_1	α_2	α_3	β_1	β_2	θ_1	θ_2
	True	-4.1	0	4.1	3	4	0.7	0.8
SBE2	mean	-3.4877	0.0476	3.4670	2.5451	3.3987	0.6248	0.7287
	median	-2.7536	0.0525	2.8786	2.0506	2.7444	0.6557	0.7387
	RMSE	2.0454	0.5563	1.8776	1.4125	1.8968	0.2505	0.1563
Naive	mean	-1.7607	-0.0016	1.7509	0.7638	1.0602	–	–
	median	-1.7594	-0.0006	1.7520	0.7628	1.0581	–	–
	RMSE	2.3435	0.1184	2.3530	2.2373	2.9407	–	–

smaller than RMSE of Naive estimators for the regression coefficients of mismeasured covariates, β_1 and β_2. Furthermore, based on RMSE, SBE2 and Naive estimators perform equally well in estimating the cut points α_1, α_2 and α_3.

3.3.3 Example 3.3

In this simulation Example 3.3, we consider a Generalized linear model (GLM). In particular, we generated the response variables Y_i from $G^{-1} = \alpha_j - \beta'\mathbf{x}$, where $j = 1, \ldots, J$, $J = 3$, $\alpha_1 = 1$, $\alpha_2 = 4$, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)' = (2, 1.5, 0.5, 1)'$, x_1 and x_2 are unobservable covariates with error. x_3 and x_4 are known predictors observed without error, and generated from

$$(x_3, x_4)' \sim N\left((1, 2)', diag(4, 1)\right).$$

To simulate x_1 and x_2 via (3.9), we assume that $\mathbf{H} = diag(1, 1)$,

$$\mathbf{U} = (u_1, u_2)' \sim N\left(\mathbf{0}, diag(\theta_1^2, \theta_2^2)\right),$$

with $\theta_1 = 0.5$ and $\theta_2 = 0.6$, and generated instrumental variables $\mathbf{z_1}$ and $\mathbf{z_2}$ from the standard normal distribution. Measurement errors $\delta = (\delta_1, \delta_2)'$ are generated from $N\left(\mathbf{0}, diag(var(\delta_1), var(\delta_2))\right)$ with different values of the variances which are reported in Tables 3.5 - 3.10. For the SBE, the density of $l(\mathbf{x})$ as in (3.15) and (3.16) is $N\left(\mathbf{0}, diag(1.96, 1.96)\right)$. $N = 1000$ Monte Carlo replications is carried out, and for the purpose of comparison of SBE2 and the Naive estimators, in each replication $n = 200, 400$, and 800 sample points $(y_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i)$ are generated.

Our simulation results for the cumulative logit link function

$$G^{-1} = logit\left[P(Y \leq j|\mathbf{x})\right] = \alpha_j - \beta'\mathbf{x},$$

are summarized in Tables 3.5, 3.6, and 3.7. In addition, results for the probit link function

$$G^{-1} = \Phi^{-1}\left[P(Y \leq j|\mathbf{x})\right] = \alpha_j - \beta'\mathbf{x},$$

are summarized in Tables 3.8, 3.9, and 3.10.

Both SBE2 and the Naive estimators seen to be a little biased. But still performing satisfactory, despite a relatively large number (eight) of unknown parameters. From Tables 3.5- 3.7 for the logit cumulative and Tables 3.8- 3.10 for the probit link when the variance of measurement error increase, our SBE2 has a much smaller bias compared to the Naive estimator.

We can also see that RMSE of SBE2 is smaller than RMSE of the Naive estimator for the regression coefficients of mismeasured covariates β_1 and β_2. Furthermore, based on RMSE, SBE2 and the Naive estimators perform equally well for the cut points and regression coefficients of observed covariates without error . The results also show that when the sample size increases the RMSE of each estimator decreases, as we expected. Note that, for large sample sizes, $n = 800$, RMSE of SBE2 is smaller than RMSE of the Naive estimators.

Table 3.5: Simulation results of the logit link function in Example 3.3, small measurement error

			var(δ_1) = 0.25 , var(δ_2) =0.36							
		n	α_1	α_2	β_1	β_2	β_3	β_4	θ_1	θ_2
	True		1	4	2	1.5	0.5	1	0.5	0.6
SBE2	mean	200	1.0446	4.2005	2.1266	1.5803	0.5283	1.0467	0.3787	0.5208
		400	0.9906	3.9999	2.0153	1.5085	0.5024	0.9946	0.3995	0.5085
		800	1.0241	4.0964	2.0648	1.5442	0.5143	1.0246	0.4416	0.5402
	median	200	1.0304	4.1647	2.0575	1.5358	0.5170	1.0326	0.3844	0.5310
		400	0.9872	3.9952	1.9856	1.4802	0.4994	0.9917	0.4421	0.5387
		800	1.0078	4.0046	2.0260	1.5121	0.5049	1.0046	0.4897	0.5915
	RMSE	200	0.5254	0.8188	0.5016	0.3954	0.1485	0.2674	0.2858	0.3666
		400	0.3526	0.5331	0.3173	0.2641	0.0980	0.1775	0.3047	0.2472
		800	0.2752	0.5843	0.3270	0.2535	0.0885	0.1759	0.2043	0.2576
Naive	mean	200	0.8261	3.3344	1.3953	0.9932	0.4198	0.8294	–	–
		400	0.8266	3.2877	1.3704	0.9698	0.4097	0.8242	–	–
		800	0.8115	3.2610	1.3590	0.9652	0.4087	0.8143	–	–
	median	200	0.9600	3.8853	1.8825	1.3658	0.4876	0.9688	–	–
		400	0.8081	3.2774	1.3655	0.9638	0.4084	0.8205	–	–
		800	0.8093	3.2470	1.3557	0.9627	0.4074	0.8132	–	–
	RMSE	200	0.4376	0.8160	0.6318	0.5287	0.1209	0.2460	–	–
		400	0.3324	0.7887	0.6427	0.5412	0.1094	0.2166	–	–
		800	0.2704	0.7749	0.6474	0.5403	0.1016	0.2045	–	–

Table 3.6: Simulation results of the logit link function in Example 3.3, moderate measurement error

				var(δ_1)	= 1.44	,	var(δ_2)	=1.69		
		n	α_1	α_2	β_1	β_2	β_3	β_4	θ_1	θ_2
	True		1	4	2	1.5	0.5	1	0.5	0.6
SBE2	mean	200	1.1770	4.5511	2.2921	1.7052	0.5680	1.1409	0.3630	0.4899
		400	1.0451	4.1624	2.0898	1.5635	0.5201	1.0405	0.3761	0.4857
		800	1.0282	4.1311	2.0845	1.5642	0.5202	1.0313	0.4107	0.5107
	median	200	1.1037	4.1737	2.1055	1.5734	0.5272	1.0658	0.3129	0.4164
		400	1.0380	4.0788	2.0195	1.5054	0.5043	1.0209	0.3635	0.4722
		800	1.0083	4.0030	2.0163	1.5111	0.5029	1.0014	0.4356	0.5269
	RMSE	200	0.6764	1.4861	0.8143	0.6000	0.2121	0.4254	0.3239	0.4206
		400	0.3924	0.7876	0.4431	0.3505	0.1236	0.2309	0.2877	0.3564
		800	0.3019	0.7298	0.3982	0.3103	0.1062	0.2057	0.3273	0.2584
Naive	mean	200	0.6287	2.5330	0.5926	0.4243	0.3196	0.6311	–	–
		400	0.6319	2.5074	0.5828	0.4148	0.3118	0.6296	–	–
		800	0.6172	2.4858	0.5785	0.4147	0.3121	0.6203	–	–
	median	200	0.6236	2.5264	0.5923	0.4193	0.3165	0.6324	–	–
		400	0.6254	2.4998	0.5802	0.4121	0.3124	0.6330	–	–
		800	0.6189	2.4801	0.5775	0.4138	0.3113	0.6228	–	–
	RMSE	200	0.5146	1.5208	1.4110	1.0795	0.1972	0.4008	–	–
		400	0.4451	1.5186	1.4190	1.0872	0.3871	0.1958	–	–
		800	0.4202	1.5267	1.4224	1.0863	0.1919	0.3872	–	–

Table 3.7: Simulation results of the logit link function in Example 3.3, large measurement error

			var(δ_1) = 2.89, var(δ_2) =3.24							
		n	α_1	α_2	β_1	β_2	β_3	β_4	θ_1	θ_2
	True		1	4	2	1.5	0.5	1	0.5	0.6
SBE2	mean	200	1.1457	4.4786	2.2463	1.6865	0.5589	1.1221	0.3419	0.4585
		400	1.0341	4.1139	2.0701	1.5506	0.5181	1.0283	0.3582	0.4661
		800	1.0255	4.1146	2.0722	1.5584	0.5162	1.0274	0.3882	0.4822
	median	200	1.0776	4.1349	2.0815	1.5480	0.5234	1.0481	0.2834	0.3849
		400	1.0280	4.0165	1.9824	1.4760	0.4975	1.0013	0.3177	0.4330
		800	1.0017	3.9798	1.9989	1.5023	0.5006	0.9934	0.3777	0.4792
	RMSE	200	0.6534	1.4962	0.8005	0.6393	0.2134	0.4094	0.3307	0.4080
		400	0.3850	0.7849	0.4483	0.3538	0.1296	0.2254	0.3010	0.3662
		800	0.3004	0.7727	0.4141	0.3244	0.1068	0.2130	0.2830	0.3576
Naive	mean	200	0.5783	2.3161	0.3528	0.2554	0.2930	0.5797	–	–
		400	0.5805	2.2956	0.3467	0.2508	0.2855	0.5770	–	–
		800	0.5661	2.2749	0.3444	0.2517	0.2859	0.5679	–	–
	median	200	0.5743	2.3042	0.3520	0.2557	0.2922	0.5780	–	–
		400	0.5760	2.2921	0.3448	0.2498	0.2863	0.5815	–	–
		800	0.5660	2.2692	0.3443	0.2515	0.2845	0.5694	–	–
	RMSE	200	0.5359	1.7247	1.6489	1.2466	0.2204	0.4452	–	–
		400	0.4839	1.7249	1.6542	1.2502	0.2206	0.4367	–	–
		800	0.4652	1.7351	1.6560	1.2488	0.2174	0.4382	–	–

Table 3.8: Simulation results of the probit link function in Example 3.3, small measurement error

				var(δ_1)	= 0.25	,	var(δ_2)	=0.36		
		n	α_1	α_2	β_1	β_2	β_3	β_4	θ_1	θ_2
	True		1	4	2	1.5	0.5	1	0.5	0.6
SBE2	mean	200	1.0579	4.1450	2.0807	1.5541	0.5184	1.0361	0.3213	0.4160
		400	1.0331	4.0510	2.0063	1.4950	0.5055	1.0137	0.4118	0.5216
		800	0.9888	4.0509	2.0449	1.5248	0.5236	0.9929	0.4660	0.5524
	median	200	0.9723	3.7097	1.8087	1.3578	0.4638	0.9250	0.2816	0.3695
		400	0.9977	3.9306	1.9374	1.4502	0.4877	0.9870	0.4603	0.5565
		800	0.9445	3.8246	1.9564	1.4356	0.5107	0.9461	0.4775	0.5801
	RMSE	200	0.5580	1.5664	0.8582	0.6510	0.2199	0.4144	0.3008	0.3743
		400	0.2938	0.6644	0.3717	0.2992	0.1073	0.1867	0.2112	0.2614
		800	0.2411	0.5871	0.3393	0.2815	0.0952	0.1505	0.1744	0.1978
Naive	mean	200	1.1384	4.6262	1.9301	1.3735	0.5807	1.1522	–	–
		400	1.1411	4.5631	1.8936	1.3429	0.5679	1.1416	–	–
		800	1.1118	4.5027	1.8769	1.3343	0.5643	1.1222	–	–
	median	200	1.1323	4.6021	1.9174	1.3563	0.5763	1.1443	–	–
		400	1.1264	4.5549	1.8873	1.3389	0.5692	1.1395	–	–
		800	1.1083	4.5009	1.8743	1.3326	0.5622	1.1239	–	–
	RMSE	200	0.4646	0.8654	0.2371	0.2219	0.1331	0.2601	–	–
		400	0.3437	0.7008	0.1920	0.2055	0.0989	0.2042	–	–
		800	0.2424	0.5743	0.1655	0.1896	0.0816	0.1557	–	–

Table 3.9: Simulation results of the probit link function in Example 3.3, moderate measurement error

			var(δ_1) = 1.44 , var(δ_2) =1.69							
		n	α_1	α_2	β_1	β_2	β_3	β_4	θ_1	θ_2
	True		1	4	2	1.5	0.5	1	0.5	0.6
SBE2	mean	200	1.0321	3.9673	1.9663	1.4680	0.4950	0.9950	0.3260	0.4260
		400	1.0576	4.1880	2.0875	1.5526	0.5210	1.0519	0.3668	0.5135
		800	1.0279	4.0784	2.0319	1.5249	0.5146	1.0265	0.4071	0.5450
	median	200	0.9981	3.7440	1.8531	1.3883	0.4692	0.9376	0.3046	0.3933
		400	1.0133	3.9607	1.9757	1.4588	0.5102	0.9952	0.3759	0.5431
		800	0.9991	4.0564	2.0147	1.4864	0.5097	0.9971	0.4203	0.5629
	RMSE	200	0.4039	0.9278	0.5349	0.4196	0.1507	0.2562	0.2928	0.3577
		400	0.3219	0.9151	0.5168	0.3998	0.1324	0.2414	0.2596	0.2993
		800	0.3072	0.9048	0.5128	0.3746	0.1307	0.2409	0.2481	0.2864
Naive	mean	200	0.7447	3.0211	0.7055	0.5051	0.3805	0.7534	–	–
		400	0.7508	2.9954	0.6926	0.4954	0.3723	0.7507	–	–
		800	0.7298	2.9590	0.6889	0.4947	0.3716	0.7370	–	–
	median	200	0.7280	3.0118	0.7055	0.5024	0.3764	0.7475	–	–
		400	0.7446	2.9867	0.6926	0.4929	0.3726	0.7489	–	–
		800	0.7269	2.9640	0.6888	0.4948	0.3709	0.7372	–	–
	RMSE	200	0.4484	1.0710	1.2988	0.9994	0.1457	0.2981	–	–
		400	0.3563	1.0461	1.3096	1.0071	0.1397	0.2752	–	–
		800	0.3255	1.0614	1.3121	1.0064	0.1347	0.2746	–	–

Table 3.10: Simulation results of the probit link function in Example 3.3, large measurement error

				var(δ_1)	= 2.89	,	var(δ_2)	=3.24		
		n	α_1	α_2	β_1	β_2	β_3	β_4	θ_1	θ_2
	True		1	4	2	1.5	0.5	1	0.5	0.6
SBE2	mean	200	1.1049	4.2800	2.1193	1.5959	0.5325	1.0648	0.3453	0.4326
		400	1.0439	4.1462	2.0632	1.5738	0.5171	1.0376	0.3674	0.4666
		800	1.0384	4.0979	2.0547	1.5540	0.5080	1.0211	0.3975	0.4796
	median	200	1.0682	3.9791	1.9780	1.4660	0.5028	0.9929	0.3316	0.4437
		400	1.0071	3.9514	1.9478	1.4869	0.4924	1.0079	0.3730	0.4559
		800	1.0042	3.9794	1.9765	1.4887	0.49741	0.9997	0.4108	0.4874
	RMSE	200	0.4246	1.0751	0.6280	0.5268	0.1566	0.2849	0.2908	0.34602
		400	0.3510	0.9462	0.4951	0.4343	0.1317	0.2436	0.2726	0.3304
		800	0.3374	0.9147	0.4791	0.4077	0.1160	0.2371	0.2679	0.3240
Naive	mean	200	0.6643	2.6828	0.4082	0.2969	0.3382	0.6702	–	–
		400	0.6699	2.6625	0.3998	0.2911	0.3311	0.6681	–	–
		800	0.6501	2.6291	0.3983	0.2918	0.3305	0.6552	–	–
	median	200	0.6548	2.6703	0.4064	0.2974	0.3345	0.6634	–	–
		400	0.6647	2.6576	0.4002	0.2892	0.3323	0.6653	–	–
		800	0.6521	2.6276	0.3977	0.2922	0.3298	0.6561	–	–
	RMSE	200	0.4852	1.3778	1.5937	1.2052	0.1802	0.3657	–	–
		400	0.4102	1.3645	1.6012	1.2101	0.1772	0.3500	–	–
		800	0.3906	1.3848	1.6021	1.2087	0.1738	0.3530	–	–

3.4 Real data example

In this section, we analyze the performance of our SBE2 for γ for the data set from the Adult Literacy and Life Skills Survey (ALL). This survey is a large-scale co-operative effort undertaken in 2003 by governments, national statistics agencies, research institutions and multi-lateral agencies. This data included civilian non-institutionalized persons 16 to 65 years old. The original collector of the data is Statistics Canada and bears no responsibility for uses of this data.

In our model we consider 3310 persons who live in Norway and have an income. In this model, the response variable Y is general satisfaction and feeling good (quartiles) in life which has 4 ordered categories ranging from 1, 'completely dissatisfied', to 4, 'completely satisfied'. X_1 is the wealth of respondent which is unobservable covariates with error. Since X_1 could not be measured precisely, we use the log of total household income of respondent as an observed variable, W_1. For the instrumental variables Z_1 and Z_2, we used the log of the total personal income of a respondent and the total household income (purchasing power parity), respectively. Furthermore, we considered the log of physical and mental component summary as known predictors observed without error, X_2 and X_3. Therefore we consider the following model

$$\begin{aligned} G^{-1} &= \alpha_j - \beta_1\, x_1 - \beta_2 x_2 - \beta_3 x_3, \qquad j = 1, 2, 3 \qquad (3.26) \\ W_1 &= X_1 + \delta \\ X_1 &= h_1 + h_2 Z_1 + h_3 Z_2 + U \end{aligned}$$

We estimate h_1, h_2, and h_3 by the least squares estimators: 2.9154, 0.0424, and 0.9016, respectively. For the SBE2 we choose the density $l(x)$ to be $N(13, 1.69)$. We

Table 3.11: SBE2 estimates and the empirical standard error for the ALL data

	α_1	α_2	α_3	β_1	β_2	$\beta 3$	θ
logit-SBE2	133.2967	137.3602	139.3601	0.0634	2.9690	30.1369	0.1823
Emp SE	0.0736	0.0410	0.0985	0.0184	0.0548	0.0208	0.1056
probit-SBE2	134.0004	137.0021	138.9996	0.0592	2.9520	29.9875	0.1920
Emp SE	0.1629	0.2350	0.1313	0.0176	0.0819	0.0415	0.1120
Naive	134.5391	136.9774	138.9440	0.0310	2.8989	30.6786	–
Emp SE	2.0173	2.0349	2.0459	0.0142	0.0796	0.4743	–

also estimated empirical standard errors of regression coefficients in (3.26) and θ, via bootstrap method using 100 independent bootstrap samples each consisting of 400 data values drawing with replacement from 3310 persons who live in Norway and have an income.

The consecutive estimates and their empirical standard errors (Emp SE) for regression coefficients in (3.26) and θ are reported in Table 3.11 for the cumulative logit link function $G^{-1} = logit\,[P(Y \leq j|\mathbf{x})] = \alpha_j - \beta_1\ x_1 - \beta_2 x_2 - \beta_3 x_3$, probit link function $G^{-1} = \Phi^{-1}\,[P(Y \leq j|\mathbf{x})] = \alpha_j - \beta_1\ x_1 - \beta_2 x_2 - \beta_3 x_3$, and the Naive estimator.

As we can see from the Table 3.11, our SBE2 for the probit and logit link functions and Naive estimator are much similar in estimating all the unknown parameters. This could be the result of small measurement error, δ, in (3.26). We can also see, all covariates of our regression model have positive slopes. This means that wealth, physical, and mental component summary of respondent have positive effect on general satisfaction and feeling good (quartiles) in life. We can also see that since the estimated coefficients of wealth, physical and mental component summary are significantly different from zero, they all have positive effect on general statisfaction and feeling good in life compared to the wealth of respondent.

3.5 Conclusions and Discussion

Regression models for binary outcomes are the foundation for studying more complex models, such as ordinal, nominal, and count models. There are limited papers about regression models with ordinal response variables and mismeasured covariates. In this chapter we extended the results obtained in Chapter 2 for the regression model with binary responses to allow for ordinal responses. We considered method of moments estimators that uses instrumental variables for therein unknown parameters that are defined via minimizing respective objective functions. In other words we studied SBE's for a cumulative logit and porbit models with classical measurement error in covariates, via using instrumental variables.

We proved consistency and asymptotic normality of the proposed SBE's under some regularity conditions. Furthermore, the performance of the SBE's is illustrated through simulation studies, and for the purpose of comparison along with the SBE's, naive estimator which ignores both the measurement error δ in (3.6) and equation (3.7) were investigated. Finally a real data application for the data set from the Adult Literacy and Life Skills (ALL) Survey was given.

Chapter 4

Conclusions and Discussion

This thesis deals with estimation in two popular regression models with a categorical response variables. In both regression models we considered method of moments estimators that uses instrumental variables for therein unknown parameters that are defined via minimizing respective objective functions. The focus of this thesis is to extend SLSE approach of Wang (2003) further to regression models with the categorical response variable and mismeasured covariates. Since the closed forms of conditional moments are not available for these models, we proposed Simulation-Based Estimators (SBE) for consistent estimation of the parameters of the model. This method does not require normality assumption on covariates and measurement errors, and can be considered as a generalization of Buzas and Stefanski (1996) with much more flexible assumptions.

Although, estimation of logistic and probit models with mismeasured covariates and non-normal random errors have being considered before but most of the authors rely on restrictive conditions to achieve consistent estimation. Furthermore, most of the methods in the literature are limited to the case where either validation or replicate data are available, which are restrictive in many real data experiments.The first contribution is to use the similar method of Abarin and Wang (2012) to propose and study SBE's for a regression model with binary response variable and classical measurement error in covariates, via using instrumental variables. This method does not require parametric assumptions for the distributions of the unobserved covariates and error components. We proved consistency and asymptotic normality of the proposed SBE's under some regularity conditions. The performance of the SBE's for the logistic and probit regression models are illustrated through simulation studies and a real data set from the AIDS Clinical Trial Group (ACTG175) study.

Regression models for binary outcomes are the foundation for studying more complex models, such as ordinal, nominal, and count models. There are limited papers about regression models with ordinal response variables and mismeasured covariates. The second contribution is to extend the results obtained in the first part for the regression model with binary responses to allow for ordinal responses. In other words we studied SBE's for a cumulative logit and porbit models with classical measurement error in covariates, via using instrumental variables. We proved consistency and asymptotic normality of the proposed SBE's under some regularity conditions. Furthermore, we used simulations to illustrate the performance of the SBE's in finite sample situation. The real data set from the Adult Literacy and Life Skills (ALL) Survey is also applied in this regard.

In the future, it would be desirable to study asymptotic confidence intervals that are based on our SBE's. The second possible extension of thesis can be an extension of SBE's for the regression model with a categorical response variable and mismeasured covariates with Berkson type measurement error. The third possible direction of future research is extending the SBE's for the regression models with discrete mismeasured covariates. It would also be desirable to develop proposed methodology to cope with missing data, since missing data and measurement error often arise simultaneously in a real world problem.

Chapter 5

Appendix

Throughout the proofs, we use the following Lemmas.

Lemma 1. Uniform Strong Law of Large Numbers (USLLN) (Jennrich (1969), Theorem 2)

Let g be a function on $\mathscr{X} \times \Theta$ where $\mathscr{X}$ is a Euclidean space and Θ is a compact subset of Euclidean space. Let $g(x, \theta)$ be a continuous function of θ for each x and a measurable function of x for each θ. Assume also that $|g(x, \theta)| \leq h(x)$ for all x and θ, where h is integrable with respect to a probability distribution function F on $\mathscr{X}$. If $x_1, x_2, \ldots$ is a random sample from F then for almost ever sequence (x_t)

$$n^{-1} \sum_{t=1}^{n} g(x_t, \theta) \longrightarrow \int g(x, \theta) dF(x)$$

uniformly for all θ in Θ.

Lemma 2. (lemma 3 of Amemiya (1973))

Let $Q_T(w,\theta)$ be a measurable function on a measurable space Ω and for each w in Ω a continuous function for θ in a compact set Θ. Then there exists a measurable function $\hat{\theta}_T(w)$ such that for all w in Ω,

$$Q_T\left(w,\hat{\theta}_T(w)\right) = \sup_{\theta\in\Theta} Q_T(w,\theta).$$

If $Q_T(w,\theta)$ converges to $Q(\theta)$ a.s. uniformly for all θ in Θ, and if $Q(\theta)$ has a unique maximum at $\theta_0 \in \Theta$, then $\hat{\theta}_T$ converges to θ_0 a.s.

Lemma 3. (lemma 4 of Amemiya (1973))

Let $f_T(w,\theta)$ be a measurable function on a measurable space Ω and for each w in Ω a continuous function for θ in a compact space set Θ. If $f_T(w,\theta)$ converges to $f(\theta)$ a.s. uniformly for all θ in Θ, and if $\hat{\theta}_T(w)$ converges to θ_0 a.s., then $f_T(w,\hat{\theta}_T(w))$ converges to $f(\theta_0)$ a.s.

5.1 Proof of Theorem 2.1

To prove the consistency, we use the Lemma 1 (USLLN). The idea is to show that $E\left[\sup_\Gamma |\rho'_{j,1}(\gamma,h_0){\Omega_j}^{-1}\rho_{j,2}(\gamma,h_0)|\right] < \infty$. For sufficiently large n, $Q_{n,s}(\gamma,\hat{h})$ has the first order Taylor expansion about $h_0 = vecH_0$,

$$\begin{aligned} Q_{n,s}(\gamma,\hat{h}) = & \sum{}_{j=1}^{n} \rho'_{j,1}(\gamma,h_0)\Omega^{-1}\rho_{j,2}(\gamma,h_0) \qquad (5.1) \\ & + \sum_{j=1}^{n}\left[\rho'_{j,1}(\gamma,\tilde{h})\Omega^{-1}\frac{\partial\rho_{j,2}(\gamma,\tilde{h})}{\partial h'} + \rho'_{j,2}(\gamma,\tilde{h})\Omega^{-1}\frac{\partial\rho_{j,1}(\gamma,\tilde{h})}{\partial h'}\right](\hat{h}-h_0) \end{aligned}$$

and $\tilde{h} = vec\tilde{H}$ satisfies $\|\tilde{h} - h_0\| \le \|\hat{h} - h_0\|$, where

$$\rho_{j,1}(\gamma, h_0) = Y_j T_j - m_1(H_0 Z_j, \gamma),$$

$$\rho_{j,2}(\gamma, h_0) = Y_j T_j - m_2(H_0 Z_j, \gamma),$$

$$\rho_{j,1}(\gamma, \tilde{h}) = Y_j T_j - m_1(\tilde{H} Z_j, \gamma),$$

$$\rho_{j,2}(\gamma, \tilde{h}) = Y_j T_j - m_2(\tilde{H} Z_j, \gamma),$$

$$\frac{\partial \rho_{j,1}(\gamma, \tilde{h})}{\partial h'} = \frac{1}{S} \sum_{s=1}^{S} \frac{\tilde{x}_{js} p(\alpha + \beta' x_{js})}{l(x_{js})} . \frac{\partial f_U(x_{js} - \tilde{H} Z_j; \theta)}{\partial u'} (Z_j \otimes I_q)',$$

and

$$\frac{\partial \rho_{j,2}(\gamma, \tilde{h})}{\partial h'} = \frac{1}{S} \sum_{s=S+1}^{2S} \frac{\tilde{x}_{js} p(\alpha + \beta' x_{js})}{l(x_{js})} . \frac{\partial f_U(x_{js} - \tilde{H} Z_j; \theta)}{\partial u'} (Z_j \otimes I_q)'.$$

By Assumptions 2.1, 2.2 and the dominated convergence theorem,

$$\begin{aligned}
& E \quad \left[\sup_{\Gamma} \|\rho_{1,1}(\gamma, h_0)\| \, |Z_1, Y_1, W_1 \right] \\
\le & \; E \quad |Y_1| + E|Y_1 W_1| \\
+ & \; \frac{1}{S} \sum_{s=1}^{S} E \left[\sup_{\Gamma} \frac{(\|x_{1s}\| + 1) p(\alpha + \beta' x_{1s}) f_U(x_{1s} - H_0 Z_1; \theta)}{l(x_{1s})} |Z_1, Y_1, W_1 \right] \\
\le & \; E \quad |Y_1| + E|Y_1 W_1| + \int \sup_{\Gamma} (\|x\| + 1) p(\alpha + \beta' x) f_U(x - H_0 Z_1; \theta) dx \\
\le & \; 1 \quad + \|W_1\| + \int (\|x\| + 1) \eta_1(x, Z_1) dx.
\end{aligned}$$

Similarly,

$$E\left[\sup_{\Gamma}\|\rho_{1,2}(\gamma,h_0)\|\,|Z_1,Y_1,W_1\right] \leq 1+\|W_1\|+\int(\|x\|+1)\eta_1(x,Z_1)dx$$

Furthermore, since $\rho_{1,1}(\gamma)$ and $\rho_{1,2}(\gamma)$ are conditionally independent given (Z_1,Y_1,W_1), using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
E & \quad \left[\sup_{\Gamma}|\rho'_{1,1}(\gamma,h_0){\Omega_1}^{-1}\rho_{1,2}(\gamma,h_0)|\right] \\
\leq & \quad E\left[\|{\Omega_1}^{-1}\|E(\sup_{\Gamma}\|\rho'_{1,1}(\gamma,h_0)\||Z_1,Y_1,W_1)E(\sup_{\Gamma}\|\rho_{1,2}(\gamma,h_0)\||Z_1,Y_1,W_1)\right] \\
\leq & \quad E\left[\|{\Omega_1}^{-1}\|\left(1+\|W_1\|+\int(\|x\|+1)p(\alpha+\beta'x)\eta_1(x,Z_1)dx\right)^2\right] \\
\leq & \quad 2E\|{\Omega_1}^{-1}\|\,(1+\|W_1\|)^2+2E\|{\Omega_1}^{-1}\|(\int\eta_1(x,Z_1)(\|x\|+1)dx)^2 \\
< & \quad \infty
\end{aligned}$$

Therefore by USLLN,

$$\sup_{\Gamma}\left\|\frac{1}{n}\sum_{j=1}^{n}\rho'_{j,1}(\gamma,h_0){\Omega_j}^{-1}\rho_{j,2}(\gamma,h_0)-E\left[\rho'_{j,1}(\gamma,h_0){\Omega_j}^{-1}\rho_{j,2}(\gamma,h_0)\right]\right\| \xrightarrow{a.s.} 0, \quad (5.2)$$

where

$$\begin{aligned}
E & \quad \left[\rho'_{1,1}(\gamma,h_0)\,{\Omega_1}^{-1}\rho_{1,2}(\gamma,h_0)\right] \\
= & \quad E\left[E(\rho'_{1,1}(\gamma,h_0)|Z_1,Y_1,W_1){\Omega_1}^{-1}E(\rho_{1,2}(\gamma,h_0)|Z_1,Y_1,W_1)\right] \\
= & \quad E\left[\rho'_1(\gamma,h_0){\Omega_1}^{-1}\rho_1(\gamma,h_0)\right] \\
= & \quad Q(\gamma,h_0)
\end{aligned}$$

Similarly by the Cauchy-Schwarz inequality and Assumption 2.2,

$$
\begin{aligned}
& \left[E \sup_{\Gamma} \left\| \rho'_{1,1}(\gamma,\tilde{h}) \Omega_1^{-1} \frac{\partial \rho_{1,2}(\gamma,\tilde{h})}{\partial h'} \right\| \right]^2 \\
\leq \quad & \left[E \left\| \Omega_1^{-1} \right\| \sup_{\Gamma,h} \left\| \rho'_{1,1}(\gamma,h) \right\| \sup_{\Gamma,h} \left\| \frac{\partial \rho_{1,2}(\gamma,h)}{\partial h'} \right\| \right]^2 \\
\leq \quad & \left[E \|\Omega_1^{-1}\| \sup_{\Gamma,h} \|\rho'_{1,1}(\gamma,h)\|^2 \right] \left[E \|\Omega_1^{-1}\| \sup_{\Gamma,h} \left\| \frac{\partial \rho_{1,2}(\gamma,h)}{\partial h'} \right\|^2 \right] \\
\leq \quad & \left[E \|\Omega_1^{-1}\| \sup_{\Gamma,h} \|\rho'_{1,1}(\gamma,h)\|^2 \right] \left[E \|\Omega_1^{-1}\| \left(\|Z_1\| \int (\|x\|+1) \eta_2(x,Z_1)\ dx \right)^2 \right] \\
< \quad & \infty,
\end{aligned}
$$

and

$$
\begin{aligned}
& \left[E \sup_{\Gamma} \left\| \rho'_{1,2}(\gamma,\tilde{h}) \Omega_1^{-1} \frac{\partial \rho_{1,1}(\gamma,\tilde{h})}{\partial h'} \right\| \right]^2 \\
\leq \quad & \left[E \left\| \Omega_1^{-1} \right\| \sup_{\Gamma,h} \left\| \rho'_{1,2}(\gamma,h) \right\| \sup_{\Gamma,h} \left\| \frac{\partial \rho_{1,1}(\gamma,h)}{\partial h'} \right\| \right]^2 \\
\leq \quad & \left[E \|\Omega_1^{-1}\| \sup_{\Gamma,h} \|\rho'_{1,2}(\gamma,h)\|^2 \right] \left[E \|\Omega_1^{-1}\| \left(\|Z_1\| \int (\|x\|+1) \eta_2(x,Z_1)\ dx \right)^2 \right] \\
< \quad & \infty.
\end{aligned}
$$

Again by Lemma 1 (USLLN), we have,

$$
\sup_{\Gamma} \left\| \frac{1}{n} \sum_{j=1}^{n} \rho'_{j,1}(\gamma,\tilde{h}) \Omega_j^{-1} \frac{\partial \rho_{j,2}(\gamma,\tilde{h})}{\partial h'} \right\| \xrightarrow{a.s.} 0,
$$

$$
\sup_{\Gamma} \left\| \frac{1}{n} \sum_{j=1}^{n} \rho'_{j,2}(\gamma,\tilde{h}) \Omega_j^{-1} \frac{\partial \rho_{j,1}(\gamma,\tilde{h})}{\partial h'} \right\| \xrightarrow{a.s.} 0.
$$

Therefore,

$$\begin{aligned}
& \sup_{\Gamma}\left\|\frac{1}{n}\sum_{j=1}^{n}\left[\rho'_{j,1}(\gamma,\tilde{h})\Omega_j^{-1}\frac{\partial\rho_{j,2}(\gamma,\tilde{h})}{\partial h'}+\rho'_{j,2}(\gamma,\tilde{h})\Omega_j^{-1}\frac{\partial\rho_{j,1}(\gamma,\tilde{h})}{\partial h'}\right](\hat{h}-h_0)\right\| \\
\leq \quad & \sup_{\Gamma}\left\|\frac{1}{n}\sum_{j=1}^{n}\rho'_{j,1}(\gamma,\tilde{h})\Omega_j^{-1}\frac{\partial\rho_{j,2}(\gamma,\tilde{h})}{\partial h'}\right\|\left\|(\hat{h}-h_0)\right\| \\
& +\sup_{\Gamma}\left\|\frac{1}{n}\sum_{j=1}^{n}\rho'_{j,2}(\gamma,\tilde{h})\Omega_j^{-1}\frac{\partial\rho_{j,1}(\gamma,\tilde{h})}{\partial h'}\right\|\left\|(\hat{h}-h_0)\right\| \\
\xrightarrow{a.s.} \quad & 0 \qquad (5.3)
\end{aligned}$$

It follows from (5.1) - (5.3) that

$$\sup_{\Gamma}\left|\frac{1}{n}Q_{n,s}(\gamma,\hat{h})-Q(\gamma,h_0)\right|\xrightarrow{a.s.}0$$

where, $Q(\gamma,h_0)=E\left[\rho_1'(\gamma,h_0){\Omega_1}^{-1}\rho_1(\gamma,h_0)\right]$.

Now for the next step, we need to show that $Q(\gamma,h_0)$ has a unique minimum at $\gamma_0\in\Gamma$. Since $E(\rho_1'(\gamma_0,h_0)|Z_1)=0$ and $\rho_1(\gamma,h_0)-\rho_1(\gamma_0,h_0)$ depends on Z_1 only, we have

$$\begin{aligned}
E & \left[\rho_1'(\gamma_0,h_0)\Omega_1^{-1}(\rho_1(\gamma,h_0)-\rho_1(\gamma_0,h_0))\right] \\
= & E\left[E(\rho_1'(\gamma_0,h_0)|Z_1)\Omega_1^{-1}(\rho_1(\gamma,h_0)-\rho_1(\gamma_0,h_0))\right] \\
= & 0,
\end{aligned}$$

which implies

$$
\begin{aligned}
& Q(\gamma, h_0) \\
= \; & E\left[\left(\rho_1'(\gamma, h_0) - \rho_1'(\gamma_0, h_0) + \rho_1'(\gamma_0, h_0)\right)\Omega_1^{-1}\left(\rho_1(\gamma, h_0) - \rho_1(\gamma_0, h_0) + \rho_1(\gamma_0, h_0)\right)\right] \\
= \; & Q(\gamma_0, h_0) + E\left[\left(\rho_1(\gamma, h_0) - \rho_1(\gamma_0, h_0)\right)'\Omega_1^{-1}\left(\rho_1(\gamma, h_0) - \rho_1(\gamma_0, h_0)\right)\right].
\end{aligned}
$$

Therefore $Q(\gamma, h_0) \geq Q(\gamma_0, h_0)$ and the equality holds if and only if $\gamma = \gamma_0$, by Assumption 2.3. Subsequently all conditions of Lemma 2 are satisfied, so we have $\hat{\gamma}_{n,s} \xrightarrow{a.s.} \gamma_0$, as $n \longrightarrow \infty$.

5.2 Proof of Theorem 2.2

By Assumption 2.4, the first derivative $\frac{\partial Q_{n,s}(\hat{\gamma}_{n,S}, \hat{h})}{\partial \gamma}$ exists and has the first order Taylor expansion in an open neighbourhood $\Gamma_0 \subset \Gamma$ of γ_0;

$$
\frac{\partial Q_{n,s}(\hat{\gamma}_{n,S}, \hat{h})}{\partial \gamma} = \frac{\partial Q_{n,s}(\gamma_0, \hat{h})}{\partial \gamma} + \frac{\partial^2 Q_{n,s}(\tilde{\gamma}, \hat{h})}{\partial \gamma \partial \gamma'}(\hat{\gamma}_{n,s} - \gamma_0) = 0 \tag{5.4}
$$

where $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma}_{n,s} - \gamma_0\|$. The first and second derivative of $Q_{n,s}$ is given by

$$
\frac{\partial Q_{n,s}(\gamma, \hat{h})}{\partial \gamma} = \sum_{j=1}^{n}\left[\frac{\partial \rho_{j,1}'(\gamma, \hat{h})}{\partial \gamma}\Omega_j^{-1}\rho_{j,2}(\gamma, \hat{h}) + \frac{\partial \rho_{j,2}'(\gamma, \hat{h})}{\partial \gamma}\Omega_j^{-1}\rho_{j,1}(\gamma, \hat{h})\right] \tag{5.5}
$$

and

$$
\begin{aligned}
& \frac{\partial^2 Q_{n,s}(\gamma,\hat{h})}{\partial\gamma\partial\gamma'} \\
= \; & \sum_{j=1}^{n}\left[\frac{\partial\rho'_{j,1}(\gamma,\hat{h})}{\partial\gamma}\Omega_j^{-1}\frac{\partial\rho_{j,2}(\gamma,\hat{h})}{\partial\gamma'} + (\rho'_{j,2}(\gamma,\hat{h})\Omega_j^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial\rho'_{j,1}(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right] \\
+ \; & \sum_{j=1}^{n}\left[\frac{\partial\rho'_{j,2}(\gamma,\hat{h})}{\partial\gamma}\Omega_j^{-1}\frac{\partial\rho_{j,1}(\gamma,\hat{h})}{\partial\gamma'} + (\rho'_{j,1}(\gamma,\hat{h})\Omega_j^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial\rho'_{j,2}(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right].
\end{aligned}
$$

Assumptions 2.1 - 2.4 imply that

$$
\begin{aligned}
E\left[\sup_{\gamma}\left\|\frac{\partial\rho'_{j,1}(\gamma,\hat{h})}{\partial\gamma}\Omega_j^{-1}\frac{\partial\rho_{j,2}(\gamma,\hat{h})}{\partial\gamma'}\right\|\right] &< \infty \\
E\left[\sup_{\gamma}\left\|\frac{\partial\rho'_{j,2}(\gamma,\hat{h})}{\partial\gamma}\Omega_j^{-1}\frac{\partial\rho_{j,1}(\gamma,\hat{h})}{\partial\gamma'}\right\|\right] &< \infty \\
E\left[\sup_{\gamma}\left\|(\rho'_{j,2}(\gamma,\hat{h})\Omega_j^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial\rho'_{j,1}(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right\|\right] &< \infty \\
E\left[\sup_{\gamma}\left\|(\rho'_{j,1}(\gamma,\hat{h})\Omega_j^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial\rho'_{j,2}(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right\|\right] &< \infty
\end{aligned}
$$

Since $\frac{\partial vec(\frac{\partial\rho'_{1,1}(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}$ depends on Z_1 only and therefore

$$
\begin{aligned}
E & \left[(\rho'_{1,2}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right] \\
= & \; E\left[E\left(\rho'_{1,2}(\gamma_0,\hat{h})|Z_1\right)\Omega_1^{-1}\otimes I_{p+k+1}\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right] \\
= & \; 0.
\end{aligned}
$$

And similarly,

$$
E\left[(\rho'_{1,1}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial \rho'_{1,2}(\gamma_0,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right] = 0.
$$

It follows from the USLLN and Lemma 3 that, as $n \longrightarrow \infty$,

$$
\begin{aligned}
& \frac{\partial^2 Q_{n,s}(\tilde{\gamma},\hat{h})}{\partial\gamma\partial\gamma'} \\
\xrightarrow{a.s.} \; & E\left[\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial\gamma}\Omega_1^{-1}\frac{\partial \rho_{1,2}(\gamma_0,\hat{h})}{\partial\gamma'} + (\rho'_{1,2}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right] \\
+ \; & E\left[\frac{\partial \rho'_{1,2}(\gamma_0,\hat{h})}{\partial\gamma}\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0,\hat{h})}{\partial\gamma'} + (\rho'_{1,1}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+1})\frac{\partial vec(\frac{\partial \rho'_{1,2}(\gamma_0,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right].
\end{aligned}
$$

Since

$$
\begin{aligned}
E\left[\frac{\partial\rho'_{1,1}(\gamma_0,\hat{h})}{\partial\gamma}\Omega_1^{-1}\frac{\partial\rho_{1,2}(\gamma_0,\hat{h})}{\partial\gamma'}\right] &= E\left[\frac{\partial\rho'_{1,2}(\gamma_0,\hat{h})}{\partial\gamma}\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma_0,\hat{h})}{\partial\gamma'}\right] \\
&= E\left[\frac{\partial\rho'_{1}(\gamma_0,\hat{h})}{\partial\gamma}\Omega_1^{-1}\frac{\partial\rho_{1}(\gamma_0,\hat{h})}{\partial\gamma'}\right] \\
&= K,
\end{aligned}
$$

therefore

$$
\frac{1}{n}\frac{\partial^2 Q_{n,s}(\tilde{\gamma},\hat{h})}{\partial\gamma\partial\gamma'} \xrightarrow{a.s.} 2K \tag{5.6}
$$

$$
\frac{1}{2n}\frac{\partial^2 Q_{n,s}(\tilde{\gamma},\hat{h})}{\partial\gamma\partial\gamma'} \xrightarrow{a.s.} K. \tag{5.7}
$$

Furthermore by Assumption 2.5, $\frac{\partial Q_{n,s}(\gamma_0,h)}{\partial\gamma}$ is continuously differentiable w.r.t. h and hence for sufficiently large n has the first Taylor expansion about h_0:

$$
\begin{aligned}
\frac{\partial Q_{n,s}(\gamma_0,\hat{h})}{\partial\gamma} = \quad & \textstyle\sum_{j=1}^{n} \left[\frac{\partial\rho'_{j,1}(\gamma_0,h_0)}{\partial\gamma}\Omega_j^{-1}\rho_{j,2}(\gamma_0,h_0) + \frac{\partial\rho'_{j,2}(\gamma_0,h_0)}{\partial\gamma}\Omega_j^{-1}\rho_{j,1}(\gamma_0,h_0)\right] \\
+ \quad & \frac{\partial^2\tilde{Q}_{n,s}(\gamma_0,\tilde{h})}{\partial\gamma\partial h'}(\hat{h}-h_0)
\end{aligned} \tag{5.8}
$$

where,

$$\frac{\partial^2 \tilde{Q}_{n,s}(\gamma_0, \tilde{h})}{\partial \gamma \partial h'}$$

$$= \sum_{j=1}^{n} \left[\frac{\partial \rho'_{j,1}(\gamma_0, \tilde{h})}{\partial \gamma} \Omega_j^{-1} \frac{\partial \rho_{j,2}(\gamma_0, \tilde{h})}{\partial h'} + (\rho'_{j,2}(\gamma_0, \tilde{h}) \Omega_j^{-1} \otimes I_{p+k+1}) \frac{\partial vec(\frac{\partial \rho'_{j,1}(\gamma_0, \tilde{h})}{\partial \gamma})}{\partial h'} \right]$$

$$+ \sum_{j=1}^{n} \left[\frac{\partial \rho'_{j,2}(\gamma_0, \tilde{h})}{\partial \gamma} \Omega_j^{-1} \frac{\partial \rho_{j,1}(\gamma_0, \tilde{h})}{\partial h'} + (\rho'_{j,1}(\gamma_0, \tilde{h}) \Omega_j^{-1} \otimes I_{p+k+1}) \frac{\partial vec(\frac{\partial \rho'_{j,2}(\gamma_0, \tilde{h})}{\partial \gamma})}{\partial h'} \right],$$

$$\frac{\partial \rho_{j,1}(\gamma_0, \tilde{h})}{\partial h'} = \frac{1}{S} \sum_{s=1}^{S} \frac{\tilde{x}_{js} p(\alpha_0 + \beta'_0 xjs)}{l(x_{js})} . \frac{\partial f_U(x_{js} - \tilde{H} Z_j; \theta)}{\partial u'} (Z_j \otimes I_q)',$$

$$\frac{\partial vec(\frac{\partial \rho'_{j,1}(\gamma_0, \tilde{h})}{\partial \theta})}{\partial h'} = \frac{1}{S} \sum_{s=1}^{S} \frac{\tilde{x}_{js} p(\alpha_0 + \beta'_0 xjs)}{l(x_{js})} . \frac{\partial^2 f_U(x_{js} - \tilde{H} Z_j; \theta)}{\partial \theta \partial u'} (Z_j \otimes I_q)',$$

$$\frac{\partial vec(\frac{\partial \rho'_{j,1}(\gamma_0, \tilde{h})}{\partial \alpha})}{\partial h'} = \frac{1}{S} \sum_{s=1}^{S} \frac{\tilde{x}_{js}}{l(x_{js})} . \frac{\partial p(\alpha_0 + \beta'_0 xjs)}{\partial \alpha} . \frac{\partial f_U(x_{js} - \tilde{H} Z_j; \theta)}{\partial u'} (Z_j \otimes I_q)',$$

$$\frac{\partial vec(\frac{\partial \rho'_{j,1}(\gamma_0, \tilde{h})}{\partial \beta})}{\partial h'} = \frac{1}{S} \sum_{s=1}^{S} \frac{\tilde{x}^2_{js}}{l(x_{js})} . \frac{\partial p(\alpha_0 + \beta'_0 xjs)}{\partial \beta} . \frac{\partial f_U(x_{js} - \tilde{H} Z_j; \theta)}{\partial u'} (Z_j \otimes I_q)',$$

and $\tilde{h} = vec\tilde{H}$ satisfies $\|\tilde{h} - h_0\| \le \|\hat{h} - h_0\|$.

By (2.5), $\hat{H} - H_0 = \left[\sum_{j=1}^{n} (W_j - H_0 Z_j) Z'_j\right] \left(\sum_{j=1}^{n} Z_j Z'_j\right)^{-1}$ which can be written as

$$\hat{h} - h_0 = vec(\hat{H} - H_0) = (\sum_{j=1}^{n} Z_j Z'_j \otimes I_q)^{-1} \sum_{j=1}^{n} Z_j \otimes (W_j - H_0 Z_j) \tag{5.9}$$

Hence (5.8) can be rewritten as

$$\frac{\partial Q_{n,s}(\gamma_0, \hat{h})}{\partial \gamma} = 2D_{n,s} \sum_{j=1}^{n} T_{j,S}, \tag{5.10}$$

where

$$D_{n,s} = \left(I_{p+k+1}, \frac{1}{2} \frac{\partial^2 \tilde{Q}_{n,s}(\gamma_0, \tilde{h})}{\partial \gamma \partial h'} \left(\sum_{j=1}^{n} Z_j Z_j' \otimes I_q \right)^{-1} \right)$$

and

$$T_{j,s} = \frac{1}{2} \left(\begin{array}{c} \frac{\partial \rho_{j,1}'(\gamma_0,h_0)}{\partial \gamma} \Omega_j^{-1} \rho_{j,2}(\gamma_0, h_0) + \frac{\partial \rho_{j,2}'(\gamma_0,h_0)}{\partial \gamma} \Omega_j^{-1} \rho_{j,1}(\gamma_0, h_0) \\ 2Z_j \otimes (W_j - H_0 z_j) \end{array} \right).$$

By the Strong Law of Large Numbers,

$$n \left(\sum_{j=1}^{n} Z_j Z_j' \otimes I_q \right)^{-1} \xrightarrow{a.s.} (EZ_1 Z_1' \otimes I_q)^{-1}, \quad n \longrightarrow \infty. \tag{5.11}$$

Similarly to (5.6) by Assumption 2.5 we can show that

$$\frac{1}{n} \frac{\partial^2 Q_{n,s}(\gamma_0, \tilde{h})}{\partial \gamma \partial h'} \xrightarrow{a.s.} 2E \left[\frac{\partial \rho_1'(\gamma_0, h_0)}{\partial \gamma} \Omega_1^{-1} \frac{\partial \rho_1(\gamma_0, h_0)}{\partial h'} \right], \quad n \longrightarrow \infty. \tag{5.12}$$

Hence by (5.11) and (5.12)

$$D_{n,s} \xrightarrow{a.s.} \left(I_{p+k+1}, E \left[\frac{\partial \rho_1'(\gamma_0, h_0)}{\partial \gamma} \Omega_1^{-1} \frac{\partial \rho_1(\gamma_0, h_0)}{\partial h'} \right] (EZ_1 Z_1' \otimes I_q)^{-1} \right) = D \tag{5.13}$$

Further by the Central Limit Theorem (CLT) we have

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} T_{j,S} \xrightarrow{d} N(0, \tau_S) \quad n \longrightarrow \infty, \tag{5.14}$$

where, $\tau_S = E(T_{1,S}T'_{1,S}) = \begin{pmatrix} \tau_{S,11} & \tau'_{S,21} \\ \tau_{S,21} & \tau_{S,22} \end{pmatrix}$ where,

$$\begin{aligned}
\tau_{S,11} = & \tfrac{1}{2} \quad E\left[\frac{\partial\rho'_{1,1}(\gamma_0,h_0)}{\partial\gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho'_{1,2}(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma_0,h_0)}{\partial\gamma'}\right] \\
& + \frac{1}{2}E\left[\frac{\partial\rho'_{1,1}(\gamma_0,h_0)}{\partial\gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho'_{1,1}(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial\rho_{1,2}(\gamma_0,h_0)}{\partial\gamma'}\right] \\
= & \tfrac{1}{2} \quad E\left[\frac{\partial\rho'_{1,1}(\gamma_0,h_0)}{\partial\gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho'_{1,2}(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma_0,h_0)}{\partial\gamma'}\right] \\
& + \frac{1}{2}E\left[\frac{\partial\rho'_{1,1}(\gamma_0,h_0)}{\partial\gamma}\Omega_1^{-1}E[\rho_{1,2}(\gamma_0,h_0)\rho'_{1,1}(\gamma_0,h_0)|Z]\Omega_1^{-1}\frac{\partial\rho_{1,2}(\gamma_0,h_0)}{\partial\gamma'}\right] \\
= & \tfrac{1}{2} \quad E\left[\frac{\partial\rho'_{1,1}(\gamma_0,h_0)}{\partial\gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho'_{1,2}(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma_0,h_0)}{\partial\gamma'}\right] + \frac{1}{2}K,
\end{aligned}$$

$$\begin{aligned}
\tau_{S,21} = & \frac{1}{2} \quad E \quad \left[(Z_1 \otimes (W_1 - H_0Z_1))\rho'_{1,2}(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma_0,h_0)}{\partial\gamma'}\right] \\
& + \frac{1}{2}E\left[(Z_1 \otimes (W_1 - H_0Z_1))\rho'_{1,1}(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial\rho_{1,2}(\gamma_0,h_0)}{\partial\gamma'}\right] \\
= & \frac{1}{2} \quad E \quad \left[(Z_1 \otimes (W_1 - H_0Z_1))\left(E[\rho'_{1,2}(\gamma_0,h_0)|Z]\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma_0,h_0)}{\partial\gamma'}\right)\right] \\
& + \frac{1}{2}E\left[(Z_1 \otimes (W_1 - H_0Z_1))\left(E[\rho'_{1,1}(\gamma_0,h_0)|Z]\Omega_1^{-1}\frac{\partial\rho_{1,2}(\gamma_0,h_0)}{\partial\gamma'}\right)\right] \\
= & 0 \quad ,
\end{aligned}$$

and $\tau_{S,22} = E\left[(Z_1 \otimes (W_1 - H_0Z_1))(Z_1 \otimes (W_1 - H_0Z_1))'\right]$.

Therefore by Slutsky's Theorem, (5.10), (5.13), and (5.14) we have

$$\frac{1}{2\sqrt{n}}\frac{\partial Q_{n,s}(\gamma_0,\hat{h})}{\partial\gamma} \xrightarrow{d} N(0, D\tau_S D') \tag{5.15}$$

Finally from (5.4), (5.7), and (5.15) we have

$$\sqrt{n}(\hat{\gamma}_{n,s}-\gamma_0) \xrightarrow{d} N(0, K^{-1}D\tau_S D'K^{-1}), \quad n \longrightarrow \infty.$$

5.3 Proof of Theorem 3.1.

To prove the consistency, we use the Lemma 1 (USLLN). The idea is to show that $E\left[\sup_\Gamma |\rho'_{i,1}(\gamma,h_0){\Omega_i}^{-1}\rho_{i,2}(\gamma,h_0)|\right] < \infty$. For sufficiently large n, $Q_{n,s}(\gamma,\hat{h})$ has the first order Taylor expansion about $h_0 = vecH_0$,

$$\begin{aligned} Q_{n,s}(\gamma,\hat{h}) &= \sum_{i=1}^{n}\rho'_{i,1}(\gamma,h_0)\Omega^{-1}\rho_{i,2}(\gamma,h_0) \qquad (5.16)\\ &+ \sum_{i=1}^{n}\left[\rho'_{i,1}(\gamma,\tilde{h})\Omega^{-1}\frac{\partial\rho_{i,2}(\gamma,\tilde{h})}{\partial h'} + \rho'_{i,2}(\gamma,\tilde{h})\Omega^{-1}\frac{\partial\rho_{i,1}(\gamma,\tilde{h})}{\partial h'}\right](\hat{h}-h_0) \end{aligned}$$

and $\tilde{h} = vec\tilde{H}$ satisfies $\|\tilde{h}-h_0\| \le \|\hat{h}-h_0\|$. where in 5.16,

$$\rho_{i,1}(\gamma,h_0) = Y_iT_i - m_1(H_0Z_i,\gamma),$$

$$\rho_{i,2}(\gamma,h_0) = Y_iT_i - m_2(H_0Z_i,\gamma),$$

$$\rho_{i,1}(\gamma,\tilde{h}) = Y_iT_i - m_1(\tilde{H}Z_i,\gamma),$$

$$\rho_{i,2}(\gamma,\tilde{h}) = Y_iT_i - m_2(\tilde{H}Z_i,\gamma),$$

$$\frac{\partial \rho_{i,1}(\gamma,\tilde{h})}{\partial h'} = \frac{1}{S}\sum_{s=1}^{S}\frac{\tilde{x}_{is} r(\beta' x_{is})}{l(x_{is})}.\frac{\partial f_U(x_{is}-\tilde{H}Z_i;\theta)}{\partial u'}(Z_i\otimes I_q)',$$

and

$$\frac{\partial \rho_{i,2}(\gamma,\tilde{h})}{\partial h'} = \frac{1}{S}\sum_{s=S+1}^{2S}\frac{\tilde{x}_{is} r(\beta' x_{is})}{l(x_{is})}.\frac{\partial f_U(x_{is}-\tilde{H}Z_i;\theta)}{\partial u'}(Z_i\otimes I_q)'.$$

By Assumptions 3.1, 3.2 and the dominated convergence theorem,

$$\begin{aligned}
E & \quad [\sup_{\Gamma}\|\rho_{1,1}(\gamma,h_0)\| | Z_1, Y_1, W_1] \\
\leq & \quad E||Y_1|| + E||Y_1W_1|| + E||Y_1^2|| \\
+ & \quad \frac{1}{S}\sum_{s=1}^{S} E\left[\sup_{\Gamma}\left[(\|x_{1s}\|+1)(r_1(\beta' x_{1s})+r_2(\beta' x_{1s}))\right]\frac{f_U(x_{1s}-H_0Z_1;\theta)}{l(x_{1s})}|Z_1,Y_1,W_1\right] \\
\leq & \quad E||Y_1|| + E||Y_1W_1|| + E||Y_1^2|| \\
+ & \quad \frac{1}{S}\sum_{s=1}^{S} E\left[\sup_{\Gamma}(\|x_{1s}\|+1)\left[-\sum_{j=1}^{J-1}G(\alpha_j-\beta' x_{1s})+J\right]\frac{f_U(x_{1s}-H_0Z_1;\theta)}{l(x_{1s})}|Z_1,Y_1,W_1\right] \\
+ & \quad \frac{1}{S}\sum_{s=1}^{S} E\left[\sup_{\Gamma}\left[-\sum_{j=1}^{J-1}(2j+1)G(\alpha_j-\beta' x_{1s})+J^2\right]\frac{f_U(x_{1s}-H_0Z_1;\theta)}{l(x_{1s})}|Z_1,Y_1,W_1\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq E||Y_1|| + E||Y_1W_1|| + E||Y_1^2|| \\
&+ \frac{1}{S}\sum_{s=1}^{S} E\left[\sup_{\Gamma}(\|x_{1s}\|+1)\left[-\sum_{j=1}^{J-1} G(\alpha_j - \beta' x_{1s}) + J\right]\frac{f_U(x_{1s} - H_0 Z_1;\theta)}{l(x_{1s})}|Z_1, Y_1, W_1\right] \\
&+ \frac{1}{S}\sum_{s=1}^{S} E\left[\sup_{\Gamma}\left[-\sum_{j=1}^{J-1}(2j+1) G(\alpha_j - \beta' x_{1s}) + J^2\right]\frac{f_U(x_{1s} - H_0 Z_1;\theta)}{l(x_{1s})}|Z_1, Y_1, W_1\right] \\
&\leq ||Y_1|| + ||Y_1W_1|| + ||Y_1^2|| \\
&+ \int \sup_{\Gamma}(\|x\|+1)\left[-\sum_{j=1}^{J-1} G(\alpha_j - \beta' x) + J\right] f_U(x - H_0 Z_1;\theta)dx \\
&+ \int \sup_{\Gamma}\left[-\sum_{j=1}^{J-1}(2j+1) G(\alpha_j - \beta' x) + J^2\right] f_U(x - H_0 Z_1;\theta)dx \\
&\leq ||Y_1|| + ||Y_1W_1|| + ||Y_1^2|| + J\int(\|x\|+1)\eta_1(x, Z_1)dx + J^2\int \eta_1(x, Z_1)dx
\end{aligned}
$$

Similarly;

$$
\begin{aligned}
E[\sup_{\Gamma}\|\rho_{1,2}(\gamma, h_0)\||Z_1, Y_1, W_1] \leq ||Y_1|| &+ ||Y_1W_1|| + ||Y_1^2|| + J\int(\|x\|+1)\eta_1(x, Z_1)dx \\
&+ J^2\int \eta_1(x, Z_1)dx
\end{aligned}
$$

Furthermore, since $\rho_{1,1}(\gamma)$ and $\rho_{1,2}(\gamma)$ are conditionally independent given (Z_1, Y_1, W_1), using Cauchy-Schwarz inequality and Assumption 3.2, we have

$$
\begin{aligned}
E & \left[\sup_{\Gamma} \|\rho'_{1,1}(\gamma, h_0)\Omega_1{}^{-1}\rho_{1,2}(\gamma, h_0)\|\right] \\
\leq & \; E\left[\|\Omega_1{}^{-1}\| E(\sup_{\Gamma} \|\rho'_{1,1}(\gamma, h_0)\| | Z_1, Y_1, W_1) E(\sup_{\Gamma} \|\rho_{1,2}(\gamma, h_0)\| | Z_1, Y_1, W_1)\right] \\
\leq & \; 3E\|\Omega_1{}^{-1}\| \left(||Y_1|| + ||Y_1 W_1|| + ||Y_1^2||\right)^2 + 3J^2 E\|\Omega_1{}^{-1}\| \left(\int (\|x\| + 1)\eta_1(x, Z_1)dx\right)^2 \\
+ & \; 3J^4 E\|\Omega_1{}^{-1}\| \left(\int \eta_1(x, Z_1)dx\right)^2 \\
\leq & \; \infty
\end{aligned}
$$

Therefore by USLLN,

$$
\sup_{\Gamma} \left\| \frac{1}{n} \sum_{j=1}^{n} \rho'_{j,1}(\gamma, h_0)\Omega_j{}^{-1}\rho_{j,2}(\gamma, h_0) - E\left[\rho'_{j,1}(\gamma, h_0)\Omega_j{}^{-1}\rho_{j,2}(\gamma, h_0)\right] \right\| \xrightarrow{a.s.} 0, \quad (5.17)
$$

where

$$
\begin{aligned}
& E\left[\rho'_{1,1}(\gamma, h_0)\Omega_1{}^{-1}\rho_{1,2}(\gamma, h_0)\right] \\
= & \; E\left[E(\rho'_{1,1}(\gamma, h_0)|Z_1, Y_1, W_1)\Omega_1{}^{-1}E(\rho_{1,2}(\gamma, h_0)|Z_1, Y_1, W_1)\right] \\
= & \; E\left[\rho'_1(\gamma, h_0)\Omega_1{}^{-1}\rho_1(\gamma, h_0)\right] \\
= & \; Q(\gamma, h_0)
\end{aligned}
$$

Similarly by Cauchy-Schwarz inequality and Assumption 3.2,

$$
\begin{aligned}
& E\left[\sup_{\Gamma}\left\|\rho_{1,1}'(\gamma,\tilde{h})\Omega_1^{-1}\frac{\partial\rho_{1,2}(\gamma,\tilde{h})}{\partial h'}\right\|\right]^2 \\
\leq \quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho_{1,1}'(\gamma,h)\|\sup_{\Gamma,h}\left\|\frac{\partial\rho_{1,2}(\gamma,h)}{\partial h'}\right\|\right]^2 \\
\leq \quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho_{1,1}'(\gamma,h)\|^2\right]\left[E\|\Omega_1^{-1}\|\|\sup_{\Gamma,h}\left\|\frac{\partial\rho_{1,2}(\gamma,h)}{\partial h'}\right\|^2\right] \\
\leq \quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho_{1,1}'(\gamma,h)\|^2\right] \\
& [E\|\Omega_1^{-1}\|\left(\int\sup_{\Gamma}(\|x\|+1)\left[-\sum_{j=1}^{J-1}G(\alpha_j-\beta'x)+J\right]\frac{\partial f_U(x-HZ_1;\theta)}{\partial h'}dx\right)^2 \\
& +E\|\Omega_1^{-1}\|\left(\int\sup_{\Gamma}\left[-\sum_{j=1}^{J-1}(2j+1)G(\alpha_j-\beta'x)+J^2\right]\frac{\partial f_U(x-HZ_1;\theta)}{\partial h'}dx\right)^2] \\
\leq \quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho_{1,1}'(\gamma,h)\|^2\right][E\|\Omega_1^{-1}\|\left(J\int\sup_{\Gamma}(\|x\|+1)\frac{\partial f_U(x-HZ_1;\theta)}{\partial h'}dx\right)^2 \\
& \qquad +E\|\Omega_1^{-1}\|\left(J^2\int\sup_{\Gamma}\frac{\partial f_U(x-HZ_1;\theta)}{\partial h'}dx\right)^2] \\
\leq \quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho_{1,1}'(\gamma,h)\|^2\right][J^2E\|\Omega_1^{-1}\|\left(\|Z_1\|\int(\|x\|+1)\eta_2(x,Z_1)dx\right)^2 \\
& \qquad +J^4E\|\Omega_1^{-1}\|\left(\|Z_1\|\int\eta_2(x,Z_1)dx\right)^2] \\
< \quad & \infty,
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left[\sup_{\Gamma}\left\|\rho'_{1,2}(\gamma,\tilde{h})\Omega_1^{-1}\frac{\partial\rho_{1,1}(\gamma,\tilde{h})}{\partial h'}\right\|\right]^2 \\
\leq\quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho'_{1,2}(\gamma,h)\|\sup_{\Gamma,h}\left\|\frac{\partial\rho_{1,1}(\gamma,h)}{\partial h'}\right\|\right]^2 \\
\leq\quad & \left[E\|\Omega_1^{-1}\|\sup_{\Gamma,h}\|\rho'_{1,2}(\gamma,h)\|^2\right][J^2E\|\Omega_1^{-1}\|\left(\|Z_1\|\int(\|x\|+1)\eta_2(x,Z_1)dx\right)^2 \\
& \qquad +J^4E\|\Omega_1^{-1}\|\left(\|Z_1\|\int\eta_2(x,Z_1)dx\right)^2] \\
<\quad & \infty.
\end{aligned}
$$

Again by Lemma 1 (USLLN) we have,

$$
\sup_{\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho'_{i,1}(\gamma,\tilde{h})\Omega_i^{-1}\frac{\partial\rho_{i,2}(\gamma,\tilde{h})}{\partial h'}\right\|\xrightarrow{a.s.}0,
$$

$$
\sup_{\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho'_{i,2}(\gamma,\tilde{h})\Omega_i^{-1}\frac{\partial\rho_{i,1}(\gamma,\tilde{h})}{\partial h'}\right\|\xrightarrow{a.s.}0.
$$

Therefore,

$$\begin{aligned}
& \sup_{\Gamma} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[\rho'_{i,1}(\gamma,\tilde{h})\Omega_i^{-1} \frac{\partial \rho_{i,2}(\gamma,\tilde{h})}{\partial h'} + \rho'_{j,2}(\gamma,\tilde{h})\Omega_i^{-1} \frac{\partial \rho_{i,1}(\gamma,\tilde{h})}{\partial h'} \right] (\hat{h} - h_0) \right\| \\
\leq \quad & \sup_{\Gamma} \left\| \frac{1}{n} \sum_{j=1}^{n} \rho'_{i,1}(\gamma,\tilde{h})\Omega_i^{-1} \frac{\partial \rho_{i,2}(\gamma,\tilde{h})}{\partial h'} \right\| \left\| (\hat{h} - h_0) \right\| \\
+ \quad & \sup_{\Gamma} \left\| \frac{1}{n} \sum_{j=1}^{n} \rho'_{i,2}(\gamma,\tilde{h})\Omega_i^{-1} \frac{\partial \rho_{i,1}(\gamma,\tilde{h})}{\partial h'} \right\| \left\| (\hat{h} - h_0) \right\| \\
\xrightarrow{a.s.} \quad & 0 \qquad (5.18)
\end{aligned}$$

It follows from (5.16) - (5.18) that

$$\sup_{\Gamma} \left\| \frac{1}{n} Q_{n,s}(\gamma,\hat{h}) - Q(\gamma,h_0) \right\| \xrightarrow{a.s.} 0$$

where, $Q(\gamma,h_0) = E\left[\rho'_1(\gamma,h_0){\Omega_1}^{-1}\rho_1(\gamma,h_0)\right]$.

Now for the next step, we need to show that $Q(\gamma,h_0)$ obtains a unique minimum at $\gamma_0 \in \Gamma$. Since $E(\rho'_1(\gamma_0,h_0)|Z_1) = 0$ and $\rho_1(\gamma,h_0) - \rho_1(\gamma_0,h_0)$ depends on Z_1 only, we have

$$\begin{aligned}
& E\left[\rho'_1(\gamma_0,h_0)\Omega^{-1}(\rho_1(\gamma,h_0) - \rho_1(\gamma_0,h_0))\right] \\
= \quad & E\left[E(\rho'_1(\gamma_0,h_0)|Z_1)\Omega^{-1}(\rho_1(\gamma,h_0) - \rho_1(\gamma_0,h_0))\right] = 0, \qquad (5.19)
\end{aligned}$$

which implies

$$
\begin{aligned}
& Q(\gamma, h_0) \\
= \;& E\left[\left(\rho_1'(\gamma, h_0) - \rho_1'(\gamma_0, h_0) + \rho_1'(\gamma_0, h_0)\right)\Omega_1^{-1}\left(\rho_1(\gamma, h_0) - \rho_1(\gamma_0, h_0) + \rho_1(\gamma_0, h_0)\right)\right] \\
= \;& Q(\gamma_0, h_0) + E\left[\left(\rho_1(\gamma, h_0) - \rho_1(\gamma_0, h_0)\right)'\Omega_1^{-1}\left(\rho_1(\gamma, h_0) - \rho_1(\gamma_0, h_0)\right)\right]. \qquad (5.20)
\end{aligned}
$$

Therefore $Q(\gamma, h_0) \geq Q(\gamma_0, h_0)$ and the equality holds if and only if $\gamma = \gamma_0$, by Assumption 3.3 Subsequently all conditions of Lemma 2 are satisfied, so we have $\hat{\gamma}_{n,s} \xrightarrow{a.s.} \gamma_0$, as $n \longrightarrow \infty$.

5.4 Proof of Theorem 3.2.

By Assumption 3.4, the first derivative $\frac{\partial Q_{n,s}(\hat{\gamma}_{n,S}, \hat{h})}{\partial\gamma}$ exists and has the first order Taylor expansion in an open neighbourhood $\Gamma_0 \subset \Gamma$ of γ_0;

$$
\frac{\partial Q_{n,s}(\hat{\gamma}_{n,S}, \hat{h})}{\partial\gamma} = \frac{\partial Q_{n,s}(\gamma_0, \hat{h})}{\partial\gamma} + \frac{\partial^2 Q_{n,s}(\tilde{\gamma}, \hat{h})}{\partial\gamma\partial\gamma'}(\hat{\gamma}_{n,s} - \gamma_0) = 0 \qquad (5.21)
$$

where $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma}_{n,s} - \gamma_0\|$. The first and second derivative of $Q_{n,s}$ is given by

$$
\frac{\partial Q_{n,s}(\gamma, \hat{h})}{\partial\gamma} = \sum_{i=1}^{n}\left[\frac{\partial\rho_{i,1}'(\gamma, \hat{h})}{\partial\gamma}\Omega_i^{-1}\rho_{i,2}(\gamma, \hat{h}) + \frac{\partial\rho_{i,2}'(\gamma, \hat{h})}{\partial\gamma}\Omega_i^{-1}\rho_{i,1}(\gamma, \hat{h})\right] \qquad (5.22)
$$

and

$$
\begin{aligned}
&\frac{\partial^2 Q_{n,s}(\gamma,\hat{h})}{\partial\gamma\partial\gamma'} \\
= \quad & \sum_{i=1}^{n}\left[\frac{\partial\rho_{i,1}'(\gamma,\hat{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho_{i,2}(\gamma,\hat{h})}{\partial\gamma'}+(\rho_{i,2}'(\gamma,\hat{h})\Omega_i^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial\rho_{j,1}'(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right] \\
+ \quad & \sum_{i=1}^{n}\left[\frac{\partial\rho_{i,2}'(\gamma,\hat{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho_{i,1}(\gamma,\hat{h})}{\partial\gamma'}+(\rho_{i,1}'(\gamma,\hat{h})\Omega_i^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial\rho_{i,2}'(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right].
\end{aligned}
$$

Assumptions 3.1-3.4 imply that

$$
\begin{aligned}
E\left[\sup_{\gamma}\left\|\frac{\partial\rho_{i,1}'(\gamma,\hat{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho_{i,2}(\gamma,\hat{h})}{\partial\gamma'}\right\|\right] &< \infty \\
E\left[\sup_{\gamma}\left\|\frac{\partial\rho_{i,2}'(\gamma,\hat{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho_{i,1}(\gamma,\hat{h})}{\partial\gamma'}\right\|\right] &< \infty \\
E\left[\sup_{\gamma}\left\|(\rho_{i,2}'(\gamma,\hat{h})\Omega_i^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial\rho_{i,1}'(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right\|\right] &< \infty \\
E\left[\sup_{\gamma}\left\|(\rho_{i,1}'(\gamma,\hat{h})\Omega_i^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial\rho_{i,2}'(\gamma,\hat{h})}{\partial\gamma})}{\partial\gamma'}\right\|\right] &< \infty
\end{aligned}
$$

Since $\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma,\hat{h})}{\partial \gamma})}{\partial \gamma'}$ depends on Z_1 only and therefore

$$\begin{aligned}
& E\left[(\rho'_{2,1}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial \gamma})}{\partial \gamma'}\right] \\
= \quad & E\left[E\left(\rho'_{2,1}(\gamma_0,\hat{h})|Z_1\right)\Omega_1^{-1}\otimes I_{p+k+J-1}\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial \gamma})}{\partial \gamma'}\right] \\
= \quad & 0.
\end{aligned}$$

And similarly,

$$E\left[(\rho'_{1,1}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial \rho'_{1,2}(\gamma_0,\hat{h})}{\partial \gamma})}{\partial \gamma'}\right] = 0.$$

It follows from the USLLN and Lemma 3 that

$$\begin{aligned}
& \frac{\partial^2 Q_{n,s}(\tilde{\gamma},\hat{h})}{\partial \gamma \partial \gamma'} \\
& \xrightarrow{a.s.} E\left[\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_{1,2}(\gamma_0,\hat{h})}{\partial \gamma'} + (\rho'_{1,2}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial \rho'_{1,1}(\gamma_0,\hat{h})}{\partial \gamma})}{\partial \gamma'}\right] \\
& \quad + E\left[\frac{\partial \rho'_{1,2}(\gamma_0,\hat{h})}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0,\hat{h})}{\partial \gamma'} + (\rho'_{1,1}(\gamma_0,\hat{h})\Omega_1^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial \rho'_{1,2}(\gamma_0,\hat{h})}{\partial \gamma})}{\partial \gamma'}\right].
\end{aligned}$$

Since

$$
\begin{aligned}
& E\left[\frac{\partial \rho'_{1,1}(\gamma_0, \hat{h})}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_{1,2}(\gamma_0, \hat{h})}{\partial \gamma'}\right] \\
= \quad & E\left[\frac{\partial \rho'_{1,2}(\gamma_0, \hat{h})}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0, \hat{h})}{\partial \gamma'}\right] \\
= \quad & E\left[\frac{\partial \rho'_{1}(\gamma_0, \hat{h})}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_{1}(\gamma_0, \hat{h})}{\partial \gamma'}\right] \\
= \quad & K,
\end{aligned}
$$

therefore

$$
\frac{1}{n}\frac{\partial^2 Q_{n,s}(\tilde{\gamma}, \hat{h})}{\partial \gamma \partial \gamma'} \xrightarrow{a.s.} 2K \tag{5.23}
$$

$$
\frac{1}{2n}\frac{\partial^2 Q_{n,s}(\tilde{\gamma}, \hat{h})}{\partial \gamma \partial \gamma'} \xrightarrow{a.s.} K. \tag{5.24}
$$

Furthermore by Assumption 3.5, $\frac{\partial Q_{n,s}(\gamma_0, h)}{\partial \gamma}$ is continuously differentiable w.r.t. h and hence for sufficiently large n has the first Taylor expansion about h_0:

$$
\begin{aligned}
\frac{\partial Q_{n,s}(\gamma_0, \hat{h})}{\partial \gamma} \quad = \quad & \sum_{i=1}^{n}\left[\frac{\partial \rho'_{i,1}(\gamma_0, h_0)}{\partial \gamma}\Omega_i^{-1}\rho_{i,2}(\gamma_0, h_0) + \frac{\partial \rho'_{i,2}(\gamma_0, h_0)}{\partial \gamma}\Omega_i^{-1}\rho_{i,1}(\gamma_0, h_0)\right] \\
+ \quad & \frac{\partial^2 \tilde{Q}_{n,s}(\gamma_0, \tilde{h})}{\partial \gamma \partial h'}(\hat{h} - h_0)
\end{aligned} \tag{5.25}
$$

where,

$$\frac{\partial^2 \tilde{Q}_{n,s}(\gamma_0,\tilde{h})}{\partial\gamma\partial h'}$$

$$= \sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,1}(\gamma_0,\tilde{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho_{i,2}(\gamma_0,\tilde{h})}{\partial h'} + (\rho'_{i,2}(\gamma_0,\tilde{h})\Omega_i^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial\rho'_{i,1}(\gamma_0,\tilde{h})}{\partial\gamma})}{\partial h'}\right]$$

$$+ \sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,2}(\gamma_0,\tilde{h})}{\partial\gamma}\Omega_i^{-1}\frac{\partial\rho_{i,1}(\gamma_0,\tilde{h})}{\partial h'} + (\rho'_{i,1}(\gamma_0,\tilde{h})\Omega_i^{-1}\otimes I_{p+k+J-1})\frac{\partial vec(\frac{\partial\rho'_{i,2}(\gamma_0,\tilde{h})}{\partial\gamma})}{\partial h'}\right],$$

$$\begin{aligned}
\frac{\partial\rho_{i,1}(\gamma,\tilde{h})}{\partial h'} &= \frac{1}{S}\sum_{s=1}^{S}\frac{\tilde{x}_{is}r(\beta' x_{is})}{l(x_{is})}.\frac{\partial f_U(x_{is}-\tilde{H}Z_i;\theta)}{\partial u'}(Z_i\otimes I_q)', \\
\frac{\partial vec(\frac{\partial\rho'_{i,1}(\gamma_0,\tilde{h})}{\partial\theta})}{\partial h'} &= \frac{1}{S}\sum_{s=1}^{S}\frac{\sum_{j=1}^{J-1}\tilde{x}_{is}r(\beta' x_{is})}{l(x_{is})}.\frac{\partial^2 f_U(x_{is}-\tilde{H}Z_i;\theta)}{\partial\theta\partial u'}(Z_i\otimes I_q)',
\end{aligned}$$

$$\begin{aligned}
\frac{\partial vec(\frac{\partial\rho'_{i,1}(\gamma_0,\tilde{h})}{\partial\alpha_j})}{\partial h'} &= \frac{1}{S}\sum_{s=1}^{S}\frac{\begin{pmatrix}1 & 0 & x_{is}\\ 0 & 2j+1 & 0\end{pmatrix}}{l(x_{is})}.\frac{\partial G(\alpha_j-\beta' x_{is};\theta)}{\partial\alpha_j} \\
&\quad .\frac{\partial f_U(x_{is}-\tilde{H}Z_i;\theta)}{\partial u'}(Z_i\otimes I_q)', \qquad j=1,\ldots,J-1 \\
\frac{\partial vec(\frac{\partial\rho'_{i,1}(\gamma_0,\tilde{h})}{\partial\beta})}{\partial h'} &= \frac{1}{S}\sum_{s=1}^{S}\frac{\sum_{j=1}^{J-1}x_{is}\begin{pmatrix}1 & 0 & x_{is}\\ 0 & 2j+1 & 0\end{pmatrix}}{l(x_{is})}.\frac{\partial G(\alpha_j-\beta' x_{is};\theta)}{\partial\beta} \\
&\quad .\frac{\partial f_U(x_{is}-\tilde{H}Z_i;\theta)}{\partial u'}(Z_i\otimes I_q)' \qquad (5.26)
\end{aligned}$$

and $\tilde{h} = vec\tilde{H}$ satisfies $\|\tilde{h} - h_0\| \leq \|\hat{h} - h_0\|$.

By (3.7), $\hat{H} - H_0 = [\sum_{i=1}^n (W_i - H_0 Z_i) Z_i'] \left(\sum_{i=1}^n Z_i Z_i'\right)^{-1}$ which can be written as

$$\hat{h} - h_0 = vec(\hat{H} - H_0) = (\sum_{i=1}^n Z_i Z_i' \otimes I_q)^{-1} \sum_{i=1}^n Z_i \otimes (W_i - H_0 Z_i) \tag{5.27}$$

Hence (5.25) can be rewritten as

$$\frac{\partial Q_{n,s}(\gamma_0, \hat{h})}{\partial \gamma} = 2D_{n,s} \sum_{i=1}^n T_{i,S}, \tag{5.28}$$

where

$$D_{n,s} = \left(I_{p+k+J-1}, \frac{1}{2} \frac{\partial^2 \tilde{Q}_{n,s}(\gamma_0, \tilde{h})}{\partial \gamma \partial h'} \left(\sum_{i=1}^n Z_i Z_i' \otimes I_q \right)^{-1} \right)$$

and

$$T_{i,s} = \frac{1}{2} \begin{pmatrix} \frac{\partial \rho_{i,1}'(\gamma_0, h_0)}{\partial \gamma} \Omega_i^{-1} \rho_{i,2}(\gamma_0, h_0) + \frac{\partial \rho_{i,2}'(\gamma_0, h_0)}{\partial \gamma} \Omega_i^{-1} \rho_{i,1}(\gamma_0, h_0) \\ 2Z_i \otimes (W_i - H_0 Z_i) \end{pmatrix}.$$

By the Strong Law of Large Numbers,

$$n \left(\sum_{i=1}^n Z_i Z_i' \otimes I_q \right)^{-1} \xrightarrow{a.s.} (EZ_1 Z_1' \otimes I_q)^{-1}, \quad n \longrightarrow \infty. \tag{5.29}$$

Similarly to (5.24) by Assumption 3.6 we can show that

$$\frac{1}{n} \frac{\partial^2 Q_{n,s}(\gamma_0, \tilde{h})}{\partial \gamma \partial h'} \xrightarrow{a.s.} 2E \left[\frac{\partial \rho_1'(\gamma_0, h_0)}{\partial \gamma} \Omega_1^{-1} \frac{\partial \rho_1(\gamma_0, h_0)}{\partial h'} \right], \quad n \longrightarrow \infty. \tag{5.30}$$

Hence by (5.29) and (5.30)

$$D_{n,s} \xrightarrow{a.s.} \left(I_{p+k+J-1}, E\left[\frac{\partial \rho_1'(\gamma_0,h_0)}{\partial \gamma}\Omega_1^{-1}\frac{\partial \rho_1(\gamma_0,h_0)}{\partial h'}\right](EZ_1Z_1' \otimes I_q)^{-1}\right) = D \quad (5.31)$$

Further by the central limit theorem we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_{i,S} \xrightarrow{d} N(0,\tau_S) \quad (5.32)$$

where, $\tau_S = E(T_{1,S}T_{1,S}') = \begin{pmatrix} \tau_{S,11} & \tau_{S,21}' \\ \tau_{S,21} & \tau_{S,22} \end{pmatrix}$ where,

$$\begin{aligned}
\tau_{S,11} = {} & \tfrac{1}{2} \quad E\left[\frac{\partial \rho_{1,1}'(\gamma_0,h_0)}{\partial \gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho_{1,2}'(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0,h_0)}{\partial \gamma'}\right] \\
& + \frac{1}{2}E\left[\frac{\partial \rho_{1,1}'(\gamma_0,h_0)}{\partial \gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho_{1,1}'(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial \rho_{1,2}(\gamma_0,h_0)}{\partial \gamma'}\right] \\
= {} & \tfrac{1}{2} \quad E\left[\frac{\partial \rho_{1,1}'(\gamma_0,h_0)}{\partial \gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho_{1,2}'(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0,h_0)}{\partial \gamma'}\right] \\
& + \frac{1}{2}E\left[\frac{\partial \rho_{1,1}'(\gamma_0,h_0)}{\partial \gamma}\Omega_1^{-1}E[\rho_{1,2}(\gamma_0,h_0)\rho_{1,1}'(\gamma_0,h_0)|Z]\Omega_1^{-1}\frac{\partial \rho_{1,2}(\gamma_0,h_0)}{\partial \gamma'}\right] \\
= {} & \tfrac{1}{2} \quad E\left[\frac{\partial \rho_{1,1}'(\gamma_0,h_0)}{\partial \gamma}\Omega_1^{-1}\rho_{1,2}(\gamma_0,h_0)\rho_{1,2}'(\gamma_0,h_0)\Omega_1^{-1}\frac{\partial \rho_{1,1}(\gamma_0,h_0)}{\partial \gamma'}\right] + \frac{1}{2}K,
\end{aligned}$$

$$
\begin{aligned}
\tau_{S,21} = \frac{1}{2} \quad & E \quad \left[(Z_1 \otimes (W_1 - H_0 Z_1)) \rho'_{1,2}(\gamma_0, h_0) \Omega_1^{-1} \frac{\partial \rho_{1,1}(\gamma_0, h_0)}{\partial \gamma'} \right] \\
& + \quad \frac{1}{2} E \left[(Z_1 \otimes (W_1 - H_0 Z_1)) \rho'_{1,1}(\gamma_0, h_0) \Omega_1^{-1} \frac{\partial \rho_{1,2}(\gamma_0, h_0)}{\partial \gamma'} \right] \\
= \frac{1}{2} \quad & E \quad \left[(Z_1 \otimes (W_1 - H_0 Z_1)) \, E[\rho'_{1,2}(\gamma_0, h_0) | Z] \Omega_1^{-1} \frac{\partial \rho_{1,1}(\gamma_0, h_0)}{\partial \gamma'} \right] \\
& + \quad \frac{1}{2} E \left[(Z_1 \otimes (W_1 - H_0 Z_1)) \, E[\rho'_{1,1}(\gamma_0, h_0) | Z] \Omega_1^{-1} \frac{\partial \rho_{1,2}(\gamma_0, h_0)}{\partial \gamma'} \right] \\
= 0 \quad & ,
\end{aligned}
$$

and $\tau_{S,22} = E\left[(Z_1 \otimes (W_1 - H_0 Z_1)) \, (Z_1 \otimes (W_1 - H_0 Z_1))' \right]$.

Therefore by Slutsky's Theorem, (5.28), (5.31), and (5.32) we have

$$
\frac{1}{2\sqrt{n}} \frac{\partial Q_{n,s}(\gamma_0, \hat{h})}{\partial \gamma} \xrightarrow{d} N(0, D\tau_S D') \tag{5.33}
$$

Finally from (5.21), (5.24), and (5.33) we have

$$
\sqrt{n}(\hat{\gamma}_{n,s} - \gamma_0) \xrightarrow{d} N(0, K^{-1} D \tau_S D' K^{-1}), \quad n \longrightarrow \infty.
$$

Bibliography

Abarin, T. and L. Wang (2006). Comparison of gmm with second-order least squares estimation in nonlinear models. *Far East Journal of Theoretical Statistics 20*(2), 179–196. (Cited on pages 13 and 53.)

Abarin, T. and L. Wang (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics 64*, 475–493. (Cited on pages 4, 12, 17, 41, 57 and 77.)

Agresti, A. (2003). *Categorical Data Analysis.* John Wiley and Sons, Inc. (Cited on pages 6 and 45.)

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica 41*, 997–1016. (Cited on page 80.)

Anderson, J. A. and R. R. Phillips (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics 30*, 22–31. (Cited on page 43.)

Bartholomew, D. and M. Knott (1999). *Latent Variable Models and Factor Analysis.* Hodder Arnold, London. (Cited on page 46.)

Bollen, K., D. Thomas, and L. Wang (2008). A consistent instrumental variable estimator for errors in covariates in ordinal probit regression models. *Working Paper, Department of Statistics, University of Manitoba*. (Cited on page 9.)

Buzas, J. and L. Stefanski (1996). Instrumental variable estimation in generalized linear measurement error models. *Journal of American Statistical Association 91*, 999–1006. (Cited on pages 3, 4, 5, 10, 41 and 76.)

Cagnone, C., S. Mignani, and A. Gardini (2004). New developments of latent variable models with ordinal data. *Proceedings of the XLI Scientific meeting of the Italian Statistical Society, Padova, Cleup,*, 221–231. (Cited on page 46.)

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu (2006). *Measurement error in nonlinear models: A modern perspective* (2nd ed.). Chapman and Hall, London. (Cited on pages 1 and 3.)

Fuller, W. A. (1987). *Measurement error models.* New York: Wiely. (Cited on pages 2 and 3.)

Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine 13*, 1665–1677. (Cited on page 45.)

Hammer, S. M., D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, W. Henry, M. M. Lederman, J. P. Phair, M. Niu, M. S. Hirsch, and T. C. Merigan (1996). A trial comparison nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine 335*(15), 1081–1090. (Cited on page 37.)

Huang, Y. and C. Wang (2000). Cox regression with accurate covariates unascertainable: a nonparametric-correction. *J. Am. Stat. Assoc. 95*, 1209–1219. (Cited on page 37.)

Huang, Y. J. and C. Wang (2001). Consistent functional methods for logistic regression with errors in covariates. *J. Am. Stat. Assoc. 96*, 1469–1482. (Cited on pages 9, 18, 37 and 42.)

Huber, P., E. Ronchetti, and V. Feser (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society, Series B 66*, 893–908. (Cited on page 46.)

Jennrich, R. (1969). Asymptotic properties of nonlinear least squares estimators. *The Annals of Mathematical Statistics 40*, 633–643. (Cited on page 79.)

Joreskog, K. (1990). New developments in lisrel: analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity 24*, 387–404. (Cited on page 46.)

Joreskog, K. and I. Moustaki (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research 36*, 347–387. (Cited on page 46.)

Ma, Y. and A. A. Tsiatis (2006). Closed form semiparametric estimators for measurement error models. *Stat. Sin. 16*, 183–193. (Cited on page 3.)

Magnus, J. and H. Neudecker (1988). *Matrix Differential with Application in Statistics and Econometrics.* Wiley, New York. (Cited on page 17.)

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological) 42*(2), 109–142. (Cited on pages 44, 45 and 46.)

Monari, P., M. Bini, D. Piccolo, and L. Salmaso (2009). *Statistical Methods for the Evaluation of Educational Services and Quality of Products*. Springer-Verlag Berlin Heidelberg. (Cited on page 45.)

Moustaki, I. (2000). A latent variable model for ordinal data. *Applied psychological measurement 24*, 211–223. (Cited on page 46.)

Moustaki, I. and M. Knott (2000). Generalized latent trait models. *Psychometrika 65*, 391–411. (Cited on page 46.)

Muthen, B. O. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika 49*, 115–132. (Cited on page 46.)

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika 77*(1), 127–137. (Cited on page 3.)

Rabe-Hesketh, S., A. Pickles, and A. Skrondal (2003). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Satatistical Modeling 3*, 215–232. (Cited on page 9.)

Schilling, S. and R. Bock (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika 70*, 533–555. (Cited on page 46.)

Stefanski, L. and R. Carroll (1985). Covariate measurement error in logistic regression. *The Annals of Statistics 13*(4), 1335–1351. (Cited on page 9.)

Stefanski, L. and R. Carroll (1991). Deconvolution-based score tests in measurement error models. *The Annals of Statistics 19*(1), 249–259. (Cited on page 3.)

Stefanski, L. A. and J. S. Buzas (1995). Instrumental variable estimation in binary regression measurement error models. *J. Am. Stat. Assoc. 90*, 541–550. (Cited on pages 1 and 9.)

van der Linden, W. and R. Hambleton (1997). *Handbook of Modern Item Response Theory*. SpringerVerlag, New York. (Cited on page 46.)

Walker, S. and D. Duncan (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika 54*, 167–179. (Cited on pages 44 and 46.)

Wang, L. (2003). Estimation of nonlinear berkson-type measurement error models. *Statistica Sinica 13*, 1201–1210. (Cited on pages 4, 5, 10, 41 and 76.)

Wang, L. (2004). Estimation of nonlinear models with berkson measurement errors. *The Annals of Statistics 32*, 2559–2579. (Cited on pages 4 and 10.)

Wang, L. (2007). A unified approach to estimation of nonlinear mixed effects and berkson measurement error models. *The Canadian Journal of Statistics 35*, 233–248. (Cited on page 4.)

Wang, L. and C. Hsiao (1995). A simulated semiparametric estimation of nonlinear errors-in-variables models. *Working Paper. Department of Economics, University of Southern California, Los Angeles.*. (Cited on pages 4 and 12.)

Wang, L. and C. Hsiao (2011). Method of moments estimation and identifiability of nonlinear semiparametric errors-in-variables models. *Journal of Econometrics 165*, 30–44. (Cited on pages 4, 8 and 12.)

Wang, L. and A. Leblanc (2008). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics 60*(4), 883–900. (Cited on page 4.)

Williams, O. D. and J. E. Grizzle (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association 67*, 55–63. (Cited on page 46.)

Xu, K., Y. Ma, and L. Wang (2015). Instrumental assisted regression for errors in variables model with binary response. *Scandinavian Journal of Statistics 42*, 104–117. (Cited on pages 8, 9 and 37.)

Index