Second-order Least Squares Estimation in Regression Models with Application to Measurement Error Problems

By Taraneh Abarin

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

© by Taraneh Abarin, 2008

Abstract

This thesis studies the Second-order Least Squares (SLS) estimation method in regression models with and without measurement error. Applications of the methodology in general quasi-likelihood and variance function models, censored models, and linear and generalized linear models are examined and strong consistency and asymptotic normality are established. To overcome the numerical difficulties of minimizing an objective function that involves multiple integrals, a simulation-based SLS estimator is used and its asymptotic properties are studied. Finite sample performances of the estimators in all of the studied models are investigated through simulation studies.

Keywords and phrases: Nonlinear regression; Censored regression model; Generalized linear models; Measurement error; Consistency; Asymptotic normality; Least squares method; Method of moments, Heterogeneity; Instrumental variable; Simulation-based estimation.

Acknowledgements

I would like to gratefully express my appreciation to Dr. Liqun Wang, my supervisor, for his patience, guidance, and encouragement, during my research. This dissertation could not have been written without his support, challenges, and generous time. He and the advisory committee members, Drs. John Brewster, James Fu, and Gady Jacoby, patiently guided me through the research process, never accepting less than my best efforts. I thank them all. My thanks go also to the external examiner, Dr. Julie Zhou, for careful reading the dissertation and helpful suggestions.

Many thanks to the Faculty of Graduate Studies for the University of Manitoba Graduate Fellowship (UMGF) and Dr. Wang who supported me financially through his Natural Sciences and Engineering Research Council of Canada (NSERC) research grant.

Last but not the least, I am grateful to my family and friends, for their

patience, love and support. This dissertation is dedicated to my mother, who never stopped believing in, and loved me unconditionally. May perpetual light shine upon her soul.

Dedication

To my mother, who loved me unconditionally…

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Recently Wang (2003, 2004) proposed a Second-order Least Squares (SLS) estimator in nonlinear measurement error models. This estimation method, which is based on the first two conditional moments of the response variable given the observed predictor variables, extends the ordinary least squares estimation by including in the criterion function the distance of the squared response variable to its second conditional moment. He showed that under some regularity conditions the SLS estimator is consistent and asymptotically normally distributed. Furthermore, Wang and Leblanc (2007) compared the SLS estimator with the Ordinary Least Squares (OLS) estimator in general nonlinear models. They showed that the SLS estimator is asymptotically more efficient than the OLS estimator when the third moment of the random error is nonzero.

This thesis contains four major extensions and studies of the second-order least squares method. First, we extend the SLS method to a wider class of Quasi-likelihood and Variance Function (QVF) models. These types of models are based on the mean and the variance of the response given the explanatory variables and include many important models such as homoscedastic and heteroscedastic linear and nonlinear models, generalized linear models and logistic regression models. Applying the same methodology of Wang (2004), we prove strong consistency and asymptotic normality of the estimator under fairly general regularity conditions.

The second work is a comparison of the SLS with Generalized Method of Moment (GMM). Since both the SLS and GMM are based on the conditional moments, an interesting question is which one is more efficient.

GMM estimation, which has been used for more than two decades by econometricians, is an estimation technique which estimates unknown parameters by matching theoretical moments with the sample moments. In recent years, there has been growing literature on finite sample behavior of GMM estimators, see, e.g., Windmeijer (2005) and references therein. Although GMM estimators are consistent and asymptotically normally distributed under general regularity conditions (Hansen 1982), it has long been recognized that this asymptotic distribution may provide a poor approximation to the finite sample distribution of the estimators (Windmeijer 2005). Identification is another issue in the GMM method (e.g., Stock and Wright 2000, and Wright 2003). In particular, the number of moment conditions needs to be equal to or greater than the number of parameters, in order for them to be identifiable. This restriction creates problems when the parametric dimension increases. It seems that adding over-identifying restrictions (moment conditions) will increase precision; however, this is not always the case (Anderson and Sorenson 1996). More recently, for a linear model with heteroscedasticity, Koenker and Machado (1999) show that an effective sample size can be given by n/q_n^3, where q_n is the number of moment conditions used. This means that a very large sample size is required to justify conventional asymptotic inference. See also Huber (2004). An interesting question that arises naturally is how SLS estimators compare with GMM estimators. This question is partially addressed in our first paper (Abarin and Wang 2006).

In Abarin and Wang (2006), we compare the asymptotic covariance matrix of the efficient SLS and GMM when both use the same number of moment conditions and we show that the SLS estimator is asymptotically more efficient than the GMM estimator. Since a theoretical comparison of these two methods is extremely difficult when SLS and GMM use different number of moment conditions, we compare these two estimators through Monte Carlo simulation studies. The simulation studies show the superiority of SLS over GMM, even when it uses more equations.

The third major work is to apply the SLS to the censored linear models, which has been widely used in many fields. Wang and Leblanc (2007) studied a second-order least squares estimator for general nonlinear regression models. However, the framework used in Wang and Leblanc (2007) does not cover an important family of models including censored regression models.

Regression models with censored response variables arise frequently in econometrics, biostatistics, and many other areas. In economics, such a model was first used by Tobin (1958) to analyze household expenditures on durable goods. Other econometric applications of the Tobit model include Heckman and MaCurdy (1986), and Killingsworth and Heckman (1986). In the last few decades various methods have been introduced to estimate the parameters for various censored regression models. In particular, Amemiya (1973) and Wang (1998) investigated maximum likelihood and moment estimators when the regression error distribution is normal, while Powell (1984, 1986) proposed semi-parametric estimators when the error distribution is symmetric or satisfies certain quantile restrictions.

In practice, many variables of interests have asymmetric distributions, e.g., income and expenditures, insurance claims and premiums, survival or failure times. Regression with asymmetric error distributions has been considered by many authors, e.g., Williams (1997), Austin et al (2003), Marazzi and Yohai (2004), and Bianco et al (2005). Most estimators considered belong to a large family of so-called M-estimators which maximize or minimize certain criteria. In our second paper (Abarin and Wang 2008a), we apply the second-order least squares estimator in censored regression models. The paper gives regularity conditions under which the proposed estimator is strongly consistent and asymptotically normal for a wide class of error distributions. Since for some distribution functions for the error term, calculation of the conditional moments may be quite difficult, a simulation-based estimator is proposed to overcome this problem. Finally, we present the results of a simulation study of the estimator for two different weighting matrices (parametric and semi-parametric). In the simulation study we investigate the finite sample behavior of the estimators and compare them with the MLE.

The fourth work relates to SLS estimation in Measurement Error models. Since Wang (2003,2004) studied the SLS estimators in measurement error models, this provides a natural motivation to apply the estimation method in other measurement error models. So we start with linear models with measurement error.

It is well-known that not all parameters of a normal linear model with measurement error in explanatory variables are identifiable (Fuller 1987). One of the methods to overcome this problem is to use instrumental variables. Similar to the two-stage method that has been suggested by Wang and Hsiao (2007), we estimate the parameters using the SLS estimation in the second stage and we show that the parameter estimators are consistent and asymptotically normally distributed, when there is no assumption on the distribution of the error terms. Compared to the more convenient two-stage method which uses OLS estimator in the second stage, we prove that the two-stage SLS estimator is asymptotically more efficient. A simulation study shows that in a moderate sample size, the Root Mean Squared Error (RMSE) reduction can be about 40%.

To generalize the previous results, we study the application of SLS in Generalized Linear Models (GLM). Generalized Linear Models, as a very important subclass of QVF models, are widely used in biostatistics, epidemiology, and many other areas. However, the real data analysis using GLM often involves covariates that are not observed directly or are measured with error. See, e.g., Franks et al. (2004), Stimmer (2003), Kiechl et al. (2004), and Carroll et al. (2006). In such cases, statistical estimation and inference become very challenging. Several researchers, such as Stefanski and Carroll (1985), Aitkin (1996), and Rabe-Hesketh et al. (2003), have studied the maximum likelihood estimation of the GLM with measurement error. However, most of the proposed approaches rely on the normality assumption for the unobserved covariates and measurement error, though some other parametric distributions have been considered (Schafer 2001, Aitkin and Rocci 2002, Kukush and Schneeweiss 2005, Roy and Banerjee 2006). A computational difficulty with the likelihood approach is that the likelihood function involves multiple integrals which do not admit closed forms in general. Another approach is based on corrected score functions (Nakamura 1990, Stefanski and Carroll 1991, and Buzas and Stefanski 1996). This approach, however, produces only approximately consistent estimators (estimators that are consistent when the measurement error approaches to zero) and therefore is applicable when the measurement error is small.

General nonlinear models have been investigated by several authors using either replicate data (Li 1998, Thoresen and Laake 2003, Li and Hsiao

2004, and Schennach 2004) or instrumental variables (Wang and Hsiao 1995, 2007; and Schennach 2007), while non- or semi-parametric approaches have been considered (Schafer 2001, Taupin 2001). In addition, Wang (2003, 2004) studied nonlinear models with Berkson measurement error. However, most of these papers deal with models with homoscedastic regression errors. In our third paper (Abarin and Wang 2008b), we consider the generalized linear models which allow very general heteroscedastic regression errors. In particular, we study the method of moments estimation combined with instrumental variables. This approach does not require the parametric assumptions for the distributions of the unobserved covariates and of the measurement error, which are difficult to check in practice.

The dissertation is organized as follows. In Chapter 2 we focus on regression models without measurement error. In Section 2.1 we begin by introducing the QVF model and the SLS estimator. There we show that the SLS estimators are strongly consistent and asymptotically normally distributed under some general regularity conditions. We present the asymptotic covariance matrix of the most efficient SLS estimator in Section 2.2. The optimum weighting matrix plays an important role in the SLS estimator, so we find the explicit form of the matrix, presenting the asymptotic covariance matrix of the most efficient SLS estimator in homoscedastic linear regression models. In Section 2.3 we develop some computational methods to deal with heteroscedastisity. We end the section with illustrations of the SLS methodology using examples. In Section 2.4 we compare the asymptotic variance covariance matrix of the GMM and the SLS estimators, using the same number of equations. We finish the section by showing some simulation results that compare the two estimators when GMM uses more equations.

In Section 2.5 we revisit the SLS in a censored model and we present the theories that show the estimator is consistent and asymptotically normally distributed under some regularity conditions. The SLS estimator is based on the first two conditional moments of the response variable given the predictor variables. In practice, it is not always straightforward to calculate the closed forms of these moments. In this case, the objective function will involve multiple integrals which makes it difficult to minimize. To overcome this numerical difficulty, we propose a simulation-based SLS estimator and give its asymptotic properties. We carry out substantial Monte Carlo simulations to study finite sample behavior of the SLS estimators and compare them with the MLE.

In Chapter 3 regression models with measurement error are studied. In

Section 3.1 we apply SLS in a linear measurement error model and present a two-stage estimator which is strongly consistent and asymptotically normally distributed and we prove the superiority of this estimator over the estimator which uses OLS in the second stage. Section 3.2 focusses on generalized linear measurement error models. We introduce the model and give some examples to motivate our estimation method. Then we introduce the method of moments estimators and derive their consistency and asymptotic normality properties. We construct simulation-based estimators for the situations where the closed forms of the moments are not available. Then we present simulation studies to illustrate the finite sample performance of the proposed estimators.

In Chapter 4 we present some issues related to the research. One of the most important parts of computation of the SLS is the optimization method. In Section 4.1 we discuss various computational methods, together with their advantages and disadvantages. The section continues with some discussion about the optimum weighting matrix, as an important part of the efficient SLS estimator. Since there are substantial simulation studies in the dissertation, in Section 4.2 we present some theoretical criteria for simulation accuracy, and show how we applied those criteria in this dissertation. We end the dissertation with a summary and discuss possible extensions for future work in Chapter 5. The proofs of the theorems are given in the Appendices.

Chapter 2

Second-order Least Squares Estimation in Regression Models

2.1 Quasi-likelihood Variance Function Models

2.1.1 Model and Estimation

Consider the model

$$Y = f(X;\beta) + \varepsilon, \tag{2.1}$$

where $Y \in I\!R$ is the response variable, $X \in I\!R^k$ is the predictor variable, and ε is the random error satisfying $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = \sigma^2 g^2(X;\beta,\theta)$. Here $\beta \in I\!R^p$, $\theta \in I\!R^q$ and $\sigma^2 \in I\!R$ are the unknown parameters. In addition, we assume that Y and ε have finite fourth moments.

Under the assumptions for model (2.1), the first two conditional moments of Y given X are

$$
\begin{aligned}
E(Y|X) &= f(X;\beta), \\
E(Y^2|X) &= \sigma^2 g^2(X;\beta,\theta) + f^2(X;\beta),
\end{aligned}
$$

and therefore, $V(Y|X) = \sigma^2 g^2(X;\beta,\theta)$. This is a typical Quasi-likelihood and Variance Function (QVF) model (Carroll et al. 2006).

This model includes many important special cases such as:

- Homoscedastic linear and nonlinear regression, with $g(X;\beta,\theta) = 1$. For linear regression, $f(X;\beta) = X'\beta$.

- Generalized linear models, including Poisson and gamma regression, with $g(X;\beta,\theta) = f^{\theta}(X;\beta)$ for some parameter θ. For Poisson regression, $\theta = 1/2$, while $\theta = 1$ for gamma regression.

- Logistic regression, where $f(X;\beta) = 1/(1+\exp(-X'\beta))$ and $g(X;\beta,\theta) = f(X;\beta)(1-f(X;\beta))$, $\sigma^2 = 1$, and there is no parameter θ.

Throughout the section we denote the parameter vector as $\gamma = (\beta', \theta', \sigma^2)'$ and the parameter space as $\Gamma = \Omega \times \Theta \times \Sigma \subset I\!R^{p+q+1}$. The true parameter value of model (2.1) is denoted by $\gamma_0 = (\beta_0', \theta_0', \sigma_0^2)' \in \Gamma$.

Suppose $(Y_i, X_i')', i = 1, 2, ..., n$ is an *i.i.d.* random sample. Then the second-order least squares (SLS) estimator $\hat{\gamma}_{SLS}$ for γ is defined as the measurable function that minimizes

$$Q_n(\gamma) = \sum_{i=1}^{n} \rho_i'(\gamma) W_i \rho_i(\gamma), \tag{2.2}$$

where $\rho_i(\gamma) = (Y_i - f(X_i;\beta), Y_i^2 - \sigma^2 g^2(X_i;\beta,\theta) - f^2(X_i;\beta))'$ and $W_i = W(X_i)$ is a 2×2 nonnegative definite matrix which may depend on X_i.

2.1.2 Consistency and Asymptotic Normality

For the consistency and asymptotic normality of $\hat{\gamma}_{SLS}$ we make the following assumptions, where μ denotes the Lebesgue measure and $\|\cdot\|$ denotes the Euclidean norm in the real space.

Assumption 1. *$f(x;\beta)$ and $g(x;\beta,\theta)$ are measurable functions of x for every $\beta \in \Omega$ and $\theta \in \Theta$.*

Assumption 2. $E\|W(X)\| (\sup_\Omega f^4(X;\beta) + 1) < \infty$ *and*
$E\|W(X)\| \left(\sup_{\Omega\times\Theta} g^4(X;\beta,\theta) + 1\right) < \infty$.

Assumption 3. *The parameter space $\Gamma \subset \mathbb{R}^{p+q+1}$ is compact.*

Assumption 4. *For any* $\gamma \in \Gamma$, $E[\rho(\gamma) - \rho(\gamma_0)]'W(X)[\rho(\gamma) - \rho(\gamma_0)] = 0$ *if and only if* $\gamma = \gamma_0$.

Assumption 5. *There exist positive functions* $h_1(x)$ *and* $h_2(x)$*, such that the first two partial derivatives of* $f(x;\beta)$ *with respect to* β *are bounded by* $h_1(x)$*, and the first two partial derivatives of* $g(x;\beta,\theta)$ *with respect to* β *and* θ *are bounded by* $h_2(x)$*. Moreover,* $E\left\|W(X)\right\|(h_1^4(X) + 1) < \infty$*, and* $E\left\|W(X)\right\|(h_2^4(X) + 1) < \infty$

Assumption 6. *The matrix* $B = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma'}\right]$ *is nonsingular.*

In the above, assumptions A1 – A3 ensure that the objective function $Q_n(\gamma)$ is uniformly convergent to $Q(\gamma)$. Additionally, assumption A4 means that the objective function $Q_n(\gamma)$ for large n attains a unique minimum at the true parameter value γ_0. Assumption A5 guarantees uniform convergence of the second derivative of $Q_n(\gamma)$. Finally, assumption A6 is necessary for the existence of the variance of the SLS $\hat{\gamma}_n$.

We first restate some existing results which are used in the proofs, which can be found in Appendix . For this purpose, let $Z = (Z_1, Z_2, ..., Z_n)$ be an *i.i.d.* random sample and $\psi \in \Psi$ a vector of unknown parameters, where the parameter space $\Psi \subset I\!R^d$ is compact. Further, suppose $Q_n(Z, \psi)$ is a measurable function for each $\psi \in \Psi$ and is continuous in $\psi \in \Psi$ for μ–almost all Z. Then Lemmas 3 and 4 of Amemiya (1973) can be stated as follows.

Lemma 2.1.1. *If, as $n \to \infty$, $Q_n(Z, \psi)$ converges a.s. to a non stochastic function $Q(\psi)$ uniformly for all $\psi \in \Psi$ and $Q(\psi)$ attains a unique minimum at $\psi_0 \in \Psi$, then $\hat{\psi}_n = \text{argmin}_{\psi \in \Psi} Q_n(Z, \psi) \xrightarrow{a.s.} \psi_0$.*

Lemma 2.1.2. *If, as $n \to \infty$, $Q_n(Z, \psi)$ converges a.s. to a non stochastic function $Q(\psi)$ uniformly for all ψ in an open neighborhood of ψ_0, then for any sequence of estimators $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$ it holds $Q_n(Z, \hat{\psi}_n) \xrightarrow{a.s.} Q(\psi_0)$.*

Theorem 2.1.3. *Under Assumptions 1 – 4, the SLS estimator $\hat{\gamma}_{SLS} \xrightarrow{a.s.} \gamma_0$, as $n \to \infty$.*

Theorem 2.1.4. *Under Assumptions 1 – 6, as $n \to \infty$, $\sqrt{n}(\hat{\gamma}_{SLS} - \gamma_0) \xrightarrow{L} N(\mathbf{0}, B^{-1}AB^{-1})$, where*

$$A = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(X) \rho(\gamma_0) \rho'(\gamma_0) W(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma'}\right], \tag{2.3}$$

and

$$B = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma'}\right]. \tag{2.4}$$

2.2 Asymptotic Covariance Matrix of the Most Efficient SLS Estimator

Asymptotic covariance of $\hat{\gamma}_{SLS}$ depends on the weighting matrix W. A natural question is how to choose W to obtain the most efficient estimator. To answer this question, we first note that, since $\partial\rho'(\gamma_0)/\partial\gamma$ does not depend on Y, matrix A in (2.3) can be written as
$A = E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}WFW\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]$, where $F = F(X) = E[\rho(\gamma_0)\rho'(\gamma_0)|X]$. Then, analog to the weighted (nonlinear) least squares estimation, we have

$$B^{-1}AB^{-1} \geq \left(E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}F^{-1}\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]\right)^{-1} \tag{2.5}$$

(in the sense that the difference of the left-hand and right-hand sides is nonnegative definite), and the lower bound is attained for $W = F^{-1}$ in both A and B . In the following we provide a proof of this fact.

Theorem 2.2.1. *Denote* $C = \frac{\partial\rho(\gamma)}{\partial\gamma'}$ *and* $D = \rho(\gamma)\rho'(\gamma)$. *Then the asymptotic covariance matrix of the most efficient SLS estimation is* $E^{-1}(C'F^{-1}C)$, *where* $F = E(D|X)$.

Proof. First, it is easy to see that $B = E(C'WC)$ and $A = E(C'WDWC) = E(C'WFWC)$, because C and W do not depend on Y. Further, let $\alpha =$

$E^{-1}(C'WFWC)E(C'WC)$. Then we have

$$\begin{aligned} & E(C - FWC\alpha)'F^{-1}(C - FWC\alpha) \\ = \;& E(C'F^{-1}C) - E(C'WC)\alpha - \alpha' E(C'WC) + \alpha' E(C'WFWC)\alpha \\ = \;& E(C'F^{-1}C) - E(C'WC)E^{-1}(C'WFWC)E(C'WC) \\ = \;& E(C'F^{-1}C) - BA^{-1}B \end{aligned}$$

which is nonnegative definite. It follows that $E^{-1}(C'F^{-1}C) \leq B^{-1}AB^{-1}$, with equality holds if $W = F^{-1}$ in both A and B. $\square$

In the following corollary, we find the optimal weighting matrix.

Corollary 2.2.2. *If* $detF \neq 0$*, then the optimal weighting matrix is given by*

$$F^{-1} = \frac{1}{detF} \begin{pmatrix} f_{22} & -f_{12} \\ -f_{12} & \sigma_0^2 g^2(X; \beta_0, \theta_0) \end{pmatrix} \tag{2.6}$$

where

$$\begin{aligned} f_{22} &= E(\varepsilon^4|X) + 4f(X;\beta_0)E(\varepsilon^3|X) \\ &+ 4\sigma_0^2 f^2(X;\beta_0)g^2(X;\beta_0,\theta_0) - \sigma_0^4 g^4(X;\beta_0,\theta_0), \end{aligned}$$

and

$$f_{12} = E(\varepsilon^3|X) + 2\sigma_0^2 f(X;\beta_0)g^2(X;\beta_0,\theta_0).$$

Proof. First, by definition the elements of F are

$$f_{11} = E\left[(Y - f(X;\beta_0))^2 \,|X\right] = \sigma_0^2 g^2(X;\beta_0,\theta_0),$$

$$\begin{aligned} f_{22} &= E\left[\left(Y^2 - \sigma_0^2 g^2(X;\beta_0,\theta_0) - f^2(X;\beta_0)\right)^2 |X\right] \\ &= E(\varepsilon^4|X) + 4f(X;\beta_0)E(\varepsilon^3|X) \\ &+ 4\sigma_0^2 f^2(X;\beta_0)g^2(X;\beta_0,\theta_0) - \sigma_0^4 g^4(X;\beta_0,\theta_0), \end{aligned}$$

and

$$\begin{aligned} f_{12} &= E\left[(Y - f(X;\beta_0))\left(Y^2 - \sigma_0^2 g^2(X;\beta_0,\theta_0) - f^2(X;\beta_0)\right)|X\right] \\ &= E(\varepsilon^3|X) + 2\sigma_0^2 f(X;\beta_0)g^2(X;\beta_0,\theta_0). \end{aligned}$$

It is straightforward to calculate the inverse of F which is given by (2.6). □

In general $F(X)$ is unknown, and it must be estimated, before the $\hat{\gamma}_{SLS}$ using $W(X) = F(X)^{-1}$ is computed. This can be done using the following two-stage procedure. First, minimize $Q_n(\gamma)$ using the identity matrix $W(X) = I_2$ to obtain the first-stage estimator $\hat{\gamma}_n$. Secondly, estimate $F(X)$ using $\hat{\gamma}_n$ and then minimize $Q_n(\gamma)$ again with $W(X) = \hat{F}(X)^{-1}$ to obtain the two-stage estimator $\hat{\hat{\gamma}}_n$. We can use the residuals to estimate $E(\epsilon^3|X)$ and $E(\epsilon^4|X)$.

It is useful to have an explicit form of the asymptotic covariance matrix of the most efficient SLS estimator for a homoscedastic linear regression mode. The following Corollary shows it.

Corollary 2.2.3. *The asymptotic covariance matrix of the most efficient SLS estimator for a homoscedastic linear regression model is given by*

$$C = \begin{pmatrix} V\left(\hat{\beta}_{SLS}\right) & \frac{\mu_3}{\mu_4-\sigma_0^4} V\left(\hat{\sigma}^2_{SLS}\right) E^{-1}(XX')EX \\ \frac{\mu_3}{\mu_4-\sigma_0^4} V\left(\hat{\sigma}^2_{SLS}\right) EX'E^{-1}(XX') & V\left(\hat{\sigma}^2_{SLS}\right) \end{pmatrix}, \tag{2.7}$$

where

$$V\left(\hat{\beta}_{SLS}\right) = \left(\sigma_0^2 - \frac{\mu_3^2}{\mu_4 - \sigma_0^4}\right)\left(E(XX') - \frac{\mu_3^2}{\sigma_0^2(\mu_4 - \sigma_0^4)} EXEX'\right)^{-1},$$

$$V\left(\hat{\sigma}^2_{SLS}\right) = \frac{(\mu_4 - \sigma_0^4)\left(\sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2\right)}{\sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2 EX'E^{-1}(XX')EX},$$

and $\mu_3 = E(\varepsilon^3|X)$, and $\mu_4 = E(\varepsilon^4|X)$.

Proof. Consider a homoscedastic linear regression model

$$Y = f(X;\beta) + \varepsilon,$$

where $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = \sigma^2$. Wang and Leblanc (2007) combined (2.6) and the lower bound of (2.5), to find the asymptotic covariance matrix of the most efficient SLS estimator for a general nonlinear regression model with homoscedasticity. Assuming that $E(\varepsilon^3|X) = \mu_3$ and $E(\varepsilon^4|X) = \mu_4$, we apply the special case for $f(X;\beta) = X'\beta$ to their asymptotic covariance matrix to find (2.7). □

It should be mentioned that when X is fixed and ε is symmetric, then (2.7) reduces to

$$C = \begin{pmatrix} \sigma^2(XX')^{-1} & 0 \\ 0 & 2\sigma^4 \end{pmatrix},$$

which is equivalent to covariance matrix for ordinary least squared estimator for the parameters.

2.3 Computational Method

If we set $Y^* = \frac{Y}{g(X;\beta_0,\theta_0)}$, then

$$E(Y^*|X) = \frac{f(X;\beta)}{g(X;\beta_0,\theta_0)}, E(Y^{*2}|X) = \sigma^2 + \frac{f^2(X;\beta)}{g^2(X;\beta_0,\theta_0)}.$$

Similar to (2.2), we can define $\rho_i^*(\gamma)$ as $\left(\frac{Y_i - f(X_i;\beta)}{g(X;\beta_0,\theta_0)}, \frac{Y_i^2 - \sigma^2 g^2(X_i;\beta,\theta) - f^2(X_i;\beta)}{g^2(X;\beta_0,\theta_0)}\right)'$ which is equivalent to

$$\begin{pmatrix} \frac{1}{g(X;\beta_0,\theta_0)} & 0 \\ 0 & \frac{1}{g^2(X;\beta_0,\theta_0)} \end{pmatrix} \begin{pmatrix} Y - f(X_i;\beta) \\ Y^2 - \sigma^2 g^2(X_i;\beta,\theta) - f^2(X_i;\beta) \end{pmatrix}.$$

The optimal weighting matrix for the new setting is

$$W^*_{opt} = E\left(\rho^*(\gamma)\rho^{*'}(\gamma)|X\right)$$

$$= \begin{pmatrix} \frac{1}{g(X;\beta_0,\theta_0)} & 0 \\ 0 & \frac{1}{g^2(X;\beta_0,\theta_0)} \end{pmatrix} E\left(\rho(\gamma)\rho'(\gamma)|X\right) \begin{pmatrix} \frac{1}{g(X;\beta_0,\theta_0)} & 0 \\ 0 & \frac{1}{g^2(X;\beta_0,\theta_0)} \end{pmatrix}.$$

Therefore,

$$W^*_{opt} = \begin{pmatrix} \frac{1}{g(X;\beta_0,\theta_0)} & 0 \\ 0 & \frac{1}{g^2(X;\beta_0,\theta_0)} \end{pmatrix} W_{opt} \begin{pmatrix} \frac{1}{g(X;\beta_0,\theta_0)} & 0 \\ 0 & \frac{1}{g^2(X;\beta_0,\theta_0)} \end{pmatrix}.$$

Assuming $Q^*_n(\gamma)$ is the objective function for the new setting, simple calculation allows us to write it in terms of $Q_n(\gamma)$. We have

$$Q^*_n(\gamma) = \sum_{i=1}^{n} \rho^{*'}_i(\gamma) W^*_{iopt} \rho^*_i(\gamma) = \sum_{i=1}^{n} \rho'_i(\gamma) W_{iopt} \rho_i(\gamma) = Q_n(\gamma).$$

It shows that minimizing $Q^*_n(\gamma)$ is equivalent to minimizing $Q_n(\gamma)$. Therefore, for computational purpose we have two options:

(i). Using the transformation $Y^* = \frac{Y}{g(X;\beta_0,\theta_0)}$ to estimate the parameters. Since β_0 and θ_0 are unknown, we can estimate them in the first step using the identity matrix for the weighting matrix. To estimate $E(\varepsilon^3|X)$ and $E(\varepsilon^4|X)$ we can use the third and fourth mean of the residuals.

(ii). Estimate the parameters using the original data and estimate of the third and forth mean of the studentized residuals $(\frac{\epsilon}{g(X;\beta_0,\theta_0)})$.

2.3.1 Examples

Example 2.3.1. **Oxidation of benzene**

Prichard et al. (1977) describes an example concerning the initial rate of oxidation of benzene over a vanadium oxide catalyst at three reaction temperatures and several benzene and oxygen concentrations.

Carrol and Rupert (1988) suggested the following highly nonlinear model to fit the data set:

$$E(Y) = 1/[\beta_1 x_4 x_1^{-1} \exp(\beta_3 x_3) + \beta_2 x_2^{-1} \exp(\beta_4 x_3)]$$

In the model, Y is the initial rate of disappearance of benzene, x_1 and x_2 are the oxygen and benzene concentrations respectively.

They defined x_3 as $2000(1/T - 1/648)$ and x_4 as the number of moles of oxygen consumed per mole of benzene, where T is the absolute temperature in Kelvins. They plotted the absolute studentized residuals against the logarithm of the predicted values, and the logarithm of the absolute residuals against the logarithm of the predicted values (pages 43– 44 Carroll and Rupert (1988)), to show that the variability varies systematically as a function of the mean response.

Table 2.1 shows the Second-order Least Squares (SLS), as well as Un-

Table 2.1: Estimated parameters and standard errors for Oxidation of benzene data set

Parameter	β_1	β_2	β_3	β_4	σ
UWNLS	0.97	3.20	7.33	5.01	-
SE	0.1	0.20	1.14	0.64	
GLS	0.86	3.42	6.20	5.66	0.11
SE	0.04	0.13	0.63	0.41	-
SLS	0.82	3.53	5.99	5.86	0.10
SE	0.04	0.12	0.62	0.38	0.04

Weighted Nonlinear Least Squares (UWNLS) and Generalized Least Squares (GLS) parameter estimates and their estimated standard errors (SE) for comparison. Note that both GLS and SLS are computed for a generalized linear model in which the standard deviation is proportional to the mean.

Example 2.3.2. **Effects of Education on Fertility**

Wooldridge (2002) studies the effects of education on women's fertility in Botswana. He considers the number of living children as the response variable, and years of school (edu), a quadratic in age (age^2), binary indicators for ever married (evermarr), living in an urban area (urban), having electricity (electric), and owning a television (tv), as explanatory variables. Table 2.2 reports the results for the Generalized Least-Squares (GLS) and Second-order least-squares (SLS), in a Poisson regression model. The next table shows the parameter estimates with their standard errors for a Poisson log-linear model.

2.4 Comparison with GMM Estimation

Consider model (2.1), where ε satisfies $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = \sigma^2$ ($g^2(X;\beta,\theta) = 1$). Under this general nonlinear regression model, the first two conditional moments of Y given X are respectively $E(Y|X) = f(X;\beta)$ and $E(Y^2|X) = \sigma^2 + f^2(X;\beta)$, where $\gamma = (\beta', \sigma^2)'$. Suppose $(Y_i, X_i')'$, $i = 1, 2, ..., n$ is an $i.i.d.$ random sample. Following Wang and Leblanc (2007), the second-order least squares estimator $\hat{\gamma}_{SLS}$ for γ is defined as the measurable function that minimizes (2.2) where

$$\rho_i(\gamma) = \left(Y_i - f(X_i;\beta), Y_i^2 - f^2(X_i;\beta) - \sigma^2\right)'.$$

It is easy to see that this estimator is an extension of the ordinary least squares estimator by adding the distance of the squared response variable to

Table 2.2: Estimated parameters and standard errors for Fertility Education data set

Variable	GLS	SE	SLS	SE
edu	-0.0217	0.0025	-0.0233	0.0023
age	0.337	0.009	0.325	0.007
age^2	-0.0041	0.0001	-0.0040	0.0001
evermarr	0.315	0.021	0.246	0.019
urban	-0.086	0.019	-0.070	0.017
electric	-0.121	0.034	-0.053	0.023
tv	-0.145	0.041	-0.145	0.037
constant	-5.375	0.141	-5.009	0.117
σ	0.867	-	.733	0.130

its second conditional moment into the criterion function.

Wang and Leblanc (2007) proved that, under some regularity conditions, the SLS estimator is consistent and has an asymptotic normal distribution with the asymptotic covariance matrix given by $B^{-1}AB^{-1}$, where A and B are given in (2.3) and (2.4), respectively.

Moreover, in this special case of model (2.1),

$$\frac{\partial \rho'(\gamma)}{\partial \gamma} = -\begin{pmatrix} \frac{\partial f(X;\beta)}{\partial \beta} & 2f(X;\beta)\frac{\partial f(X;\beta)}{\partial \beta} \\ 0 & 1 \end{pmatrix}.$$

In Theorem 2.2.1 we proved that the best choice for the weighting matrix is $W = F^{-1}$, where $F = E(\rho(\gamma)\rho'(\gamma)|X)$, which gives the smallest variance covariance matrix

$$\left(E\left[\frac{\partial \rho'(\gamma)}{\partial \gamma} F^{-1} \frac{\partial \rho(\gamma)}{\partial \gamma'}\right]\right)^{-1}.$$

In the rest of this section, SLS always refers to the most efficient second-order least squares estimator using the weight $W = F^{-1}$.

Now we compare the SLS with GMM estimator when both use the same set of moment conditions. Given the *i.i.d.* random sample $(Y_i, X_i')', i = 1, 2, ..., n$, the GMM estimator using the first two conditional moments is defined as the measurable function which minimizes

$$G_n(\gamma) = \left(\sum_{i=1}^{n} \rho_i(\gamma)\right)' W_n \left(\sum_{i=1}^{n} \rho_i(\gamma)\right),$$

where $\rho_i(\gamma)$ is defined in (2.2) and W_n is a nonnegative definite weighting matrix. It can be shown (e.g., Mátyás (1999)) that under some regularity conditions, the GMM estimator has an asymptotic normal distribution and the asymptotic covariance matrix of the efficient GMM is given by

$$\left[E\left(\frac{\partial \rho'(\gamma)}{\partial \gamma}\right) V^{-1} E\left(\frac{\partial \rho(\gamma)}{\partial \gamma'}\right)\right]^{-1},$$

where $V = E[\rho(\gamma)\rho'(\gamma)] = E(F)$ is the optimum weighting matrix. The next theorem compares the asymptotic covariance matrices of the SLS and GMM estimators.

Theorem 2.4.1. *The SLS estimator is asymptotically more efficient than the GMM estimator using the first two moment conditions, i.e.* $E^{-1}(C'F^{-1}C) \leq [E(C')V^{-1}E(C)]^{-1}$.

Proof. The proof is similar to that of Theorem 2.2.1. Let $\alpha = V^{-1}E(C)$. Then the result follows from

$$\begin{aligned} E(C - F\alpha)'F^{-1}(C - F\alpha) &= E(C'F^{-1}C) - E(C')\alpha - \alpha'E(C) + \alpha'E(F)\alpha \\ &= E(C'F^{-1}C) - E(C')V^{-1}E(C) \geq 0. \end{aligned}$$

The proof is completed. $\square$

If $\hat{\beta}$ and $\hat{\sigma}^2$ denote the estimators of the regression and variance parameters respectively, then the above theorem implies that $V\left(\hat{\theta}_{GMM}\right) \geq V\left(\hat{\beta}_{SLS}\right)$ and $V\left(\hat{\sigma}^2_{GMM}\right) \geq V\left(\hat{\sigma}^2_{SLS}\right)$ asymptotically. Given these theoretical comparison results, an interesting question is GMM performs better than SLS if using more than two moment conditions. This question is examined in the next part.

2.4.1 Monte Carlo Simulation Studies

In order to study the finite sample behavior of the second-order least squares (SLS) estimation approach and the generalized method of moment (GMM) estimation, several simulation scenarios are considered.

We consider the following exponential, logistic and linear-exponential models, each with two or three parameters.

1. Exponential model with two parameters $Y = 10e^{\beta X} + \varepsilon$, where $X \sim U(0.1, 10)$, $\varepsilon \sim N(0, \sigma^2)$, and true parameter values are $\beta = -0.5$ and $\sigma^2 = 1$.

2. Exponential model with three parameters $Y = \beta_1 e^{\beta_2 X} + \varepsilon$, where $X \sim$

$U(0.1, 10)$, $\varepsilon \sim N(0, \sigma^2)$, and true parameters are $\beta_1 = 10$, $\beta_2 = -0.5$ and $\sigma^2 = 1$.

3. Logistic model with two parameters $Y = \frac{20}{1+\exp[-(X-\beta)/34]} + \varepsilon$, where $X \sim U(20, 100)$, $\varepsilon \sim N(0, \sigma^2)$, and true parameters are $\beta = 50$ and $\sigma^2 = 1$.

4. Logistic model with three parameters $Y = \frac{20}{1+\exp[-(X-\beta_1)/\beta_2]} + \varepsilon$, where $X \sim U(20, 80)$, $\varepsilon \sim N(0, \sigma^2)$, and true parameters are $\beta_1 = 50$, $\beta_2 = 34$ and $\sigma^2 = 1$.

5. Linear-exponential model with two parameters $Y = 5e^{\beta_1 X} + 10e^{\beta_2 X} + \varepsilon$, where $X \sim U(0.1, 10)$, $\varepsilon \sim N(0, 1)$, and true parameters are $\beta_1 = -3$ and $\beta_2 = -1$.

6. Linear-exponential model with three parameters $Y = \beta_3 e^{\beta_1 X} + 10e^{\beta_2 X} + \varepsilon$, where $X \sim U(0.1, 10)$, $\varepsilon \sim N(0, 1)$, and true parameters are $\beta_1 = -3$ and $\beta_2 = -1$ and $\beta_3 = 5$.

In each model we compare SLS with two versions of GMM estimator, one (GMM3) using the first three and another (GMM4) using the first four moment conditions. Both SLS and GMM estimators are computed in two steps. In the first step, identity weighting matrix is used to obtain initial parameter estimates. In the second step, first the optimal weight for the SLS is calculated according to formula (7) in Wang and Leblanc (2007), and optimal weight for GMM is calculated as $W_n = n^{-1} \sum_{i=1}^{n} \rho_i(\hat{\gamma})\rho_i(\hat{\gamma})'$. Then the final estimates are computed using the estimated weights.

As is frequently the case in nonlinear numerical optimization, convergence, numerical complaints, and other problems will be encountered. To avoid potential optimization problems involved in the iterative procedures, a direct grid search method is applied. In particular, $n_0 = 5000$ grid points per parameter are generated in each iteration. For each model, 1000 Monte Carlo repetitions are carried out for each of the sample sizes $n = 10, 20, 30, 50, 70, 100, 200, 300, 500$. The Monte Carlo means (SLS, GMM3, GMM4) and their root mean squared errors (RMSE) are computed. The numerical computation is done using the statistical computer language R for Windows on a PC with standard configuration.

Tables 2.3 – 2.17 report the results for $n = 100, 200, 300, 500$. We present the Monte Carlo means, accuracy, root mean squared errors of the estimators, and 95% confidence interval for the parameters. The accuracy for the estimators is calculated using $A = z_{\alpha/2}\sqrt{Var(\hat{\beta})/R}$, where $R = 1000$

is the Monte Carlo replication number, and $z_{\alpha/2}$ is the $1-\alpha/2$ quantile on the right tail of the distribution $N(0,1)$.

According to the results, SLS shows its asymptotic properties for sample size larger than 100. However, both GMM3 and GMM4 converge slower. Figure 2.1 suggests a downward bias in the GMM3 and GMM4 for σ^2 but not in SLS. The confidence intervals show that the true value of parameters fall in the intervals for SLS when sample size increases. However, the same results for GMM are not satisfying because of its finite sample bias. Moreover, the accuracy values show that SLS has more accurate mean values than GMM.

Figure 2.2 shows the RMSE of the estimates for β , versus sample size. We used these values to compare the efficiency of the estimators. Clearly SLS is more efficient than GMM3 and GMM4. This fact is more obvious, when we compare the corresponding confidence intervals (Confidence intervals for GMM are outside of the confidence intervals for SLS). As sample size increases we can see that RMSE decreases and the confidence intervals become smaller.

Although the overall results show that SLS has smaller RMSE than

GMM, in logistic model with three parameters (Tables 2.13 – 2.14), SLS estimator for β_2 and σ^2 has larger RMSE for small sample sizes than GMMs. As sample size increases, SLS starts to dominate both GMM3 and GMM4. Figure 2.3 shows this result. Table 2.15 shows the case ($\beta_1 = -3$) that SLS has larger RMSE than GMM, even for sample size equal to 500. It seems that in this case, SLS needs more sample size to perform better than GMMs.

Another point in the results is that GMM3 has smaller RMSE than GMM4. Since GMM4 uses more information than GMM3, we expect to have more precise estimators for GMM4 than GMM3. However this is not always true. In Koenker and Machado (1999), they imply that GMM with higher number of moment equations needs more sample sizes to justify conventional asymptotic inference.

Figure 2.1: Monte Carlo estimates of σ^2 in the Exponential model with two parameters

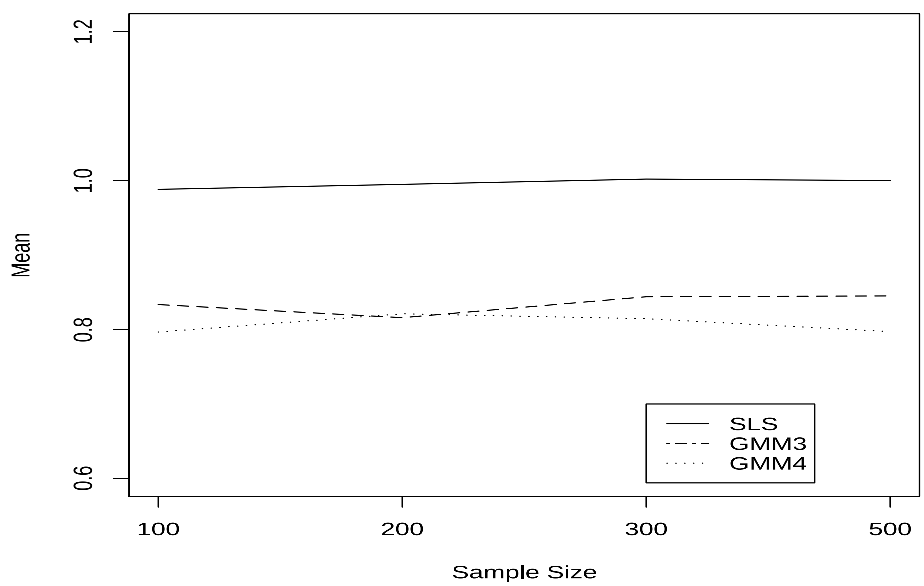


Figure 2.2: RMSE of Monte Carlo estimates of β for the Exponential model with two parameters

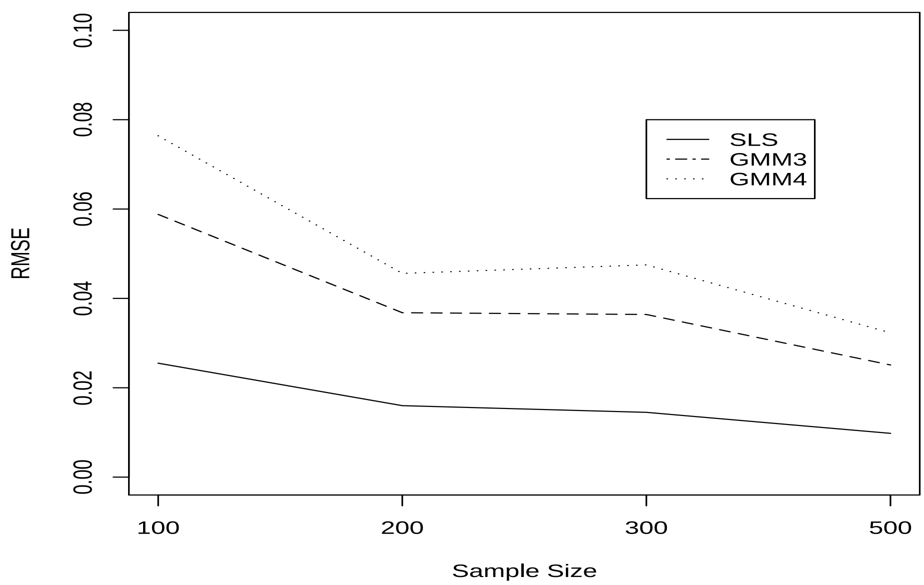


Figure 2.3: RMSE of Monte Carlo estimates of β_2 in Logistic model with three parameters

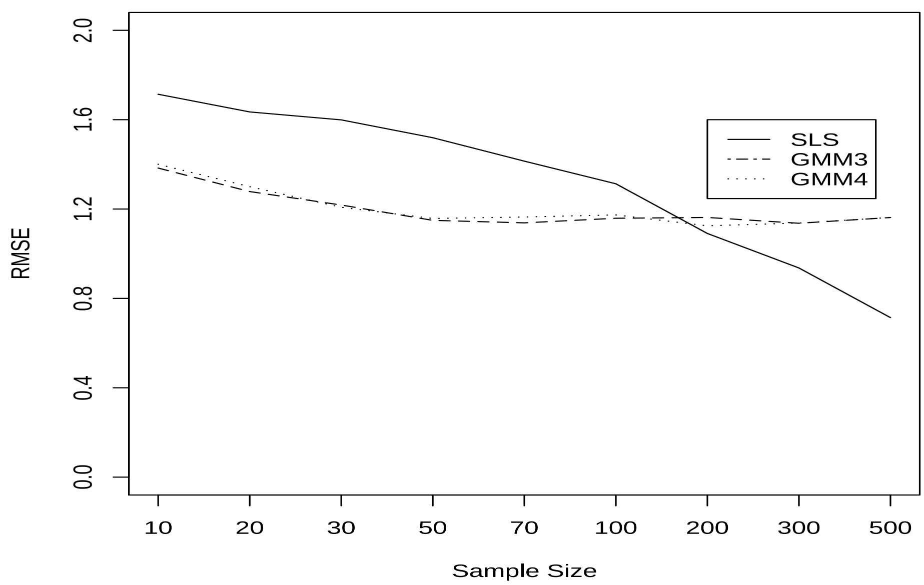


Table 2.3: Simulation results of the Exponential model with two parameters

	n = 100	n = 200	n = 300	n = 500
		$\beta = -0.5$		
SLS	-0.5031	-0.5001	-0.5000	-0.5004
RMSE	0.0255	0.0160	0.0145	0.0098
Accuracy	0.0016	0.0010	0.0009	0.0006
95% C.I.	-0.505,-0.502	-0.501,-0.499	-0.501,-0.499	-0.501,-0.500
GMM3	-0.4988	-0.4933	-0.4942	-0.4949
RMSE	0.0588	0.0368	0.0364	0.0251
Accuracy	0.0037	0.0022	0.0022	0.0015
95% C.I.	-0.502,-0.495	-0.496,-0.491	-0.496,-0.492	-0.496,-0.493
GMM4	-0.4986	-0.4925	-0.4918	-0.4914
RMSE	0.0764	0.0456	0.0475	0.0323
Accuracy	0.0047	0.0028	0.0029	0.0019
95% C.I.	-0.503,-0.494	-0.495,-0.490	-0.495,-0.489	-0.493,-0.489

Table 2.4: Simulation results of the Exponential model with two parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\sigma^2 = 1$		
SLS	0.9882	0.9949	1.0020	1.0008
RMSE	0.1427	0.1003	0.0851	0.0622
Accuracy	0.0088	0.0062	0.0053	0.0039
95% C.I.	0.979,0.997	0.989,1.001	0.997,1.007	0.996,1.004
GMM3	0.8335	0.8159	0.8440	0.8451
RMSE	0.4309	0.4399	0.4169	0.4371
Accuracy	0.0246	0.0248	0.0240	0.0254
95% C.I.	0.809,0.858	0.791,0.841	0.820,0.868	0.820,0.870
GMM4	0.7966	0.8211	0.8145	0.7972
RMSE	0.4567	0.4398	0.4449	0.4493
Accuracy	0.0254	0.0249	0.0251	0.0249
95% C.I.	0.771,0.822	0.796,0.846	0.789,0.840	0.772,0.822

Table 2.5: Simulation results of the Logistic model with two parameters

	n = 100	n = 200	n=300	n = 500
		$\beta = 50$		
SLS	50.0112	50.0315	50.0021	50.0101
RMSE	0.7832	0.5426	0.4357	0.3419
Accuracy	0.0486	0.0336	0.0270	0.0212
95% C.I.	49.963,50.060	49.998,50.065	49.975,50.029	49.989,50.031
GMM3	49.8564	49.9216	49.8726	49.8778
RMSE	0.8998	0.6909	0.5931	0.4849
Accuracy	0.0551	0.0426	0.0359	0.0291
95% C.I.	49.801,49.911	49.879,49.964	49.837,49.909	49.849,49.907
GMM4	49.7912	49.8827	49.8076	49.8232
RMSE	0.9950	0.7768	0.7113	0.5993
Accuracy	0.0603	0.0476	0.0425	0.0355
95% C.I.	49.731,49.852	49.835,49.930	49.765,49.850	49.788,49.859

Table 2.6: Simulation results of the Logistic model with two parameters - continued

	n = 100	n = 200	n=300	n = 500
		$\sigma^2 = 1$		
SLS	0.9927	0.9974	0.9934	0.9956
RMSE	0.1357	0.0940	0.0820	0.0628
Accuracy	0.0084	0.0058	0.0051	0.0039
95% C.I.	0.984,1.001	0.992,1.003	0.988,0.998	0.992,0.999
GMM3	0.8370	0.8097	0.8233	0.7939
RMSE	0.4290	0.4482	0.4490	0.4452
Accuracy	0.0246	0.0252	0.0256	0.0245
95% C.I.	0.812,0.862	0.785,0.835	0.798,0.849	0.769,0.818
GMM4	0.8150	0.8193	0.8077	0.7978
RMSE	0.4463	0.4464	0.4453	0.4543
Accuracy	0.0252	0.0253	0.0249	0.0252
95% C.I.	0.790,0.840	0.794,0.845	0.783,0.833	0.773,0.823

Table 2.7: Simulation results of the Linear-Exponential model with two parameters

	n = 100	n = 200	n = 300	n = 500
		$\beta_1 = -3$		
SLS	-3.0766	-3.0452	-3.0168	-3.0235
RMSE	0.7899	0.5711	0.4997	0.3840
Accuracy	0.0487	0.0353	0.0310	0.0238
95% C.I.	-3.125,-3.028	-3.080,-3.010	-3.048,-2.986	-3.047,-3.000
GMM3	-3.0474	-2.9455	-2.9732	-2.9497
RMSE	0.5817	0.5728	0.5627	0.5463
Accuracy	0.0360	0.0354	0.0349	0.0337
95% C.I.	-3.083,-3.011	-2.981,-2.910	-3.008,-2.938	-2.983,-2.916
GMM4	-3.0470	-2.9862	-2.9290	-2.9405
RMSE	0.5833	0.5677	0.5797	0.5693
Accuracy	0.0361	0.0352	0.0357	0.0351
95% C.I.	-3.083,-3.011	-3.021,-2.951	-2.965,-2.893	-2.976,-2.905

Table 2.8: Simulation results of the Linear-Exponential model with two parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\beta_2 = -1$		
SLS	-1.0182	-1.0104	-1.0103	-1.0027
RMSE	0.0798	0.0722	0.0538	0.0381
Accuracy	0.0048	0.0044	0.0033	0.0024
95% C.I.	-1.023,-1.013	-1.015,-1.006	-1.014,-1.007	-1.005,-1.000
GMM3	-1.0180	-1.0351	-1.0293	-1.0275
RMSE	0.0950	0.1417	0.1372	0.1137
Accuracy	0.0058	0.0085	0.0083	0.0068
95% C.I.	-1.024,-1.012	-1.044,-1.027	-1.038,-1.021	-1.034,-1.021
GMM4	-1.0268	-1.0354	-1.0508	-1.0393
RMSE	0.1205	0.1883	0.2011	0.1668
Accuracy	0.0073	0.0115	0.0121	0.0101
95% C.I.	-1.034,-1.020	-1.047,-1.024	-1.063,-1.039	-1.049,-1.029

Table 2.9: Simulation results of the Exponential model with three parameters

	n = 100	n = 200	n = 300	n = 500
		$\beta_1 = 10$		
SLS	10.0228	10.0055	10.0182	9.9832
RMSE	0.6031	0.3615	0.3346	0.2789
Accuracy	0.0374	0.0224	0.0207	0.0173
95% C.I.	9.985,10.060	9.983,10.028	9.997,10.039	9.966,10.001
GMM3	10.1560	10.2193	10.2312	10.1981
RMSE	0.9316	0.8114	0.9176	0.9007
Accuracy	0.0570	0.0484	0.0551	0.0545
95% C.I.	10.099,10.213	10.171,10.268	10.176,10.286	10.144,10.253
GMM4	10.2152	10.4046	10.2722	10.2912
RMSE	1.0557	0.9777	1.0554	1.0189
Accuracy	0.0641	0.0552	0.0632	0.0606
95% C.I.	10.151,10.279	10.349,10.460	10.209,10.335	10.231,10.352

Table 2.10: Simulation results of the Exponential model with three parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\beta_2 = -0.5$		
SLS	-0.5042	-0.5017	-0.5026	-0.4997
RMSE	0.0381	0.0257	0.0234	0.0192
Accuracy	0.0024	0.0016	0.0014	0.0012
95% C.I.	-0.507,-0.502	-0.503,-0.500	-0.504,-0.501	-0.501,-0.499
GMM3	-0.5150	-0.5365	-0.5261	-0.5196
RMSE	0.0940	0.1239	0.1094	0.0993
Accuracy	0.0058	0.0073	0.0066	0.0060
95% C.I.	-0.521,-0.509	-0.544,-0.529	-0.533,-0.519	-0.526,-0.514
GMM4	-0.5270	-0.5916	-0.5419	-0.5391
RMSE	0.1317	0.2020	0.1591	0.1411
Accuracy	0.0080	0.0112	0.0095	0.0084
95% C.I.	-0.535,-0.519	-0.603,-0.580	-0.551,-0.532	-0.548, -0.531

Table 2.11: Simulation results of the Exponential model with three parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\sigma^2 = 1$		
SLS	0.9743	0.9843	0.9887	0.9967
RMSE	0.1492	0.1082	0.0954	0.0774
Accuracy	0.0091	0.0066	0.0059	0.0048
95% C.I.	0.965,0.983	0.978,0.991	0.983,0.995	0.992,1.001
GMM3	0.8518	0.8719	0.8424	0.8427
RMSE	0.4360	0.4245	0.4260	0.4319
Accuracy	0.0254	0.0251	0.0245	0.0249
95% C.I.	0.826,0.877	0.847,0.897	0.818,0.867	0.818,0.868
GMM4	0.8217	0.7826	0.7897	0.8082
RMSE	0.4424	0.4558	0.4527	0.4490
Accuracy	0.0251	0.0248	0.0249	0.0252
95% C.I.	0.797,0.847	0.758,0.807	0.765,0.815	0.783,0.833

Table 2.12: Simulation results of the Logistic model with three parameters

	n = 100	n = 200	n = 300	n = 500
		$\beta_1 = 50$		
SLS	50.0984	50.0689	50.0330	50.0295
RMSE	0.8201	0.5725	0.4620	0.3400
Accuracy	0.0505	0.0352	0.0286	0.0210
95% C.I.	50.048,50.149	50.034,50.104	50.004,50.062	50.009,50.051
GMM3	49.9090	49.8856	49.8797	49.8675
RMSE	0.8464	0.6756	0.5956	0.5430
Accuracy	0.0522	0.0413	0.0362	0.0327
95% C.I.	49.857,49.961	49.844,49.927	49.844,49.916	49.835,49.900
GMM4	49.8544	49.8440	49.8297	49.8212
RMSE	0.9465	0.8296	0.7164	0.6844
Accuracy	0.0580	0.0505	0.0431	0.0410
95% C.I.	49.796,49.912	49.793,49.895	49.787,49.873	49.780,49.862

Table 2.13: Simulation results of the Logistic model with three parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\beta_2 = 34$		
SLS	33.9483	33.9557	33.9568	33.9424
RMSE	1.3131	1.0907	0.9363	0.7140
Accuracy	0.0814	0.0676	0.0580	0.0441
%95 C.I.	33.867,34.030	33.888,34.023	33.899,34.015	33.898,33.987
GMM3	33.9386	34.0087	33.9917	34.0285
RMSE	1.1588	1.1623	1.1364	1.1621
Accuracy	0.0718	0.0721	0.0705	0.0720
95% C.I.	33.867,34.010	33.937,34.081	33.921,34.062	33.956,34.101
GMM4	33.9530	33.9826	34.0543	34.0038
RMSE	1.1738	1.1250	1.1373	1.1623
Accuracy	0.0727	0.0698	0.0704	0.0721
95% C.I.	33.880,34.026	33.913,34.052	33.984,34.125	33.932,34.076

Table 2.14: Simulation results of the Logistic model with three parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\sigma^2 = 1$		
SLS	0.9695	0.9848	0.9862	0.9944
RMSE	0.1452	0.1003	0.0833	0.0637
Accuracy	0.0088	0.0061	0.0051	0.0039
95% C.I.	0.961,0.978	0.979,0.991	0.981,0.991	0.990,0.998
GMM3	0.7941	0.8108	0.8136	0.8099
RMSE	0.4502	0.4361	0.4397	0.4385
Accuracy	0.0248	0.0244	0.0247	0.0245
95% C.I.	0.769,0.819	0.786,0.835	0.789,0.838	0.785,0.834
GMM4	0.7884	0.8137	0.8339	0.8125
RMSE	0.4581	0.4311	0.4341	0.4496
Accuracy	0.0252	0.0241	0.0249	0.0253
95% C.I.	0.763,0.814	0.790,0.838	0.809,0.859	0.787,0.838

Table 2.15: Simulation results of the Linear-Exponential model with three parameters

	n = 100	n = 200	n = 300	n = 500
		$\beta_1 = -3$		
SLS	-3.0091	-2.9845	-3.0173	-3.0040
RMSE	0.7339	0.6662	0.6601	0.5772
Accuracy	0.0455	0.0413	0.0409	0.0358
95% C.I.	-3.055,-2.964	-3.026,-2.943	-3.058,-2.976	-3.040,-2.968
GMM3	-3.0118	-2.9210	-2.9855	-2.9919
RMSE	0.5904	0.5760	0.5572	0.5631
Accuracy	0.0366	0.0354	0.0345	0.0349
95% C.I.	-3.048,-2.975	-2.956,-2.886	-3.020,-2.951	-3.027,-2.957
GMM4	-3.0021	-2.9519	-2.9551	-2.9644
RMSE	0.5685	0.5684	0.5713	0.5921
Accuracy	0.0353	0.0351	0.0353	0.0367
95% C.I.	-3.037,-2.967	-2.987,-2.917	-2.990,-2.920	-3.001,-2.928

Table 2.16: Simulation results of the Linear-Exponential model with three parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\beta_2 = -1$		
SLS	-1.0175	-1.0133	-1.0172	-1.0057
RMSE	0.0916	0.0647	0.0528	0.0396
Accuracy	0.0056	0.0039	0.0032	0.0024
95% C.I.	-1.023,-1.012	-1.017,-1.009	-1.020,-1.014	-1.008,-1.003
GMM3	-1.0242	-1.0826	-1.0614	-1.0096
RMSE	0.1327	0.2371	0.1980	0.1792
Accuracy	0.0081	0.0138	0.0117	0.0107
95% C.I.	-1.032,-1.016	-1.096,-1.069	-1.073,-1.050	-1.020,-0.999
GMM4	-1.0499	-1.1092	-1.1312	-1.0841
RMSE	0.2035	0.3153	0.3364	0.2681
Accuracy	0.0122	0.0183	0.0192	0.0158
95% C.I.	-1.062,-1.038	-1.128,-1.091	-1.150,-1.112	-1.100,-1.068

Table 2.17: Simulation results of the Linear-Exponential model with three parameters - continued

	n = 100	n = 200	n = 300	n = 500
		$\beta_3 = 5$		
SLS	4.9739	5.0194	5.0969	5.0629
RMSE	1.4251	0.9281	0.8949	0.8469
Accuracy	0.0884	0.0575	0.0552	0.0524
95% C.I.	4.886,5.062	4.962,5.077	5.042,5.152	5.011,5.115
GMM3	4.9434	5.1693	5.2010	5.1909
RMSE	1.1403	1.0668	0.9990	1.0453
Accuracy	0.0706	0.0653	0.0607	0.0637
95% C.I.	4.873,5.014	5.104,5.235	5.140,5.262	5.127,5.255
GMM4	5.0251	5.2054	5.2798	5.2024
RMSE	1.1481	1.1331	1.1033	1.0956
Accuracy	0.0712	0.0691	0.0662	0.0668
95% C.I.	4.954,5.096	5.136,5.275	5.214,5.346	5.136,5.269

2.5 Censored Regression Models

2.5.1 Model and Estimation

Consider the censored regression model

$$Y^* = X'\beta + \varepsilon, \qquad Y = \max(Y^*, 0) \tag{2.8}$$

where Y^* and $Y \in I\!R$ are latent and observed response variables, $X \in I\!R^p$ is a vector of observed predictors, and ε is the random error with density $f(\varepsilon; \phi)$ which satisfies $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) < \infty$. The unknown parameters are $\beta \in \Omega \subset I\!R^p$ and $\phi \in \Phi \subset I\!R^q$. Although we consider zero as the left censoring value, it can be easily replaced by any other constant.

Under model (2.8), the first two conditional moments of Y given X are respectively given by

$$E(Y|X) = \int I(X'\beta + \varepsilon)(X'\beta + \varepsilon) f(\varepsilon; \phi) d\varepsilon, \tag{2.9}$$

$$E(Y^2|X) = \int I(X'\beta + \varepsilon)(X'\beta + \varepsilon)^2 f(\varepsilon; \phi) d\varepsilon, \tag{2.10}$$

where $I(x) = 1$ for $x > 0$ and $I(x) = 0$ for $x \leq 0$. Let $\gamma = (\beta', \phi')'$ denote the vector of model parameters and $\Gamma = \Omega \times \Phi \subset I\!R^{p+q}$ the parameter space.

For every $x \in I\!R^p$ and $\gamma \in \Gamma$, define

$$m_1(x;\gamma) = \int I(x'\beta + \varepsilon)(x'\beta + \varepsilon) f(\varepsilon;\phi) d\varepsilon \tag{2.11}$$

$$m_2(x;\gamma) = \int I(x'\beta + \varepsilon)(x'\beta + \varepsilon)^2 f(\varepsilon;\phi) d\varepsilon. \tag{2.12}$$

Now suppose $(Y_i, X_i')', i = 1, 2, ..., n$ is an $i.i.d.$ random sample and

$$\rho_i(\gamma) = \left(Y_i - m_1(X_i;\gamma), Y_i^2 - m_2(X_i;\gamma)\right)'.$$

Then the second-order least squares (SLS) estimator $\hat{\gamma}_n$ for γ is defined as the measurable function that minimizes (2.2). For some error distribution functions $f(\varepsilon;\phi)$, explicit forms of the integrals in the first two conditional moments may be quite troublesome. We consider some examples where explicit forms of these moments can be obtained.

Example 2.5.1. First consider the model (2.8) where ε has a normal distribution $N(0, \sigma_\varepsilon^2)$. This is a standard Tobit model. For this model $\gamma = (\beta', \sigma_\varepsilon^2)'$, and the first two conditional moments are given by

$$E(Y|X) = X'\beta\Phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right) + \sigma_\varepsilon\phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right)$$

and

$$E(Y^2|X) = \left[(X'\beta)^2 + \sigma_\varepsilon^2\right]\Phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right) + \sigma_\varepsilon(X'\beta)\phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right),$$

where Φ and ϕ are the standard normal distribution and density function respectively.

Example 2.5.2. Now consider the model (2.8) where $\frac{\varepsilon}{\sigma_\varepsilon}\sqrt{\frac{k}{k-2}}$ has a t distribution $t(k)$ with $k > 2$. Then we have

$$E(Y|X) = (X'\beta)F_k\left(\frac{X'\beta}{\sigma_\varepsilon}\right) + \frac{\sigma_\varepsilon\sqrt{k}\Gamma(\frac{k-1}{2})}{2\sqrt{\pi}\Gamma(\frac{k}{2})}\left(1+\frac{(X'\beta)^2}{k\sigma_\varepsilon^2}\right)^{-(k-1)/2}$$

and

$$\begin{aligned} E(Y^2|X) &= \left[(X'\beta)^2 + k\sigma_\varepsilon^2\right] F_k\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \\ &+ \frac{\sigma_\varepsilon(X'\beta)\sqrt{k}\Gamma(\frac{k-1}{2})}{\sqrt{\pi}\Gamma(\frac{k}{2})}\left(1+\frac{(X'\beta)^2}{k\sigma_\varepsilon^2}\right)^{-(k-1)/2} \\ &+ \frac{k(k-1)\sigma_\varepsilon^2}{k-2}F_{k-2}\left(\sqrt{\frac{k-2}{k}}\frac{X'\beta}{\sigma_\varepsilon}\right), \end{aligned}$$

where F_k is the distribution function of $t(k)$.

Example 2.5.3. Now consider the model (2.8) where $\frac{\varepsilon\sqrt{2k}}{\sigma_\varepsilon}+k$ has a chi-square distribution $\chi^2(k), k > 1$. Then we have

$$\begin{aligned} E(Y|X) &= (X'\beta + k\sigma_\varepsilon) \\ &- (1-I(X'\beta))\left[(X'\beta)F_k\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) + k\sigma_\varepsilon F_{k+2}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right] \end{aligned}$$

and

$$\begin{aligned} E(Y^2|X) &= (X'\beta + k\sigma_\varepsilon)^2 + 2k\sigma_\varepsilon^2 - (1-I(X'\beta))\left[(X'\beta)^2F_k\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right. \\ &+ \left. 2k\sigma_\varepsilon(X'\beta)F_{k+2}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) + \sigma_\varepsilon^2 k(k+2)F_{k+4}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right], \end{aligned}$$

where F_k is the distribution function of $\chi^2(k)$.

The above three models and the corresponding moments will be used in our Monte Carlo simulation studies.

2.5.2 Consistency and Asymptotic Normality

To prove the consistency and asymptotic normal distribution of the SLS $\hat{\gamma}_n$, we need to assume some regularity conditions. Here we adopt the setup of Amemiya (1973) and express these regularity conditions in terms of error distribution $f(\varepsilon; \phi)$.

Assumption 7. *$f(\varepsilon; \phi)$ is continuous in $\phi \in \Phi$ for μ-almost all ε.*

Assumption 8. *The parameter space $\Gamma \subset I\!R^{p+q}$ is compact.*

Assumption 9. *The weight $W(X)$ is nonnegative definite with probability one and satisfies $E\|W(X)\|(Y^4 + \|X\|^4 + 1) < \infty$.*

Assumption 10. $E\|W(X)\| \int(\varepsilon^4 + 1)\sup_\Phi f(\varepsilon, \phi)d\varepsilon < \infty$.

Assumption 11. *For any $\gamma \in \Gamma$, $E[\rho(\gamma) - \rho(\gamma_0)]'W(X)[\rho(\gamma) - \rho(\gamma_0)] = 0$ if and only if $\gamma = \gamma_0$.*

Assumption 12. *There exists an open subset $\phi_0 \in \Phi_0 \subset \Phi$ in which $f(\varepsilon; \phi)$ is twice continuously differentiable with respect to ϕ, for μ-almost* *all ε. Furthermore, there exists positive function $K(\varepsilon)$, such that the first two partial derivatives of $f(\varepsilon;\phi)$ with respect to ϕ are bounded by $K(\varepsilon)$, and $E\,\|W(X)\|\int(\varepsilon^4+1)K^2(\varepsilon)d\varepsilon<\infty$.*

Assumption 13. *The matrix $B=E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}W(X)\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]$ is nonsingular.*

Assumption A12 guaranties uniform convergence of the second derivative of $Q_n(\gamma)$. This assumption and the Dominated Convergence Theorem together imply that the first derivatives $\partial m_1(X;\gamma)/\partial\gamma$ and $\partial m_2(X;\gamma)/\partial\gamma$ exist and their elements are respectively given by

$$\frac{\partial m_1(x;\gamma)}{\partial\beta}=x\int_{-x'\beta}^{\infty}f(\varepsilon;\phi)d\varepsilon$$
$$\frac{\partial m_1(x;\gamma)}{\partial\phi}=\int_{-x'\beta}^{\infty}(x'\beta+\varepsilon)\frac{\partial f(\varepsilon;\phi)}{\partial\phi}d\varepsilon$$

and

$$\frac{\partial m_2(x;\gamma)}{\partial\beta}=2x\int_{-x'\beta}^{\infty}(x'\beta+\varepsilon)f(\varepsilon;\phi)d\varepsilon$$
$$\frac{\partial m_2(x;\gamma)}{\partial\phi}=\int_{-x'\beta}^{\infty}(x'\beta+\varepsilon)^2\frac{\partial f(\varepsilon;\phi)}{\partial\phi}d\varepsilon.$$

Theorem 2.5.4. *Under the assumptions A7 – A11, the SLS $\hat{\gamma}_n\xrightarrow{a.s.}\gamma_0$, as $n\to\infty$.*

Theorem 2.5.5. *Under assumptions A7 – A13, as $n\to\infty$, $\sqrt{n}(\hat{\gamma}_n-\gamma_0)\xrightarrow{L}N(0,B^{-1}AB^{-1})$, where A and B are given in (2.3) and (2.4), respectively.*

2.5.3 Efficient Choice of the Weighting Matrix

The SLS $\hat{\gamma}_n$ depends on the weighting matrix $W(X)$. To answer the question of how to choose $W(X)$ to obtain the most efficient estimator, we first note that, since $\partial\rho'(\gamma_0)/\partial\gamma$ does not depend on Y matrix A in (2.3) can be written as

$$A = E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma} W(X) F(X) W(X) \frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right],$$

where

$$F(X) = E[\rho(\gamma_0)\rho'(\gamma_0)|X]$$

and has elements

$$F_{11} = E[(Y - m_1(X;\gamma_0))^2|X] = m_2 - m_1^2$$

$$F_{22} = E[(Y^2 - m_2(X;\gamma_0))^2|X] = m_4 - m_2^2$$

and

$$F_{12} = E[(Y - m_1(X;\gamma_0))(Y^2 - m_2(X;\gamma_0))|X] = m_3 - m_1 m_2,$$

where

$$m_3(x;\gamma) \;=\; \int I(x'\beta + \varepsilon)(x'\beta + \varepsilon)^3 f(\varepsilon;\phi) d\varepsilon$$

and

$$m_4(x;\gamma) \;=\; \int I(x'\beta + \varepsilon)(x'\beta + \varepsilon)^4 f(\varepsilon;\phi) d\varepsilon.$$

We calculated the third and the fourth conditional moments for the examples we presented at the beginning of this section. We will refer to these moments later in the simulation study.

For the model in Example 2.5.1:

$$\begin{aligned} E(Y^3|X) &= [(X'\beta)^3 + 3(X'\beta)\sigma_\varepsilon^2]\Phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \\ &+ \sigma_\varepsilon[2\sigma_\varepsilon^2 + (X'\beta)^2]\phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \end{aligned}$$

and

$$\begin{aligned} E(Y^4|X) &= [(X'\beta)^4 + 6(X'\beta)^2\sigma_\varepsilon^2 + 3\sigma_\varepsilon^4]\Phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \\ &+ \sigma_\varepsilon(X'\beta)[5\sigma_\varepsilon^2 + (X'\beta)^2]\phi\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \end{aligned}$$

For the model in Example 2.5.2 with $k > 4$:

$$\begin{aligned} E(Y^3|X) &= (X'\beta)[(X'\beta)^2 - 3k\sigma_\varepsilon^2]F_k\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \\ &- \frac{4\sigma_\varepsilon(X'\beta)^2\sqrt{k/\pi}\,\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})(1-k)}\left(1 + \frac{(X'\beta)^2}{k\sigma_\varepsilon^2}\right)^{\frac{1-k}{2}} \\ &+ \frac{3k(k-1)\sigma_\varepsilon^2}{k-2}(X'\beta)F_{k-2}\left(\sqrt{\frac{k-2}{k}}\frac{X'\beta}{\sigma_\varepsilon}\right) \\ &+ \frac{2\sigma_\varepsilon^3 k\sqrt{k/\pi}\,\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})(1-k)(3-k)}\left(1 + \frac{(X'\beta)^2}{k\sigma_\varepsilon^2}\right)^{\frac{3-k}{2}} \end{aligned}$$

and

$$\begin{aligned}
E(Y^4|X) &= [(X'\beta)^4 + k^2\sigma_\varepsilon^4 - 6k^2\sigma_\varepsilon^4(X'\beta)^2]F_k\left(\frac{X'\beta}{\sigma_\varepsilon}\right) \\
&- \frac{8\sigma_\varepsilon(X'\beta)^3\sqrt{k/\pi}\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})(1-k)}\left(1+\frac{(X'\beta)^2}{k\sigma_\varepsilon^2}\right)^{\frac{1-k}{2}} \\
&+ \frac{2k^2\sigma_\varepsilon^2(k-1)}{k-2}(3(X'\beta)^2-1)F_{k-2}\left(\sqrt{\frac{k-2}{k}}\frac{X'\beta}{\sigma_\varepsilon}\right) \\
&+ \frac{8\sigma_\varepsilon^3(X'\beta)k\sqrt{k/\pi}\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})(1-k)(3-k)}\left(1+\frac{(X'\beta)^2}{k\sigma_\varepsilon^2}\right)^{\frac{3-k}{2}} \\
&+ k^2\sigma_\varepsilon^4\left[\frac{(k-1)(k-3)}{(k-2)(k-4)}F_{k-4}\left(\sqrt{\frac{k-4}{k}}\frac{X'\beta}{\sigma_\varepsilon}\right)\right]
\end{aligned}$$

For the model in Example 2.5.3:

$$\begin{aligned}
E(Y^3|X) &= (X'\beta)^3 + 3k\sigma_\varepsilon(X'\beta)^2 + k(k+2)\sigma_\varepsilon^2[3(X'\beta) + (k+4)\sigma_\varepsilon] \\
&+ (1 - I(X'\beta))\left[(X'\beta)^3F_k\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right. \\
&+ 3k\sigma_\varepsilon(X'\beta)^2F_{k+2}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) \\
&+ 3k(k+2)\sigma_\varepsilon^2(X'\beta)F_{k+4}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) \\
&+ \left. k(k+2)(k+4)\sigma_\varepsilon^3F_{k+6}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
E(Y^4|X) &= (X'\beta)^4 + 4k\sigma_\varepsilon(X'\beta)^3 + 6k(k+2)\sigma_\varepsilon^2(X'\beta)^2 \\
&+ k(k+2)(k+4)\sigma_\varepsilon^3[4(X'\beta) + \sigma_\varepsilon(k+6)] \\
&+ (1 - I(X'\beta))\left[(X'\beta)^4 F_k\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right. \\
&+ 4k\sigma_\varepsilon(X'\beta)^3 F_{k+2}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) \\
&+ 6k(k+2)\sigma_\varepsilon^2(X'\beta)^2 F_{k+4}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) \\
&+ 4k(k+2)(k+4)\sigma_\varepsilon^3(X'\beta)F_{k+6}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right) \\
&+ \left. k(k+2)(k+4)(k+6)\sigma_\varepsilon^4 F_{k+8}\left(\frac{-X'\beta}{\sigma_\varepsilon}\right)\right]
\end{aligned}$$

We proved in Theorem 2.2.1 that

$$B^{-1}AB^{-1} \geq E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}F(X)^{-1}\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]^{-1} \tag{2.13}$$

and the lower bound is attained for $W(X) = F(X)^{-1}$ in B and A. The matrix $F(X)$ is invertible if its determinant $F_{11}F_{22} - F_{12}^2 > 0$. As we will see later in the simulation study, the weighting matrix $W(X)$ plays a crucial role in the efficiency of the SLS estimator.

2.5.4 Simulation-Based Estimator

The numerical computation of the SLS estimator of the last section can be done using standard numerical procedures when closed forms of the first two conditional moments are available. However, sometimes explicit forms of the integrals in (2.11) and (2.12) may be difficult or impossible to derive. In this case numerical minimization of $Q_n(\gamma)$ will be troublesome, especially when the dimension of parameters $p+q$ is greater than two or three. To overcome this computational difficulty, in this section we consider a simulation-based approach in which the integrals are simulated by Monte Carlo methods. First note that by a change of variables the integrals in (2.11) and (2.12) can be written as

$$m_1(x;\gamma) = \int I(\varepsilon)\varepsilon f(\varepsilon - x'\beta;\phi)d\varepsilon \tag{2.14}$$

$$m_2(x;\gamma) = \int I(\varepsilon)\varepsilon^2 f(\varepsilon - x'\beta;\phi)d\varepsilon. \tag{2.15}$$

The simulation-based estimator can be constructed in the following way. First, choose a known density $l(t)$ with support in $[0,+\infty)$ and generate an *i.i.d.* random sample $\{t_{ij}, j=1,2,...,2S, i=1,2,...,n\}$ from $l(t)$. Then approximate $m_1(x;\gamma)$ and $m_2(x;\gamma)$ by the Monte Carlo simulators

$$m_{1,S}(x_i;\gamma) = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}f(t_{ij}-x_i'\beta;\phi)}{l(t_{ij})}$$

$$m_{1,2S}(x_i;\gamma) = \frac{1}{S}\sum_{j=S+1}^{2S}\frac{t_{ij}f(t_{ij}-x_i'\beta;\phi)}{l(t_{ij})}$$

and

$$m_{2,S}(x_i;\gamma) = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}^2 f(t_{ij}-x_i'\beta;\phi)}{l(t_{ij})}$$

$$m_{2,2S}(x_i;\gamma) = \frac{1}{S}\sum_{j=S+1}^{2S}\frac{t_{ij}^2 f(t_{ij}-x_i'\beta;\phi)}{l(t_{ij})}.$$

Hence, a simulated version of the objective function $Q_n(\gamma)$ can be defined as

$$Q_{n,S}(\gamma) = \sum_{i=1}^{n}\rho_{i,S}^{'}(\gamma)W_i\rho_{i,2S}(\gamma), \tag{2.16}$$

where

$$\rho_{i,S}(\gamma) = \left(Y_i - m_{1,S}(X_i;\gamma), Y_i^2 - m_{2,S}(X_i;\gamma)\right)'$$

$$\rho_{i,2S}(\gamma) = \left(Y_i - m_{1,2S}(X_i;\gamma), Y_i^2 - m_{2,2S}(X_i;\gamma)\right)'.$$

Since $\rho_{i,S}(\gamma)$ and $\rho_{i,2S}(\gamma)$ are conditionally independent given Y_i, X_i, $Q_{n,S}(\gamma)$ is an unbiased simulator for $Q_n(\gamma)$. Finally, the Simulation-Based estimator (SBE) for γ can be defined by

$$\hat{\gamma}_{n,S} = \text{argmin}_{\gamma\in\Gamma}Q_{n,S}(\gamma).$$

Note that $Q_{n,S}(\gamma)$ is continuous in, and differentiable with respect to γ, as long as function $f(\varepsilon - X'\beta;\phi)$ has these properties. In particular, the first derivative of $\rho_{i,S}(\gamma)$ becomes

$$\frac{\partial\rho_{i,S}^{'}(\gamma)}{\partial\gamma} = -\left(\frac{\partial m_{1,S}(X_i;\gamma)}{\partial\gamma}, \frac{\partial m_{2,S}(X_i;\gamma)}{\partial\gamma}\right),$$

where $\partial m_{1,S}(X_i;\gamma)/\partial\gamma$ is a column vector with elements

$$\frac{\partial m_{1,S}(X_i;\gamma)}{\partial\beta} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}}{l(t_{ij})}\frac{\partial f(t_{ij}-X_i'\beta;\phi)}{\partial\beta}$$

$$\frac{\partial m_{1,S}(X_i;\gamma)}{\partial\phi} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}}{l(t_{ij})}\frac{\partial f(t_{ij}-X_i'\beta;\phi)}{\partial\phi}$$

and $\partial m_{2,S}(X_i;\gamma)/\partial\gamma$ is a column vector with elements

$$\frac{\partial m_{2,S}(X_i;\gamma)}{\partial\beta} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}^2}{l(t_{ij})}\frac{\partial f(t_{ij}-X_i'\beta;\phi)}{\partial\beta}$$

$$\frac{\partial m_{2,S}(X_i;\gamma)}{\partial\phi} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}^2}{l(t_{ij})}\frac{\partial f(t_{ij}-X_i'\beta;\phi)}{\partial\phi}.$$

The derivatives $\partial m_{1,2S}(X_i;\gamma)/\partial\gamma$ and $\partial m_{2,2S}(X_i;\gamma)/\partial\gamma$ can be given similarly.

With the modified expression (2.14) and (2.15), Assumption A12 can be substituted by the following assumption.

Assumption 14. *There exist an open subset $\gamma_0 \in \Gamma_0 \subset \Omega$, in which $f(\varepsilon - x'\beta;\phi)$ is twice continuously differentiable with respect to γ and ε. Furthermore, there exists positive function $K(\varepsilon, x)$, such that the first two partial derivatives of $f(\varepsilon - x'\beta;\phi)$ with respect to β and ϕ are bounded by $K(\varepsilon, x)$, and $E\,\|W(X)\| \int(\varepsilon^4+1)K^2(\varepsilon, X)d\varepsilon < \infty$.*

With the new assumption, the first derivatives $\partial m_1(x;\gamma)/\partial\gamma$ and $\partial m_2(x;\gamma)/\partial\gamma$ exist and their elements are respectively given by

$$\frac{\partial m_1(x;\gamma)}{\partial\beta} = \int_0^\infty \varepsilon\frac{\partial f(\varepsilon - x'\beta;\phi)}{\partial\beta}d\varepsilon, \frac{\partial m_1(x;\gamma)}{\partial\phi} = \int_0^\infty \varepsilon\frac{\partial f(\varepsilon - x'\beta;\phi)}{\partial\phi}d\varepsilon$$

and

$$\frac{\partial m_2(x;\gamma)}{\partial\beta} = \int_0^\infty \varepsilon^2\frac{\partial f(\varepsilon - x'\beta;\phi)}{\partial\beta}d\varepsilon, \frac{\partial m_2(x;\gamma)}{\partial\phi} = \int_0^\infty \varepsilon^2\frac{\partial f(\varepsilon - x'\beta;\phi)}{\partial\phi}d\varepsilon.$$

Therefore, Theorems 2.5.4 and 2.5.5 can be proved in the same way. We have the following theorem for the simulation-based estimator.

Theorem 2.5.6. *Suppose the* $Supp[l(t)] \supseteq [0,+\infty) \cap Supp[f(\varepsilon - X'\beta;\phi)]]$ *for all* $\gamma \in \Gamma$ *and* $x \in \mathbb{R}^p$. *Then as* $n \to \infty$,

(i). Under assumptions A7 – A11, $\hat{\gamma}_{n,S} \xrightarrow{a.s.} \gamma_0$.

(ii). Under assumptions A7 – A11, and A13 – A14,

$\sqrt{n}(\hat{\gamma}_{n,S} - \gamma_0) \xrightarrow{L} N(0, B^{-1}A_S B^{-1})$, *where*

$$\begin{aligned} 2A_S &= E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W(X)\rho_{1,2S}(\gamma_0)\rho'_{1,2S}(\gamma_0)W(X)\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'}\right] \\ &+ E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W(X)\rho_{1,2S}(\gamma_0)\rho'_{1,S}(\gamma_0)W(X)\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right]. \end{aligned}$$

The proof for Theorem 2.5.6 is analogous to Theorem 3 in Wang (2004) and is therefore omitted. In general, the simulation-based estimator $\hat{\gamma}_{n,S}$ is less efficient than the SLS $\hat{\gamma}_n$, due to the simulation approximation of $\rho_i(\gamma)$ by $\rho_{i,S}(\gamma)$ and $\rho_{i,2S}(\gamma)$. Wang (2004) showed that the efficiency loss caused by simulation has a magnitude of $O(1/S)$. Therefore, the larger the simulation size S, the smaller the efficiency loss. Note also that the above asymptotic results do not require the simulation size S tends to infinity.

2.5.5 Monte Carlo Simulation Studies

In this section we study finite sample behavior of the SLS estimator with both identity and optimal weighting matrices, and compare them with the maximum likelihood estimator (MLE). We conducted substantial simulation studies using a variety of configurations with respect to sample size, degree of censoring, and error distribution.

In particular, we simulate the three models in Examples 2.5.1 – 2.5.3 with $\beta = (\beta_1, \beta_2)'$ and $\phi = \sigma_\varepsilon^2$. For each model, we consider the amount of censoring of 31% and 60% respectively. The covariate X has a normal distribution. For each model, 1000 Monte Carlo repetitions are carried out for each of the sample sizes $n = 50, 70, 100, 200, 300, 500$. We computed the Monte Carlo means of the SLS estimator using the identity weight (SLSIDEN) and the optimal weight (SLSOPT), together with their root mean squared errors

(RMSE).

The simulation results for all sample sizes are summarized in Figures 2.4 and 2.5. In particular, Figure 2.4 contains the estimates of the model with normal errors, 60% censoring, and the true parameter values $\beta_1 = -6, \beta_2 = 1.5$ and $\sigma_\varepsilon^2 = 16$. Figure 2.5 compares the root mean squared errors (RMSE) of the estimators of σ_ε^2 for the three models with 60% censoring. Both figures show that the SLSOPT and MLE perform very similarly. Furthermore, Figure 2.4 shows that all three estimators achieve their large-sample properties with moderate sample sizes, even for a relatively high amount of censoring. Figure 2.5 shows that the SLSIDEN clearly has smaller RMSE than the other two estimators in all models.

Next we present more detailed numerical results for the sample size $n = 200$. In particular, Tables 2.18 – 2.23 present the simulation results for the models with normal, t, and chi-square error distributions and 31% and 60% censoring respectively. Moreover, we report 95% confidence intervals for the parameters as well.

Again, these results show that in general the SLS with optimal weight (SLSOPT) performs very closely to the MLE. This pattern holds for both

Figure 2.4: Monte Carlo estimates of the normal model with 60% censoring and various sample sizes.

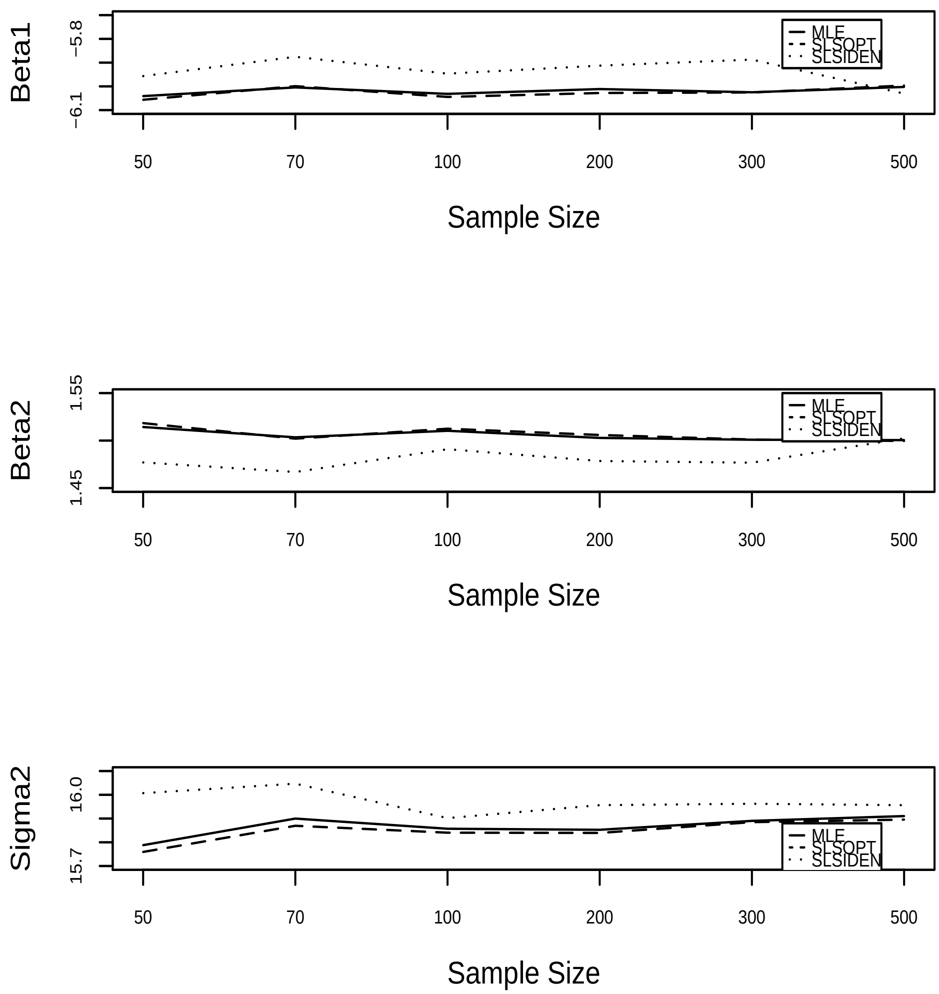
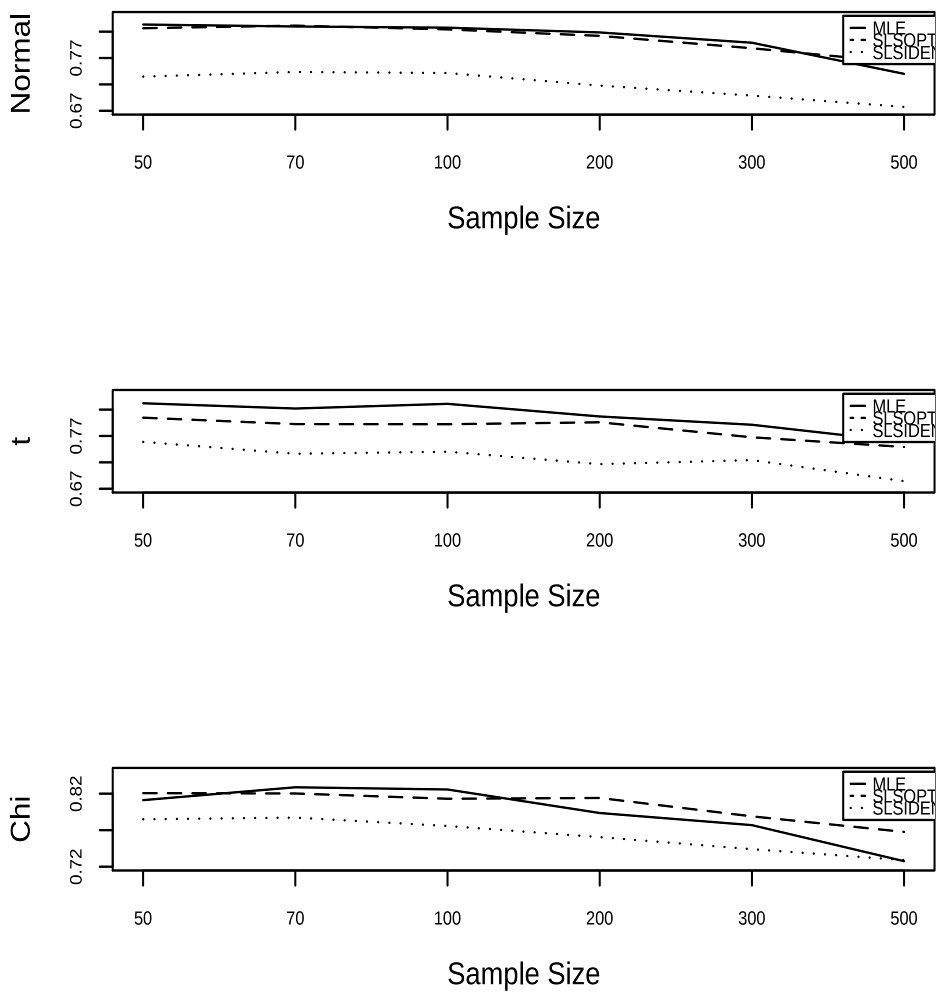


Figure 2.5: RMSE of the estimators of σ_ε^2 in three models with 60% censoring and various sample sizes.



amounts of censoring. Generally, the bias of the estimators increases as the error distribution changes from symmetric to asymmetric. In the case of χ^2 error distribution, MLE shows more bias in both β_1 and β_2. Comparing Tables 2.18 – 2.23 reveals that as the proportion of censored observation declines, the RMSE of the estimators decreases. This is because of the decrease in the variance of the estimators. It is also apparent that the 95% confidence intervals for β_2 are generally shorter than the confidence intervals for β_1 and σ_ε^2. As usual, estimating σ_ε^2 with a similar accuracy or precision as regression coefficients would need more Monte Carlo iterations.

We also examine the behavior of the estimators under misspecified distributions. In particular, we calculate the estimators assuming the normal random errors, while the data is generated using t or χ^2 distributions. In each case, the error term ε is normalized to have zero mean and variance σ_ε^2. The simulated mean estimates, the RMSE, and the biases for the sample size $n = 200$ and 31% censoring are presented in Table 2.24. As we can see from Table 2.24 that, under misspecification, both the MLE and the SLSOPT have relatively high biases. In contrast to previous simulation results, in this case, the SLSIDEN performs dramatically better in terms of bias for all the parameters and RMSE for σ_ε^2. This is due to the fact that in

Table 2.18: Simulation results of the models with sample size $n = 200$ and 31% censoring

Error	Normal	$t_{(5)}$	$\chi^2_{(4)}$
		$\beta_1 = -1.5$	
SLSIDEN	-1.4860	-1.4483	-1.4806
RMSE	0.7428	0.7754	0.8069
Accuracy	0.0461	0.0480	0.0500
95% C.I.	(-1.532,-1.440)	(-1.496,-1.400)	(-1.531,-1.431)
MLE	-1.5097	-1.5188	-1.5695
RMSE	0.4614	0.4298	0.4814
Accuracy	0.0286	0.0266	0.0295
95% C.I.	(-1.538,-1.481)	(-1.545,-1.492)	(-1.599,-1.540)
SLSOPT	-1.5016	-1.5155	-1.5472
RMSE	0.4749	0.4959	0.5770
Accuracy	0.0294	0.0307	0.0357
95% C.I.	(-1.531,-1.472)	(-1.546,-1.485)	(-1.583,-1.512)

Table 2.19: Simulation results of the models with sample size $n = 200$ and 31% censoring - continued

Error	Normal	$t_{(5)}$	$\chi^2_{(4)}$
		$\beta_2 = 1.5$	
SLSIDEN	1.4924	1.4869	1.4947
RMSE	0.1632	0.1538	0.1829
Accuracy	0.0101	0.0095	0.0113
95% C.I.	(1.482,1.503)	(1.477,1.496)	(1.483,1.506)
MLE	1.4999	1.5026	1.5219
RMSE	0.1091	0.0948	0.0923
Accuracy	0.0068	0.0059	0.0056
95% C.I.	(1.493,1.507)	(1.497,1.508)	(1.516,1.527)
SLSOPT	1.4988	1.5021	1.5092
RMSE	0.1106	0.1103	0.1161
Accuracy	0.0069	0.0068	0.0072
95% C.I.	(1.492,1.506)	(1.495,1.509)	(1.502,1.516)

Table 2.20: Simulation results of the models with sample size $n = 200$ and 31% censoring - continued

Error	Normal	$t_{(5)}$	$\chi^2_{(4)}$
		$\sigma_\varepsilon^2 = 16$	
SLSIDEN	15.9419	15.9363	15.9210
RMSE	0.6625	0.6878	0.7301
Accuracy	0.0409	0.0425	0.0450
95% C.I.	(15.901,15.983)	(15.894,15.979)	(15.876,15.966)
MLE	15.9249	15.9388	15.9287
RMSE	0.7835	0.7879	0.7549
Accuracy	0.0484	0.0487	0.0466
95% C.I.	(15.877,15.973)	(15.890,15.987)	(15.882,15.975)
SLSOPT	15.9075	15.8561	15.9408
RMSE	0.7830	0.7968	0.7995
Accuracy	0.0482	0.0486	0.0494
95% C.I.	(15.859,15.956)	(15.807,15.905)	(15.891,15.990)

Table 2.21: Simulation results of the models with sample size $n = 200$ and 60% censoring

Dist.	Normal	$t_{(5)}$	$\chi^2_{(4)}$
		$\beta_1 = -6$	
SLSIDEN	-5.9127	-5.9714	-5.9146
RMSE	0.8471	0.8548	0.8592
Accuracy	0.0523	0.0530	0.0530
95% C.I.	(-5.965,-5.860)	(-6.024,-5.918)	(-5.968,-5.862)
MLE	-6.0113	-6.0388	-6.1176
RMSE	0.5910	0.6093	0.6660
Accuracy	0.0366	0.0377	0.0407
95% C.I.	(-6.048,-5.975)	(-6.077,-6.001)	(-6.158,-6.077)
SLSOPT	-6.0279	-6.0418	-6.0361
RMSE	0.6102	0.6775	0.6992
Accuracy	0.0378	0.0419	0.0433
95% C.I.	(-6.066,-5.990)	(-6.084,-6.000)	(-6.079,-5.993)

Table 2.22: Simulation results of the models with sample size $n = 200$ and 60% censoring - continued

Dist.	Normal	$t_{(5)}$	$\chi^2_{(4)}$
		$\beta_2 = 1.5$	
SLSIDEN	1.4785	1.4957	1.4752
RMSE	0.2024	0.1796	0.2136
Accuracy	0.0125	0.0111	0.0132
95% C.I.	(1.466,1.491)	(1.485,1.507)	(1.462,1.488)
MLE	1.5028	1.5051	1.5287
RMSE	0.1325	0.1180	0.1086
Accuracy	0.0082	0.0073	0.0065
95% C.I.	(1.495,1.511)	(1.498,1.512)	(1.522,1.535)
SLSOPT	1.5059	1.5064	1.5073
RMSE	0.1346	0.1288	0.1271
Accuracy	0.0083	0.0080	0.0079
95% C.I.	(1.498,1.514)	(1.498,1.514)	(1.499,1.515)

Table 2.23: Simulation results of the models with sample size $n = 200$ and 60% censoring - continued

Dist.	Normal	$t_{(5)}$	$\chi^2_{(4)}$
		$\sigma_\varepsilon^2 = 16$	
SLSIDEN	15.9563	15.9154	16.0025
RMSE	0.7177	0.7166	0.7605
Accuracy	0.0444	0.0441	0.0472
95% C.I.	(15.912,16.001)	(15.871,15.960)	(15.955,16.050)
MLE	15.8526	15.8927	15.9100
RMSE	0.8185	0.8070	0.7935
Accuracy	0.0499	0.0496	0.0489
95% C.I.	(15.803,15.903)	(15.843,15.942)	(15.861,15.959)
SLSOPT	15.8393	15.8426	15.9053
RMSE	0.8120	0.7959	0.8141
Accuracy	0.0494	0.0484	0.0501
95% C.I.	(15.790,15.889)	(15.794,15.891)	(15.855,15.955)

Table 2.24: Simulation results of the misspecified models with sample size $n = 200$ and 31% censoring

	$t_{(5)}$			$\chi^2_{(4)}$		
	Mean	RMSE	Bias	Mean	RMSE	Bias
			$\beta_1 = -1.5$			
SLSIDEN	-1.7644	0.7664	0.2644	-1.3148	0.8223	0.1852
MLE	-1.7792	0.5409	0.2792	-1.8527	0.6024	0.3527
SLSOPT	-1.9290	0.6285	0.4290	-2.0917	0.7539	0.5917
			$\beta_2 = 1.5$			
SLSIDEN	1.5153	0.1583	0.0153	1.4671	0.2041	0.0329
MLE	1.5560	0.1204	0.0560	1.5430	0.1309	0.0430
SLSOPT	1.5748	0.1447	0.0748	1.5582	0.1377	0.0582
			$\sigma^2_\varepsilon = 16$			
SLSIDEN	15.7071	0.6819	0.2929	16.1360	0.6900	0.1360
MLE	15.2305	0.9006	0.7695	16.6462	0.8713	0.6462
SLSOPT	15.3186	0.8878	0.6814	16.7860	0.8889	0.7860

SLSIDEN, the weighting matrix does not depend on the parameters which are poorly estimated because of misspecification. Although SLSIDEN is not as efficient as MLE and SLSOPT in the correct specification cases, it shows more robustness in misspecified cases.

Chapter 3

Second-order Least Squares Estimation in Measurement Error Models

3.1 Linear Measurement Error models

Consider a linear regression model

$$Y = X'\beta + \varepsilon, \tag{3.1}$$

where $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$ are response and predictor variables respectively, $\beta \in \mathbb{R}^p$ is the unknown parameters, and ε is the random error. The observed predictor variable is

$$Z = X + \delta, \tag{3.2}$$

where δ is the random measurement error. We assume that X, δ, and ε are mutually uncorrelated and have means $\mu_x, 0, 0$ and covariances $\Sigma_x, \Sigma_\delta, \sigma_\varepsilon^2$

respectively, where Σ_x has full rank but Σ_δ can be singular to allow some components of X to be measured without error. In addition, suppose that an instrumental variable (IV) $V \in I\!R^q$ is available which is correlated with X but uncorrelated with δ and ε. We can see that (3.1) and (3.2) give a conventional linear errors-in-variables model

$$Y = Z'\beta + \varepsilon - \delta'\beta.$$

It is well-known that the ordinary least squares estimators for β based on (Z, Y) will be inconsistent because Z is correlated with the error term $\varepsilon - \delta'\beta$. Our goal is to find consistent estimators for parameters β and σ_ε^2 in model (3.1) – (3.2). The observed data is $(Z_i, V_i, Y_i), i = 1, 2, ..., n$, which are supposed to be independent but not necessarily identically distributed. Wang and Hsiao (2007) proposed a method that can yield consistent estimators, both for β and σ_ε^2. First, since $\Sigma_{vx} \neq 0$, we have

$$X = HV + U, \tag{3.3}$$

where U is uncorrelated with δ, ε and satisfies $E(VU') = 0$ by construction. Furthermore, because V is uncorrelated with U and δ, substituting (3.3) into (3.2) results in a standard linear regression equation

$$Z = HV + U + \delta. \tag{3.4}$$

It follows that H can be consistently estimated using the least squares method. On the other hand, substituting (3.3) into (3.1) we obtain

$$Y = V'\omega + \nu, \tag{3.5}$$

where $\omega = H'\beta$ and $\nu = \varepsilon + U'\beta$ is uncorrelated with V. Hence ω can be consistently estimated using the data (V_i, Y_i). We note that in (3.4) H can be estimated by

$$\hat{H} = \left(\sum_{i=1}^{n} Z_i V_i' \right) \left(\sum_{i=1}^{n} V_i V_i' \right)^{-1}. \tag{3.6}$$

We can see that ω in (3.5) can be estimated either by OLS or SLS. If we use OLS to estimate ω we have

$$\hat{\omega}_{OLS} = (VV')^{-1} VY. \tag{3.7}$$

The second-order least squares estimator $\hat{\gamma}_{SLS}$ for γ is defined as the measurable function that minimizes (2.2), where

$\gamma = (\omega', \sigma_\nu^2)'$, $\rho_i(\gamma) = (Y_i - V_i'\omega, Y_i^2 - (V_i'\omega)^2 - \sigma_\nu^2)'$, and $\sigma_\nu^2 = \beta'\Sigma_u\beta + \sigma_\varepsilon^2$.

Given the consistent estimators $\hat{H}$ and $\hat{\omega}$, consistent estimator for β can be obtained by minimizing $(\hat{\omega} - \hat{H}'\beta)' A_n (\hat{\omega} - \hat{H}'\beta)$, where A_n is a nonnegative definite weighting matrix which may depend on the data. The minimum distance estimator (MDE) is given by

$$\hat{\beta}_{OLS} = (\hat{H} A_n \hat{H}')^{-1} \hat{H} A_n \hat{\omega}_{OLS}. \tag{3.8}$$

Similarly,

$$\hat{\beta}_{SLS} = (\hat{H}A_n\hat{H}')^{-1}\hat{H}A_n\hat{\omega}_{SLS}. \tag{3.9}$$

Furthermore, since $\sqrt{n}\,(\hat{\omega}_{OLS} - \omega) \xrightarrow{L} N\,(0, \Sigma_{\hat{\omega}_{OLS}})$ (Wang and Leblanc (2007)), then the delta-method implies

$$\sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) \xrightarrow{L} N\left(0, (HAH')^{-1}HA\Sigma_{\hat{\omega}_{OLS}}AH'(HAH')^{-1}\right),$$

where $A = \text{plim}(A_n/n)$. Similarly for SLS we have $\sqrt{n}\,(\hat{\omega}_{SLS} - \omega) \xrightarrow{L} N\,(0, \Sigma_{\hat{\omega}_{SLS}})$, then

$$\sqrt{n}\left(\hat{\beta}_{SLS} - \beta\right) \xrightarrow{L} N\left(0, (HAH')^{-1}HA\Sigma_{\hat{\omega}_{SLS}}AH'(HAH')^{-1}\right).$$

Here $\Sigma_{\hat{\omega}_{OLS}} = \sigma_\nu^2 G_2^{-1}$ and

$$\Sigma_{\hat{\omega}_{SLS}} = \left(\sigma_\nu^2 - \frac{\mu_3^2}{\mu_4 - \sigma_{\delta_1}^4}\right)\left(G_2 - \frac{\mu_3^2}{\sigma_\nu^2(\mu_4 - \sigma_\nu^4)}G_1G_1'\right)^{-1},$$

where $G_1 = E(V)$, $G_2 = E(VV')$, $\mu_3 = E(\nu^3|V)$, and $\mu_4 = E(\nu^4|V)$. To estimate σ_ε^2, we have

$$\sigma_y^2 = \beta'\Sigma_{zy} + \sigma_\varepsilon^2,$$

hence $\hat{\sigma_\varepsilon^2} = S_{yy} - \hat{\beta}'S_{zy}$. The next theorem shows that under some conditions SLS estimator of ω dominates OLS estimator.

Theorem 3.1.1. *$V(\hat{\beta}_{OLS}) - V(\hat{\beta}_{SLS})$ is nonnegative definite if $\mu_3 \neq 0$ and $G_1'G_2^{-1}G_1 = 1$, and is positive definite if $G_1'G_2^{-1}G_1 \neq 1$.*

Proof. Let $B = AH'(HAH')^{-1}$, so for every nonzero vector y, we have

$$y'(V(\hat{\beta}_{OLS}) - V(\hat{\beta}_{SLS}))y = y'B'(V(\hat{\omega}_{OLS}) - V(\hat{\omega}_{SLS}))By = x'\Sigma_D x$$

Depends on either $\Sigma_D = V(\hat{\omega}_{OLS}) - V(\hat{\omega}_{OLS})$ is n.n.d. or p.d., $V(\hat{\beta}_{OLS}) - V(\hat{\beta}_{SLS})$ will be either n.n.d or p.d. (Wang and Leblanc (2007)), and $x = By$ is a nonzero vector.

□

In a simulation study, under the condition that the error terms have chi squared distribution, we compared the OLS estimation with SLS estimation. Table 3.1 shows the results of the study. It shows that in a moderate sample size the RMSE reduction was about 40%.

Table 3.1: Simulation results of the Linear Measurement error model with instrumental variable.

	n = 50	n = 70	n = 100	n = 200	n = 300	n = 400
			$\beta_1 = 0.5$			
OLS	0.6613	0.5356	0.5412	0.5048	0.5125	0.5038
RMSE	3.2698	0.5291	0.2500	0.1345	0.1083	0.0932
SLS	0.6356	0.5282	0.5311	0.5047	0.5116	0.5037
RMSE	2.3860	0.5027	0.2179	0.1176	0.0977	0.0831
			$\beta_2 = -0.5$			
OLS	-0.7058	-0.5504	-0.5460	-0.5105	-0.5085	-0.5100
RMSE	4.0054	0.7919	0.3952	0.1997	0.1611	0.1407
SLS	-0.6896	-0.5630	-0.5476	-0.5116	-0.5098	-0.5086
RMSE	3.5007	0.7536	0.3817	0.1822	0.1411	0.1235
			$\sigma_\varepsilon^2 = 1$			
OLS	0.8392	0.9217	0.9566	0.9728	0.9843	0.9900
RMSE	3.8355	0.8642	0.4054	0.2522	0.2027	0.1822
SLS	0.8719	0.9295	0.9714	0.9741	0.9877	0.9924
RMSE	2.3432	0.8264	0.4038	0.2425	0.1980	0.1767

3.2 Generalized Linear Measurement Error Models

3.2.1 Method of Moments Estimator

In a generalized linear model (GLM, McCullagh and Nelder 1989), the first two conditional moments of the response variable $Y \in \mathbb{R}$ given the covariates $X \in \mathbb{R}^p$ can be written as

$$\begin{aligned} E(Y|X) &= G^{-1}(X'\beta) \\ V(Y|X) &= \theta K(G^{-1}(X'\beta)), \end{aligned}$$

where $\beta \in \mathbb{R}^p$ and $\theta \in \mathbb{R}$ are unknown parameters, G is the link function, and K is a known function. It follows that

$$E(Y^2|X) = \theta K(G^{-1}(X'\beta)) + (G^{-1}(X'\beta))^2.$$

In order to have uniform notations with the second chapter, we consider the model of the form

$$E(Y \mid X) = f(X'\beta), \tag{3.10}$$

$$E(Y^2 \mid X) = f^2(X'\beta) + g^2(X'\beta, \theta). \tag{3.11}$$

Here we assume that $V(Y \mid X) = g^2(X'\beta, \theta)$. Furthermore, suppose that X is unobservable, instead we observe

$$Z = X + \delta, \tag{3.12}$$

where δ is a random measurement error. We also assume that an instrumental variable $V \in \mathbb{R}^q$ is available and is related to X through

$$X = HV + U, \tag{3.13}$$

where H is a $p \times q$ matrix of unknown parameters with rank p and U is independent of V has mean zero and distribution $f_U(u; \phi)$ with unknown parameters $\phi \in \Phi \subset \mathbb{R}^k$. In addition, assume that $E(Y^j|X) = E(Y^j|X, V)$, $j = 1, 2$ and the measurement error δ is independent of X, V, Y. There is no assumption concerning the functional forms of the distributions of X and δ. In this sense, model (3.10) – (3.13) is semi parametric. In this model the observed variables are (Y, Z, V). Our interest is to estimate $\gamma = (\beta', \theta, \phi')'$. We propose the method of moments estimators as follows.

First, substituting (3.13) into (3.12) results in a usual linear regression equation

$$E(Z \mid V) = HV. \tag{3.14}$$

Furthermore, by model assumptions we have

$$E(Y \mid V) = \int f(V'H'\beta' + u'\beta) f_U(u;\phi) du, \tag{3.15}$$

$$\begin{aligned} E(Y^2 \mid V) &= \int f^2(V'H'\beta' + u'\beta) f_U(u;\phi) du \\ &+ \int g^2(V'H'\beta' + u'\beta;\theta) f_U(u;\phi) du, \end{aligned} \tag{3.16}$$

and

$$E(YZ \mid V) = \int (HV + u) f(V'H'\beta' + u'\beta) f_U(u;\phi) du. \tag{3.17}$$

Throughout the section, all integrals are taken over the space $I\!R^p$. It follows that H can be consistently estimated by (3.14) and the least squares method, and β, θ, and ϕ can be consistently estimated using (3.15) – (3.17) and nonlinear least squares method, provided that they are identifiable by these equations. Indeed, the identifiability of the unknown parameters by these equations have been shown by Wang and Hsiao (1995) for the case where f is integrable and g is a constant function, Schennach (2007), and Wang and Hsiao (2007) for the more general cases. Since it is straightforward to estimate H using (3.14), in the following we assume that H is known and focus on the estimation of β, θ, and ϕ. In practice, one can estimate H using an external sample, or a subset of the main sample, and estimate other parameters using the rest of the main sample. Now we use some examples to demonstrate that the mentioned parameters may indeed be estimated using (3.15) – (3.17). To simplify notations, we consider the case where all variables are scalars and $U \sim N(0, \phi)$.

Example 3.2.1. Consider a Gamma loglinear model where Y has a continuous distribution with the first two conditional moments $E(Y \mid X) = \exp(\beta_1 + \beta_2 X)$ and $V(Y \mid X) = \theta \exp[2(\beta_1 + \beta_2 X)]$. Here θ is the dispersion parameter. This type of model has wide applications in finance, radio ligand assays, and kinetic reaction experiments. Using the assumptions for the model, we find

$$\begin{aligned} E(Y \mid V) &= E\left[\exp(\beta_1 + \beta_2 X) \mid V\right] \\ &= \exp(\beta_1) E\left[\exp(\beta_2(HV + U)) \mid V\right] \\ &= \exp\left(\beta_1 + \beta_2 HV + \frac{\beta_2^2 \phi}{2}\right), \end{aligned} \tag{3.18}$$

$$\begin{aligned} E(Y^2 \mid V) &= E\left[(\theta + 1)\exp(2(\beta_1 + \beta_2 X) \mid V)\right] \\ &= (\theta + 1)\exp(2\beta_1)\exp(2\beta_2 HV) E\left[\exp(2\beta_2 U \mid V)\right] \\ &= (\theta + 1)\exp\left[2(\beta_1 + \beta_2 HV + \beta_2^2 \phi)\right], \end{aligned} \tag{3.19}$$

and

$$
\begin{aligned}
E(YZ \mid V) &= \exp(\beta_1)E[(HV+U)\exp(\beta_2(HV+U)) \mid V] \\
&= \exp(\beta_1+\beta_2 HV)\,[HVE(\exp(\beta_2 U) \mid V) \\
&+ E(U\exp(\beta_2 U) \mid V)] \\
&= \exp\left(\beta_1+\beta_2 HV+\frac{\beta_2^2\phi}{2}\right)(HV+\beta_2\phi). \qquad (3.20)
\end{aligned}
$$

Example 3.2.2. Consider a Poisson loglinear model where Y is a count variable with moments $E(Y \mid X) = \exp(\beta_1+\beta_2 X)$ and $V(Y \mid X) = \exp(\beta_1 + \beta_2 X)$. This model has applications in biology, demographics, and survival analysis. For this model, we find

$$
\begin{aligned}
E(Y \mid V) &= E\left[\exp(\beta_1+\beta_2 X) \mid V\right] \\
&= \exp(\beta_1)E\left[\exp(\beta_2(HV+U)) \mid V\right] \\
&= \exp\left(\beta_1+\beta_2 HV+\frac{\beta_2^2\phi}{2}\right), \qquad (3.21)
\end{aligned}
$$

$$
\begin{aligned}
E(Y^2 \mid V) &= E(\exp(\beta_1+\beta_2 X \mid V)) + E(\exp(2(\beta_1+\beta_2 X \mid V))) \qquad (3.22) \\
&= \exp\left[2(\beta_1+\beta_2 HV+\beta_2^2\phi)\right] + \exp\left(\beta_1+\beta_2 HV+\frac{\beta_2^2\phi}{2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
E(YZ \mid V) &= \exp(\beta_1) E[(HV+U)\exp(\beta_2(HV+U)) \mid V] \\
&= \exp(\beta_1 + \beta_2 HV)\,[HV E(\exp(\beta_2 U) \mid V) \\
&+ E(U \exp(\beta_2 U) \mid V)] \\
&= \exp\left(\beta_1 + \beta_2 HV + \frac{\beta_2^2 \phi}{2}\right)(HV + \beta_2 \phi). \qquad (3.23)
\end{aligned}
$$

Example 3.2.3. Consider a logistic model where

$$E(Y \mid X) = 1/[1 + \exp(-\beta_1 - \beta_2 X)]$$

and

$$V(Y \mid X) = \exp(-\beta_1 - \beta_2 X)/[1 + \exp(-\beta_1 - \beta_2 X)]^2.$$

In this model, Y is binary and its second moment is equal to the mean. Logistic regression has been used extensively in medical and social sciences as well as marketing applications. Using the assumptions for the model, we find

$$
\begin{aligned}
E(Y \mid V) &= E(Y^2 \mid V) \\
&= \frac{1}{\sqrt{2\pi\phi}} \int \frac{\exp(-u^2/2\phi)du}{[1 + \exp(-\beta_1 - \beta_2 HV - \beta_2 u)]} \qquad (3.24)
\end{aligned}
$$

and

$$E(YZ \mid V) = \frac{1}{\sqrt{2\pi\phi}} \int \frac{(HV+u)\exp(-u^2/2\phi)du}{[1 + \exp(-\beta_1 - \beta_2 HV - \beta_2 u)]}. \qquad (3.25)$$

These moments do not have closed forms. However, the unknown parameters can still be estimated by the simulation-based approach in Section 3.2.2.

We define $\gamma = (\beta', \theta, \phi')'$ and the parameter space to be $\Gamma = \Omega \times \Theta \times \Phi$. The true parameter value of the model is denoted by $\gamma_0 \in \Gamma$. For every $\tau \in I\!R^p$ and $\gamma \in \Gamma$, define $m(\tau;\gamma) = (m_1(\tau;\gamma), m_2(\tau;\gamma), m_3(\tau;\gamma))'$, where

$$m_1(\tau;\gamma) = \int f(x'\beta) f_U(x-\tau;\phi)dx, \tag{3.26}$$

$$m_2(\tau;\gamma) = \int f^2(x'\beta) f_U(x-\tau;\phi)dx + \int g^2(x'\beta;\theta) f_U(x-\tau;\phi)dx, \tag{3.27}$$

and

$$m_3(\tau;\gamma) = \int x f(x'\beta) f_U(x-\tau;\phi)dx. \tag{3.28}$$

Suppose $T_i = (Y_i, Z_i, V_i), i = 1, 2, ..., n$, is an *i.i.d.* random sample, and $\rho_i(\gamma) = Y_i T_i - m(HV_i;\gamma)$, then the method of moments estimator (MME) for γ is defined as

$$\hat{\gamma}_n = \text{argmin}_{\gamma\in\Gamma} Q_n(\gamma) = \text{argmin}_{\gamma\in\Gamma} \sum_{i=1}^{n} \rho_i'(\gamma) W_i \rho_i(\gamma), \tag{3.29}$$

where $W_i = W(V_i)$ is a nonnegative definite matrix which may depend on V_i. Note that for the binary response Y, we have $E(Y|V) = E(Y^2|V)$. In this case the first two elements of $\rho_i(\gamma)$ are identical. This redundancy can be eliminated by setting the first row and first column of $W(V_i)$ to zero. However, to simplify presentation in the following we present results in terms of general response Y.

3.2.2 Consistency and Asymptotic Normality

To prove the consistency of $\hat{\gamma}_n$, we assume the following regularity conditions, where μ denotes Lebesgue measure.

Assumption 15. *The parameter space* $\Gamma = \Omega \times \Theta \times \Phi$ *is compact in* $\mathbb{R}^{p+k+1}$, *and also parameter space is a compact subset of* $\mathbb{R}^{pq}$, *containing the true value* $\psi_0 = vecH_0$. *Furthermore,* $E\|W(V)\|(\|YZ\|^2 + Y^4) < \infty$, *and* EVV' *is nonsingular.*

Assumption 16. $f(x'\beta)$ *and* $g(x'\beta;\theta)$ *are measurable functions of* x *for each* $\beta \in \Omega$ *and* $\theta \in \Theta$. *Furthermore,* $f^2(x'\beta)f_U(x-\tau;\phi)$ *and* $g^2(x'\beta;\theta)f_U(x-\tau;\phi)$ *are uniformly bounded by a function* $K(x,v)$, *which satisfies* $E\|W(V)\|\left(\int K(x,V)(\|x\|+1)dx\right)^2 < \infty$.

Assumption 17. $E[\rho(\gamma) - \rho(\gamma_0)]'W(V)[\rho(\gamma) - \rho(\gamma_0)] = 0$ *if and only if* $\gamma = \gamma_0$, *where* $\rho(\gamma) = YT - m(HV;\gamma)$.

Theorem 3.2.4. *Under* Assumptions *15 – 17,* $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, *as* $n \to \infty$.

To derive the asymptotic normality for $\hat{\gamma}_n$, we assume additional regularity conditions as follows.

Assumption 18. *There exists an open subset* $\gamma_0 \in \Gamma_0 \subset \Gamma$, *in which the first two partial derivatives of* $f^2(x'\beta)f_U(x-\tau;\phi)$ *and* $g^2(x'\beta;\theta)f_U(x-\tau;\phi)$ *w.r.t.* γ *and* ψ *are uniformly bounded by a function* $K(x,v)$, *which satisfies* $E\,\|W(V)\|\left(\|V\|\int K(x,V)(\|x\|+1)dx\right)^2 < \infty$.

Assumption 19. *The matrix*

$$B = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(V) \frac{\partial \rho(\gamma_0)}{\partial \gamma'}\right]$$

is nonsingular.

These assumptions are equivalent to the ones that are needed for consistency and asymptotic normality of an M-estimator (see e.g., van der Vaart 2000, Sections 5.2 and 5.3).

Theorem 3.2.5. *Under* Assumptions *15 – 19,*
$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{L} N(0, B^{-1}AB^{-1})$ *as* $n \to \infty$, *where*

$$A = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(V)\rho(\gamma_0)\rho'(\gamma_0)W(V)\frac{\partial \rho(\gamma_0)}{\partial \gamma'}\right].$$

When H is unknown, we can estimate it using

$$\hat{H} = \left(\sum_{i=1}^{n} Z_i V_i'\right)\left(\sum_{i=1}^{n} V_i V_i'\right)^{-1}. \tag{3.30}$$

In that case the method of moments estimator for γ is defined as $\hat{\gamma}_n = \text{argmin}_{\gamma\in\Gamma} Q_n(\gamma)$, where

$$Q_n(\gamma) = \sum_{i=1}^{n} \hat{\rho}_i'(\gamma) W_i \hat{\rho}_i(\gamma), \tag{3.31}$$

and $\hat{\rho}_i(\gamma) = Y_i T_i - m(\hat{H}V_i;\gamma)$, then Theorem 3.2.5 can be modified to the following theorem.

Theorem 3.2.6. *Under* ASSUMPTIONS *15 - 19, as* $n \to \infty$, $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{L} N(0, B^{-1}CAC'B^{-1})$, *where*

$$C = \left[I_{p+k+1}, E\left(\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(V) \frac{\partial \rho(\gamma_0)}{\partial \psi'} \right) (EVV' \otimes I_p)^{-1} \right],$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix},$$

$$A_{11} = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(V) \rho(\gamma_0) \rho'(\gamma_0) W(V) \frac{\partial \rho(\gamma_0)}{\partial \gamma'} \right],$$

$$A_{12} = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(V) \rho(\gamma_0) \left((Z - HV)' \otimes V' \right) \right]$$

and $A_{22} = E\left[VV' \otimes (Z - HV)(Z - HV)' \right]$.

The asymptotic covariance of $\hat{\gamma}_n$ depends on the weight $W(V)$. A natural question is how to choose $W(V)$ to obtain the most efficient estimator. To answer this question, we first write $C = (I_{p+k+1}, G)$, where $G =$

$E\left(\frac{\partial\rho'(_0)}{\partial\gamma}W(V)\frac{\partial\rho(\gamma_0)}{\partial\psi'}\right)(EWW'\otimes I_p)^{-1}$. Therefore, $CAC' = A_{11} + GA'_{12} + A_{12}G' + GA_{22}G'$. Since

$$\frac{\partial\rho(\gamma)}{\partial\psi'} = -\frac{\partial m(HV;\gamma)}{\partial\psi'},$$

and

$$\begin{aligned}
\frac{\partial m_1(HV;\gamma)}{\partial\psi'} &= -\int f(x'\beta)\frac{\partial f_U(x-HV;\phi)}{\partial u'}dx(V\otimes I_p)', \\
\frac{\partial m_2(HV;\gamma)}{\partial\psi'} &= -\int f^2(x'\beta)\frac{\partial f_U(x-HV;\phi)}{\partial u'}dx(V\otimes I_p)' \\
&- \int g^2(x'\beta;\theta)\frac{\partial f_U(x-HV;\phi)}{\partial u'}dx(V\otimes I_p)',
\end{aligned}$$

$$\frac{\partial m_3(HV;\gamma)}{\partial\psi'} = -\int xf(x'\beta)\frac{\partial f_U(x-HV;\phi)}{\partial u'}dx(V\otimes I_p)',$$

the last three terms in CAC' are due to the least squares estimation of ψ. To simplify discussion, assume for the moment that ψ is known, so that these three terms do not appear in CAC'. The following discussion remains valid, when ψ is unknown and estimated using a part of the sample (Y_i, Z_i, V_i), $i = 1, 2, ..., n$, while $Q_n(\gamma)$ is constructed using the rest of the sample points. Then the independence of the sample points implies that $A_{12} = 0$. Since $\partial\rho'(\gamma_0)/\partial\gamma$ depends on V only, matrix A_{11} can be written as

$$A_{11} = E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}W(V)FW(V)\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right],$$

where $F = F(V) = E[\rho(\gamma_0)\rho'(\gamma_0)|V]$. Then, analogous to the weighted (nonlinear) least squares estimation, we have

$$B^{-1}A_{11}B^{-1} \geq E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}F^{-1}\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]^{-1} \tag{3.32}$$

(in the sense that the difference of the left-hand and right-hand sides is nonnegative definite), and the lower bound is attained for $W = F^{-1}$ in both B and A_{11} (Hansen (1982), Abarin and Wang 2006).

In practice, F is a function of unknown parameters and therefore needs to be estimated. This can be done using the following two-stage procedure. First, minimize $Q_n(\gamma)$ with identity matrix $W = I_{p+2}$ to obtain the first-stage estimator $\hat{\gamma}_n$. Secondly, estimate F by $\hat{F} = \frac{1}{n}\sum_{i=1}^{n}\rho_i(\hat{\gamma}_n)\rho_i'(\hat{\gamma}_n)$ or alternatively by a nonparametric estimator, and then minimize $Q_n(\gamma)$ again with $W = \hat{F}^{-1}$ to obtain the second-stage estimator $\hat{\hat{\gamma}}_n$. Since $\hat{F}$ is consistent for F, the asymptotic covariance of $\hat{\hat{\gamma}}_n$ is given by the right hand side of (3.32). Consequently $\hat{\hat{\gamma}}_n$ is asymptotically more efficient than the first-stage estimator $\hat{\gamma}_n$.

3.2.3 Simulation-Based Estimator

When the explicit form of $m(\tau;\gamma)$ exists, the numerical computation of MME $\hat{\gamma}_n$ can be done using the usual optimization methods. However, sometimes the integrals in (3.26) – (3.28) do not have explicit forms. In this section, we use a simulation-based approach to overcome this problem. The simulation-based approach is used to approximate the multiple integrals in which they are simulated by Monte Carlo methods such as importance sampling.

We start with choosing a known density $l(x)$ and generate independent random points $\{x_{ij}, j = 1, 2, ..., 2S, i = 1, 2, ..., n\}$ from $l(x)$. Then we can approximate $m(HV_i; \gamma)$ by Monte Carlo simulators

$$m_{1,S}(HV_i;\gamma) = \frac{1}{S}\sum_{j=1}^{S}\frac{f(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \quad (3.33)$$

$$\begin{aligned} m_{2,S}(HV_i;\gamma) &= \frac{1}{S}\sum_{j=1}^{S}\frac{f^2(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \\ &+ \frac{1}{S}\sum_{j=1}^{S}\frac{g^2(x'_{ij}\beta;\theta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \end{aligned} \quad (3.34)$$

$$m_{3,S}(HV_i;\gamma) = \frac{1}{S}\sum_{j=1}^{S}\frac{x_{ij}f(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \quad (3.35)$$

and

$$m_{1,2S}(HV_i;\gamma) = \frac{1}{S}\sum_{j=S+1}^{2S}\frac{f(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \quad (3.36)$$

$$\begin{aligned} m_{2,2S}(HV_i;\gamma) &= \frac{1}{S}\sum_{j=S+1}^{2S}\frac{f^2(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \\ &+ \frac{1}{S}\sum_{j=S+1}^{2S}\frac{g^2(x'_{ij}\beta;\theta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})} \end{aligned} \quad (3.37)$$

$$m_{3,2S}(HV_i;\gamma) \;=\; \frac{1}{S}\sum_{j=S+1}^{2S}\frac{x_{ij}f(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)}{l(x_{ij})}. \qquad (3.38)$$

Finally, the Simulation-Based Estimator (SBE) for γ is defined by

$$\hat{\gamma}_{n,S} = \text{argmin}_{\gamma\in\Gamma} Q_{n,S}(\gamma) = \text{argmin}_{\gamma\in\Gamma}\sum_{i=1}^{n}\rho'_{i,S}(\gamma)W_i\rho_{i,2S}(\gamma), \qquad (3.39)$$

where $\rho_{i,S}(\gamma) = Y_iT_i - m_S(HV_i;\gamma)$ and $\rho_{i,2S}(\gamma) = Y_iT_i - m_{2S}(HV_i;\gamma)$.

Here we define

$$m_S(HV_i;\gamma) = (m_{1,S}(HV_i;\gamma), m_{2,S}(HV_i;\gamma), m_{3,S}(HV_i;\gamma))'$$

and

$$m_{2S}(HV_i;\gamma) = (m_{1,2S}(HV_i;\gamma), m_{2,2S}(HV_i;\gamma), m_{3,2S}(HV_i;\gamma))'.$$

We notice that by construction,

$E[m_S(HV_i;\gamma)|V_i] = E[m_{2S}(HV_i;\gamma)|V_i] = m(HV_i;\gamma)$, and therefore $m_S(HV_i;\gamma)$ and $m_{2S}(HV_i;\gamma)$ are unbiased simulators for $m(HV_i;\gamma)$. Moreover, $Q_{n,S}(\gamma)$ is an unbiased simulator for $Q_n(\gamma)$, because $Q_{n,S}(\gamma)$ and $Q_n(\gamma)$ have the same conditional expectations given the sample (Y_i, Z_i, V_i).

Alternatives to (3.33) – (3.38) which generally yield more stable estimates are

$$m'_{1,S}(HV_i;\gamma) \;=\; \sum_{j=1}^{S}\frac{f(x'_{ij}\beta)f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=1}^{S}f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \qquad (3.40)$$

$$
\begin{aligned}
m'_{2,S}(HV_i;\gamma) &= \sum_{j=1}^{S} \frac{f^2(x'_{ij}\beta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=1}^{S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \\
&+ \sum_{j=1}^{S} \frac{g^2(x'_{ij}\beta;\theta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=1}^{S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \qquad (3.41)
\end{aligned}
$$

$$
m'_{3,S}(HV_i;\gamma) = \sum_{j=1}^{S} \frac{x_{ij} f(x'_{ij}\beta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=1}^{S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \qquad (3.42)
$$

and

$$
m'_{1,2S}(HV_i;\gamma) = \sum_{j=S+1}^{2S} \frac{f(x'_{ij}\beta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=S+1}^{2S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \qquad (3.43)
$$

$$
\begin{aligned}
m'_{2,2S}(HV_i;\gamma) &= \sum_{j=S+1}^{2S} \frac{f^2(x'_{ij}\beta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=S+1}^{2S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \\
&+ \sum_{j=S+1}^{2S} \frac{g^2(x'_{ij}\beta;\theta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=S+1}^{2S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})} \qquad (3.44)
\end{aligned}
$$

$$
m'_{3,2S}(HV_i;\gamma) = \sum_{j=S+1}^{2S} \frac{x_{ij} f(x'_{ij}\beta) f_U(x_{ij}-HV_i;\phi)/l(x_{ij})}{\sum_{j=S+1}^{2S} f_U(x_{ij}-HV_j;\phi)/l(x_{ij})}, \qquad (3.45)
$$

where we have replaced S with the sum of the weights.

Since $(1/S)\sum_{j=S+1}^{2S} f_U(x_{ij}-HV_i;\phi)/l(x_{ij}) \approx 1$, when S is large enough, $m'_S(HV_i;\gamma) \approx m_S(HV_i;\gamma)$ and $m'_{2S}(HV_i;\gamma) \approx m_{2S}(HV_i;\gamma)$. Although these simulators are biased, the biases are small and the improvement in variance makes them preferred alternatives to (3.33) – (3.38). (See Lemma 4.3 of Robert and Cassella 2004.)

Theorem 3.2.7. *Suppose that the support of* $l(x)$ *covers the support of* $g^2(x'\beta;\theta)f_U(x-\tau;\phi)$ *and* $f^2(x'\beta)f_U(x-\tau;\phi)$, *for all* $\tau \in I\!R^p$ *and* $\gamma \in \Gamma$. *Then the simulation estimator* $\hat{\gamma}_{n,S}$ *has the following properties:*

(i). Under ASSUMPTIONS *15 – 17,* $\hat{\gamma}_{n,S} \xrightarrow{a.s.} \gamma_0$, *as* $n \to \infty$.

(ii). Under ASSUMPTIONS *15 – 19,* $\sqrt{n}(\hat{\gamma}_{n,S}-\gamma_0) \xrightarrow{L} N(0, B^{-1}A_S B^{-1})$,

where

$$\begin{aligned} 2A_S &= E\left[\frac{\partial \rho'_{1,S}(\gamma_0)}{\partial \gamma} W \rho_{1,2S}(\gamma_0)\rho'_{1,2S}(\gamma_0) W \frac{\partial \rho_{1,S}(\gamma_0)}{\partial \gamma'}\right] \\ &+ E\left[\frac{\partial \rho'_{1,S}(\gamma_0)}{\partial \gamma} W \rho_{1,2S}(\gamma_0)\rho'_{1,S}(\gamma_0) W \frac{\partial \rho_{1,2S}(\gamma_0)}{\partial \gamma'}\right]. \end{aligned}$$

When H is unknown and estimated using a part of the sample (Y_i, Z_i, V_i), $i = 1, 2, ..., n$, while $Q_n(\gamma)$ is constructed using the rest of the sample points, we can still define the simulation estimator simply by replacing H with $\hat{H}$. In this case, Theorem 3.2.7 will be modified to the following theorem.

Theorem 3.2.8. *Suppose that the support of* $l(x)$ *covers the support of* $g^2(x'\beta;\theta)f_U(x-\tau;\phi)$ *and* $f^2(x'\beta)f_U(x-\tau;\phi)$, *for all* $\tau \in I\!R^p$ *and* $\gamma \in \Gamma$. *Then the simulation estimator* $\hat{\gamma}_{n,S}$ *has the following properties:*

(i). Under ASSUMPTIONS *15 – 17,* $\hat{\gamma}_{n,S} \xrightarrow{a.s.} \gamma_0$, *as* $n \to \infty$.

(ii). Under ASSUMPTION *15 - 19,* $\sqrt{n}(\hat{\gamma}_{n,S}-\gamma_0) \xrightarrow{L} N(0, B^{-1}CA_SC'B^{-1})$*,*

where

$$A_S = \begin{pmatrix} A_{S,11} & A_{S,12} \\ A'_{S,12} & A_{S,22} \end{pmatrix},$$

$$C = \left[I_{p+k+1}, E\left(\frac{\partial\rho'(\gamma_0)}{\partial\gamma} W(V) \frac{\partial\rho(\gamma_0)}{\partial\psi'}\right)(EVV' \otimes I_p)^{-1} \right],$$

and

$$\begin{aligned} 2A_{S,11} &= E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma} W\rho_{1,2S}(\gamma_0)\rho'_{1,2S}(\gamma_0) W \frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'}\right] \\ &+ E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma} W\rho_{1,2S}(\Gamma_0)\rho'_{1,S}(\psi_0) W \frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right]. \end{aligned}$$

Although asymptotically, the importance density $l(x)$ does not have an effect on the efficiency of $\hat{\gamma}_{n,S}$, however, the choice of $l(x)$ will affect the finite sample variances of the simulators $m_S(HV_i;\gamma)$ and $m_{2S}(HV_i;\gamma)$. In addition, Wang (2004) showed that the efficiency loss caused by simulation is of magnitude $O(1/S)$.

3.2.4 Simulation Studies

In this section we present simulation studies on three generalized linear models of Examples 3.2.1 – 3.2.3, to demonstrate how the proposed estimators can be calculated and their performance in finite sample sizes.

First, consider the **Gamma loglinear model** in Example 3.2.1 where $U \sim N(0, \phi)$. We calculated conditional moments (3.18) - (3.20) for this model. Therefore, the method of moments estimators (MME) can be computed by minimizing $Q_n(\gamma)$ in (3.29). Specifically, the estimators are computed in two steps, using the identity and the estimated optimal weighting matrix, respectively.

To compute the Simulation-Based Estimators (SBE), we choose the density of $N(0, 2)$ to be $l(x_{ij})$, and generate independent points x_{ij} using $S = 2000$.

Furthermore, the simulated moments $m_S(HV; \gamma)$ and $m_{2S}(HV; \gamma)$, and $m'_S(HV; \gamma)$ and $m'_{2S}(HV_i; \gamma)$ are calculated according to (3.33) – (3.38), and (3.40) – (3.45) respectively. The two-step SBE1 $\hat{\gamma}_{n,s}$ is calculated by minimizing $Q_{n,S}(\gamma)$, using the identity and the estimated optimal weighting matrix, and $m_S(HV; \gamma)$ and $m_{2S}(HV; \gamma)$. Similarly the two-step SBE2 $\hat{\gamma}_{n,s}$ is calculated by minimizing $Q_{n,S}(\gamma)$, using the identity and the estimated optimal weighting matrix, and $m'_S(HV; \gamma)$ and $m'_{2S}(HV; \gamma)$. The data have been generated using V and δ from a standard normal distribution, and parameter values $\beta_1 = -0.2, \beta_2 = 0.3, \phi = 0.8, \theta = 0.1$. We also generated the response variable from a Gamma distribution with parameters $1/\theta$

Table 3.2: Simulation results of the Gamma Loglinear model with Measurement error in covariate

	$\beta_1 = -0.2$	$\beta_2 = 0.3$	$\phi = 0.8$	$\theta = 0.1$
MME	-0.206	0.299	0.810	0.101
RMSE	0.045	0.045	0.182	0.032
Accuracy	0.003	0.003	0.012	0.002
95% C.I.	(-0.209,-0.203)	(0.297,0.301)	(0.798,0.821)	(0.089,0.103)
SBE1	-0.206	0.299	0.814	0.101
RMSE	0.045	0.045	0.184	0.032
Accuracy	0.003	0.003	0.013	0.002
95% C.I.	(-0.209,-0.203)	(0.297,0.301)	(0.803,0.826)	(0.089,0.102)

and $\theta \exp(\beta_1 + \beta_2 X)$, respectively. In the simulation, we assumed that H is known. $N = 1000$ Monte Carlo replications have been carried out and in each replication, $n = 400$ sample points (Y_i, Z_i, V_i) have been generated. The computation has been done using MATLAB on a workstation running LINUX operating system.

Tables 3.2 and 3.3 show the summaries of the results for the Gamma

Table 3.3: Simulation results of the Gamma Loglinear model with Measurement error in covariate - continued

	$\beta_1 = -0.2$	$\beta_2 = 0.3$	$\phi = 0.8$	$\theta = 0.1$
SBE2	-0.206	0.298	0.815	0.101
RMSE	0.045	0.045	0.182	0.032
Accuracy	0.003	0.003	0.012	0.002
95% C.I.	(-0.209,-0.203)	(0.296,0.300)	(0.803,0.826)	(0.089,0.102)
z_1	0.450	0.515	0.312	0.058
z_2	0.223	0.134	1.355	0.333

Loglinear model. As we can see, estimators show their asymptotic properties. However, both SBE1 and SBE2 converge slower for ϕ. The tables present the accuracy corresponding to each estimation value, using $A = z_{\alpha/2}\sqrt{Var(\hat{\theta})/R}$ and with confidence 95%, where θ is the notation for a parameter in general. They show that the three estimators have the same amount of accuracy. The tables also show 95% confidence intervals for the parameters. They show that the true value of parameters fall in the intervals for β_2, ϕ, and θ, although because of the larger bias of the estimators, this is not the case for SBE1 and SBE2. The same results for β_1 is not satisfying and it shows that estimators have finite sample bias.

We used the root mean squared errors (RMSE) to compare the efficiency of the estimators. There is no significant difference between the estimators efficiencies.

To test the difference between MME and SBE1, we calculated z statistic for the test. As we can see in the tables, z_1 shows that there is no significant difference between MME and SBE1 for each parameter. Similarly, we calculated z_2 to test the difference between SBE1 and SBE2. The results show that there is no significant difference between them.

Next, we consider the Poisson loglinear model in Example 3.2.2, where $U \sim N(0, \phi)$. We calculated conditional moments (3.21) – (3.23) for this model. Parameters are $\beta_1 = -0.2$, $\beta_2 = 0.3$, and $\phi = 1$, and we choose the density of $N(0, 2)$ to be $l(x_{ij})$. We generated the response variable from a Poisson distribution with parameter $\exp(\beta_1 + \beta_2 X)$.

Tables 3.4 – 3.5 show the summaries of the results for the Poisson Loglinear model. As we can see in the tables, all the estimators show their asymptotic properties. All the estimators show the same accuracy and 95% confidence intervals for the parameters. The results show that the true value of parameters fall in the intervals. The root mean squared errors (RMSE) of the estimates show that MME and both SB estimators perform equally, for all the parameters. Furthermore, the test statistic values imply that there is no significant difference between the estimators.

Finally, we consider the Logistic model in Example 3.2.3, where $U \sim N(0, \phi)$. For this model, conditional moments (3.24) – (3.25) do not have closed forms. Therefore, we were only able to compute SBE1 and SBE2. Parameter values for this model are $\beta_1 = 0.5$, $\beta_2 = 0.3$, and $\phi = 0.8$, and we choose the density of $N(0, 2)$ to be $l(x_{ij})$. We generated the response variable from a Binary distribution with parameter $(1 + \exp(-(\beta_1 + \beta_2 X)))^{-1}$.

Table 3.4: Simulation results of the Poisson Loglinear model with Measurement error in covariate

	$\beta_1 = -0.2$	$\beta_2 = 0.3$	$\phi = 1$
MME	-0.201	0.295	0.998
RMSE	0.055	0.071	0.030
Accuracy	0.004	0.005	0.173
95% C.I.	(-0.205,-0.197)	(0.291,0.300)	(0.987,1.011)
SBE1	-0.201	0.296	0.999
RMSE	0.055	0.071	0.176
Accuracy	0.004	0.005	0.011
95% C.I.	(-0.205,-0.197)	(0.291,0.300)	(0.987,1.011)

Table 3.5: Simulation results of the Poisson Loglinear model with Measurement error in covariate - continued

	$\beta_1 = -0.2$	$\beta_2 = 0.3$	$\phi = 1$
SBE2	-0.201	0.295	0.999
RMSE	0.055	0.071	0.173
Accuracy	0.004	0.005	0.011
95% C.I.	(-0.205,-0.197)	(0.291,0.300)	(0.987,1.011)
z_1	0.281	0.641	0.445
z_2	0.224	0.570	0.411

Table 3.6: Simulation results of the Logistic model with Measurement error in covariate

	$\beta_1 = 0.5$	$\beta_2 = 0.3$	$\phi = 0.8$
SBE1	0.506	0.306	0.811
RMSE	0.134	0.118	0.192
Accuracy	0.008	0.007	0.015
95% C.I.	(0.498,0.515)	(0.299, 0.313)	(0.795,0.827)
SBE2	0.507	0.307	0.809
RMSE	0.130	0.118	0.187
Accuracy	0.008	0.007	0.015
95% C.I.	(0.499,0.515)	(0.299,0.314)	(0.795,0.823)
z	0.515	0.294	0.446

Table 3.6 shows the summaries of the results for the Logistic model. All the estimators show their asymptotic properties. Both estimators show the same accuracy and 95% confidence intervals for the parameters. The results show that the true value of parameters fall in the intervals. The root mean squared errors (RMSE) of the estimates show that SBE1 and SBE2 perform equally well, except for slightly more efficiency in SBE2, for ϕ. Furthermore, the test statistic values imply that there is no significant difference between the estimators.

Chapter 4

Related Issues

4.1 Computational Issues

4.1.1 Optimization Methods

Considering model (2.1), we know that the second-order least squares estimator (SLSE) $\hat{\gamma}_{SLS}$ for γ is defined as the measurable function that minimizes

$$Q_n(\gamma) = \sum_{i=1}^{n} \rho_i'(\gamma) W_i \rho_i(\gamma). \tag{4.1}$$

The first derivative of $Q_n(\gamma)$ exists and is given by

$$\frac{\partial Q_n(\gamma)}{\partial \gamma} = 2 \sum_{i=1}^{n} \frac{\partial \rho_i'(\gamma)}{\partial \gamma} W_i \rho_i(\gamma),$$

and the second derivative of $Q_n(\gamma)$ is given by

$$\frac{\partial^2 Q_n(\gamma)}{\partial \gamma \partial \gamma'} = 2 \sum_{i=1}^{n} \left[\frac{\partial \rho_i'(\gamma)}{\partial \gamma} W_i \frac{\partial \rho_i(\gamma)}{\partial \gamma'} + (\rho_i'(\gamma) W_i \otimes I_{p+q+1}) \frac{\partial \mathrm{vec}(\partial \rho_i'(\gamma)/\partial \gamma)}{\partial \gamma'} \right]. \tag{4.2}$$

There are numerous optimization methods which are suitable for nonlinear objective functions. Under some regularity conditions, we know that the first and the second derivatives of $Q_n(\gamma)$ exist. Therefore, we start with the methods based on the gradient of the objective function.

(i). **Gradient Descent (Steepest Descent) Method**

This method is based on the observation that if the real-valued function $Q_n(\gamma)$ is defined and differentiable in a neighborhood of a point $\hat{\gamma}_0$, then $Q_n(\gamma)$ decreases fastest if one goes from $\hat{\gamma}_0$ in the direction of the negative gradient of $Q_n(\gamma)$ at $\hat{\gamma}_0$. It follows that if $\hat{\gamma}_1 = \hat{\gamma}_0 - \varepsilon_0 \frac{\partial Q_n(\hat{\gamma}_0)}{\partial \gamma}$ for $\varepsilon_0 > 0$, then $Q_n(\hat{\gamma}_0) \geq Q_n(\hat{\gamma}_1)$. If one starts with a guess $\hat{\gamma}_0$ for a local minimum of $Q_n(\gamma)$, given the nth iterate $\hat{\gamma}_n$, the $(n+1)$th iterate will be

$$\hat{\gamma}_{n+1} = \hat{\gamma}_n - \varepsilon_n \frac{\partial Q_n(\hat{\gamma}_n)}{\partial \gamma}, n \in N.$$

The iteration will be repeated until convergence is achieved. The idea of this method is simple and it works in space of any number of dimensions. However, the algorithm can take many iterations to converge towards a local minimum, if the curvature in different directions is very different. Since the value of the step size ε changes at every iteration, finding the optimal ε per step can be time-consuming. Conversely, using a fixed ε can yield poor results.

(ii). **Gauss-Newton Method**

This method is used for minimization problems for which the objective function is a sum of squares. Because of this property, this method is used to solve nonlinear least squares problems. It is a modification of Newton's method that does not use second derivatives. The method starts with linearization of $f(X_i;\beta)$ by a Taylor series around the initial estimate $\hat{\beta}_0$ as

$$f(X_i;\beta) \approx f(X_i;\hat{\beta}_0) + \frac{\partial f(X_i;\hat{\beta}_0)}{\partial \beta}(\beta - \hat{\beta}_0).$$

Then using least squares method, the next estimator $\hat{\beta}_1$ is attained by minimizing $\sum_{i=1}^{n}[Y_i - \frac{\partial f(X_i;\hat{\beta}_0)}{\partial \beta}(\beta - \hat{\beta}_0)]^2$ with respect to β as

$$\hat{\beta}_1 = \hat{\beta}_0 + [\sum_{i=1}^{n} \frac{\partial f(X_i;\hat{\beta}_0)}{\partial \beta} \frac{\partial f(X_i;\hat{\beta}_0)}{\partial \beta'}]^{-1} \sum_{i=1}^{n}(Y_i - f(X_i;\hat{\beta}_0)) \frac{\partial f(X_i;\hat{\beta}_0)}{\partial \beta},$$

and like the previous method, the iteration will be repeated until convergence is achieved.

Although Gauss-Newton Method is a popular method for least squares problems, it can not be applied to SLS method. The objective function

(4.1) can not be expressed as a sum of squares. The quadratic form of $Q_n(\gamma)$ is

$$\sum_{i=1}^{n} \left[W_{i11}(Y_i - f(X;\beta))^2 + W_{i22}(Y_i^2 - f^2(X;\beta) - \sigma^2 g^2(X_i;\beta,\theta))^2 \right.$$
$$\left. + 2W_{i12}(Y_i - f(X;\beta))(Y_i^2 - f^2(X;\beta) - \sigma^2 g^2(X_i;\beta,\theta)) \right].$$

Since $W(X)$ is symmetric, we can write the spectral decomposition form of $Q_n(\gamma)$ as

$$Q_n(\gamma) = \sum_{i=1}^{n} \lambda_{1i}(\rho_i' e_{1i})^2 + \sum_{i=1}^{n} \lambda_{2i}(\rho_i' e_{2i})^2,$$

when λ_{1i} and λ_{2i} are nonnegative eigenvalues and e_{1i} and e_{2i} are eigenvectors of matrix $W(X_i)$.

Even if we use the linearization stage, the minimization stage cannot be applied because the first derivative of $Q_n(\gamma)$ does not have a linear form. So the minimization problem remains.

(iii). **Newton-Raphson Method**

Newton-Raphson is an algorithm for finding approximations to the roots of a real-valued function. The Taylor series of $\frac{\partial Q_n(\gamma)}{\partial \gamma}$ about the point $\hat{\gamma}_0 + \varepsilon$ is given by (keeping terms only to first order)

$$\frac{\partial Q_n(\hat{\gamma}_0 + \varepsilon)}{\partial \gamma} = \frac{\partial Q_n(\hat{\gamma}_0)}{\partial \gamma} + \frac{\partial^2 Q_n(\hat{\gamma}_0)}{\partial \gamma \partial \gamma'} \varepsilon.$$

Setting $\frac{\partial Q_n(\hat{\gamma}_0+\varepsilon)}{\partial\gamma}=0$, and solving the equation for $\varepsilon=\varepsilon_0$, we have

$$\varepsilon_0=-(\frac{\partial^2 Q_n(\hat{\gamma}_0)}{\partial\gamma\partial\gamma'})^{-1}\frac{\partial Q_n(\hat{\gamma}_0)}{\partial\gamma}.$$

With an initial choice of ε_0, the algorithm can be applied iteratively to obtain

$$\hat{\gamma}_{n+1}=\hat{\gamma}_n-(\frac{\partial^2 Q_n(\hat{\gamma}_n)}{\partial\gamma\partial\gamma'})^{-1}\frac{\partial Q_n(\hat{\gamma}_n)}{\partial\gamma}. \tag{4.3}$$

In application it is common to multiply a controlling factor (step size) by the second factor of (4.3) to control the steps toward the minimum value. There are other Newton's methods that use the second derivatives of $\frac{\partial Q_n(\gamma)}{\partial\gamma}$ which are more complicated in SLS case.

Comparing to Gradient descent method, it potentially speeds the convergence significantly. If the objective function $(\frac{\partial Q_n(\gamma)}{\partial\gamma})$ is continuously differentiable and its derivative does not vanish at the root and it has a second derivative at the root, then the convergence is quadratic or faster. If the derivative does vanish at the root, then the convergence is usually only linear.

The method can be used with an approximation to calculate the second derivative of $Q_n(\gamma)$. As we can see in (4.2), calculating the second part of the equation is troublesome. In application, we ignore the second part of the equation, so for minimization of the objective function, we need only the first derivative.

A problem with the Newton-Raphson method is that global optimization is not guaranteed and the procedure can be unstable near a horizontal asymptote or a local extremum. If the criterion function and parameter space are convex, then the criterion function has a unique local minimum, which is also the global minimum. For non convex problems however, there can be many local minima.

For such problems, a multi-start algorithm is used. In this algorithm, one starts a local optimization algorithm from initial values of the parameters to converge to a local minimum, and then one repeats the process a number of times with different initial values. The estimator is taken to be the parameter values that correspond to the small value of the criterion function obtained during the multi-start process.

There are some issues that are common between methods based on the gradient of the objective function. Here we mention some of them.

(a) The first common issue is the start point. The SLS estimator is defined as a global minimizer of $Q_n(\gamma)$, and its consistency and asymptotic normality depends on this assumption. A good initial value for minimization method is crucial because a poor starting value may obtain a local minimum instead of a global one. If the initial value is too far from the true parameter, the method converges very slowly or may even fail to converge. In SLS minimization problem with homoscedasticity, OLS estimator can be a good start. If we look at the quadratic form of $Q_n(\gamma)$, we see that OLS estimator minimizes the first two components, so it may be a good start. In the same way, for regression models with heteroscedasticity, weighted LS can be a proper start.

(b) Another issue is determining the step size. Too large of a step may miss the target and too small of a step may slow down the speed of convergence. It seems for regression models with homoscedasticity, fixed steps work. If we look at the form of $\rho_i(\gamma)$ and $\frac{\partial \rho'(\gamma_0)}{\partial \gamma}$, we see that in homoscedasticity case, $g(X; \beta, \theta) = 1$ and this makes the form of the matrices much simpler and their variations more controlled. However, for the models with heteroscedasticity, not only different step sizes for each iteration looks necessary, but also different step sizes for each component of $\hat{\gamma}$ can improve the convergence procedure. This improvement makes the minimization algorithm more complicated.

(c) Computation and evaluation of the gradients is another issue. If $Q_n(\gamma)$ has many extremum points, or it's too flat, the optimization procedure can be unstable. The second derivative matrix of the objective function can be exact or near singular.

Another group of optimization methods is Search Methods group. One of the simplest and most direct search methods is Random Grid Search method. This method usually handles the cases that the objective function is not differentiable, the derivatives are not continuous, or their evaluation is much more difficult than the evaluation of the objective function itself, or the second derivative matrix is exact or near singular. The method relies only on evaluating the objective function at a sequences of random points and comparing values in order to reach the optimum point. Comparing to the optimization methods based on the gradient of the objective function, this method is more stable, and because it does not need an initial value, it avoids the false convergence problem. The most important issue about this method is when the dimension of parameter space is high, the procedure can be very slow, or it may cause memory constraints problem.

Another search method that enhances the performance of a local method by using memory structure is Tabu Search. A more recent method that mimics the improvisation process of musicians is Harmonic Search. Although it seems that the method shows some advantages comparing with traditional mathematical optimization techniques, there are still no applications of this method in statistical literature. There are other optimization methods for convex functions, called Simplex Methods such as Nelder-Mead method or Downhill simplex method. There are also some methods based on smoothness of the objective function when existence of derivatives is not assumed. Generally, no matter what method of optimization is used to minimize a criterion function, it should be accompanied by evaluating the behavior of the function using graphical techniques (in small dimensional cases), exploratory steps, and analyzing the derivatives.

4.1.2 Optimum Weighting Matrix

One of the most important components in SLS estimation procedure is computation of the optimum weighting matrix. As we mentioned in the previous chapters, the optimum weighting matrix depends on unknown parameters, and it must be estimated. In nonlinear regression models with constant variance (homoscedastic), computation of the unknown parameters can be done in two steps using moments of the residuals. Li (2005), and Wang and Leblanc (2007) showed that this method works quite satisfactory. The problem arises when the variance of the response is not constant and may cause wild behavior of the weighting matrix. Using the Pearson (Studentized) residuals seem to solve the problem. We applied this method in examples 2.3.1 and 2.3.2 and the results were quite acceptable.

Another issue in computation of the weighting matrix is singularity. In chapter 2 we proved that the asymptotic covariance matrix of the most efficient SLS estimation is

$$\left(E \left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} F^{-1} \frac{\partial \rho(\gamma_0)}{\partial \gamma'} \right] \right)^{-1},$$

where $F = F(X) = E[\rho(\gamma_0)\rho'(\gamma_0)|X]$. This is based on the assumption that F is nonsingular and therefore it is invertible. However, sometimes in real application, the matrix might be singular. In this case, using the diagonal matrix with the same variance components as the optimum matrix, instead of the optimum matrix, can solve the problem. Although this matrix is not optimum anymore, but it is applicable.

A good initial value for SLS minimization method is crucial because a poor starting value may obtain a local minimum instead of a global one. In section 2.2 we stated that the optimum matrix can be computed using a two-stage procedure. In the first stage we minimize $Q_n(\gamma)$ using the identity matrix $W(X) = I_2$ to obtain the first-stage estimator $\hat{\gamma}_n$. Although this is an easy choice for the weighting matrix, however, in nonlinear regression models with homoscedasticity, calculating the optimum weighting matrix using the ordinary least squares estimation in the first stage, usually leads to a faster convergence. Similarly, for the nonlinear regression methods with heteroscedasticity, the generalized least squares estimation can be used in the first stage.

4.2 Simulation Accuracy

Mathematical theories have been able to show asymptotic behavior of estimators such as maximum (quasi) likelihood, (generalized) method of moment, and least squares. In this case, when there is no strong theory to verify the finite sample behavior of estimators, simulation study is a solution. Monte Carlo simulation offers an alternative to analytical mathematics for understanding a statistic's sampling distribution and evaluating its behavior in random samples. Monte Carlo simulation does this empirically using random samples from a known population of simulated data to track an estimator's behavior.

Consider model (2.1) where $\beta \in I\!R$ is the unknown regression parameter. Here we consider the case that β is a scalar, however it can be generalized to a vector case with more complicated notations. Suppose $(Y_i, X_i')', i = 1, 2, ..., n$ is an *i.i.d.* random sample. Our goal is to estimate the parameter β and measure the quality of the estimator.

A simulation study is usually undertaken to determine the value of β connected with a stochastic model like model (2.1). It estimates the parameter according to a criterion (such as minimizing sum of squared of residuals for least squares), and results in an output data $\hat{\beta}$, a random variable whose expected value is β. A second independent simulation - that is, a second simulation run - provides a new and independent random variable having mean β. This continues until we have amassed a total R runs - and R independent random variables $\hat{\beta}_1, ..., \hat{\beta}_R$ - all of which are identically distributed with mean β. The average of these R values, $\hat{\beta} = \sum_{i=1}^{R} \hat{\beta}_i / R$, is then used as an estimator of β. If $\hat{\beta}$ has a finite variance then the Law of Large Numbers implies that this estimate converges to β as R tends to infinity. Under the same conditions, the Central Limit Theorem allows us to refine this conclusion by asserting that the estimator $\hat{\beta} = \sum_{i=1}^{R} \hat{\beta}_i / R$ is approximately normally distributed about β with standard deviation $\sqrt{Var(\hat{\beta})/R}$ (Lange(1999)). Since a random variable is unlikely to be many standard deviations from its mean, it follows that $\hat{\beta}$ is a *good* estimator of β when $\sqrt{Var(\hat{\beta})/R}$ is small. Ross (1997) justifies the above statement from both the Chebyshev's inequality and more importantly for simulation studies, from the Central Limit Theorem. Indeed, for any $c > 0$, Chebyshev's inequality yields the rather conservative bound

$$P\left(|\hat{\beta} - \beta| > c\sqrt{Var(\hat{\beta})/R}\right) \leq \frac{1}{c^2}.$$

However, when R is large, as will usually be the case in simulations, we can apply the Central Limit Theorem to assert that $(\hat{\beta} - \beta)/\sqrt{Var(\hat{\beta})/R}$ is approximately distributed as a standard normal random variable; and thus

$$P\left(|\hat{\beta} - \beta| > c\sqrt{Var(\hat{\beta})/R}\right) \approx P\left(|Z| > c\right) = 2(1 - \Phi(c)), \tag{4.4}$$

where Φ is the standard normal distribution function. In practice, we estimate the Monte Carlo error as $\sqrt{v/R}$, where $v = \frac{1}{R-1}\sum_{i=1}^{R}[\hat{\beta}_i - \frac{1}{R}\sum_{i=1}^{R}\hat{\beta}_i]^2$ is the usual unbiased estimator of $Var(\hat{\beta})$. Finster (1987) defines the following concept of an *accurate* estimate.

Definition: $\hat{\beta}$ is an *accurate* estimate for β with accuracy A and confidence $1-\alpha$ (with $0 < \alpha < 1$), if

$$P(|\hat{\beta} - \beta| \leq A) \geq 1 - \alpha,$$

where $[-A, A]$ is the set of acceptable simulation errors. Similar to (4.4), applying the Central Limit Theorem for $\hat{\beta}$, we have

$$P\left(|\hat{\beta} - \beta| \leq z_{\alpha/2}\sqrt{Var(\hat{\beta})/R}\right) \approx 1 - \alpha, \tag{4.5}$$

when $z_{\alpha/2}$ is the $1-\alpha/2$ quantile on the right tail of the distribution $N(0,1)$. This provides a way of relating the number of replications with the accuracy

$$A = z_{\alpha/2}\sqrt{Var(\hat{\beta})/R}. \tag{4.6}$$

Solving for R, and replacing v as an unbiased estimator for $Var(\hat{\beta})$, we have

$$R = (\frac{z_{\alpha/2}}{A})^2 v.$$

Therefore, if we have a desirable value for accuracy A, we can calculate the (minimum) number of iterations. In literature, it is common to use $R = 1000$

iterations for simulation studies. However, Diaz-Emparanza (2000) argues that this size of sample for some models is small.

Once a variable is defined in terms of its distribution function, a variety of useful attributes of that variable are determined. One of the simplest ways to observe the asymptotic behavior of an estimator is to draw a histogram of the estimated values. It is useful to visualize the variable's distributional form of $\hat{\beta}$, and compare it with the bell-shaped normal curve with mean β and variance $Var(\hat{\beta})$. The skewness and kurtosis of a distribution function often are important in Monte Carlo work. Mooney (1997) presents

$$\frac{\sum_{i=1}^{R}(\hat{\beta}_i - \hat{\beta})^3/R}{[\sum_{i=1}^{R}(\hat{\beta}_i - \hat{\beta})^2/R]^{3/2}}$$

as a measure of skewness, and

$$\frac{\sum_{i=1}^{R}(\hat{\beta}_i - \hat{\beta})^4/R}{[\sum_{i=1}^{R}(\hat{\beta}_i - \hat{\beta})^2/R]^{2}}$$

as a measure of kurtosis for a simulated estimator. Comparing these values to the corresponding numbers for a normal distribution (zero for skewness and 3 for kurtosis), gives a better evaluation about the distribution of an estimator. Comparison of the distribution of an estimator to normal distribution can be done by a formal test. For example Jarque and Bera (1987) combine the skewness and kurtosis estimators to allow probability-based inferences about the normality of $\hat{\beta}$, based on a chi squared distribution. Moreover, Ross (1997) considers the Kolmogorov-Smirnov test to evaluate that whether an assumed probability distribution (in our case normal) is consistent with a given data set. Although these goodness-of-fit tests are quite common in simulation studies, however, Mooney (1997) argues that they have a low power issue.

There are other basic characteristics that have been used in literature as *measures of accuracy* in a Monte Carlo study. The central tendency of a distribution lets us estimate the estimator's bias by $\hat{\beta} - \beta$. Similarly, the root mean-squared error of an estimator can be estimated by $\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{\beta} - \beta)^2}$. The characteristics that we mentioned in here are more common, however, other measures of accuracy, such as median absolute error ($median|\hat{\beta} - \beta|$), and the bias of estimator's standard error in percentage ($\frac{100}{R}\sum_{i=1}^{R}[\hat{\sigma}_i(\hat{\beta}_i) - \sigma(\hat{\beta})]/\sigma(\hat{\beta})$) can be found as well (M*á*ty*á*s (1999)), where $\hat{\sigma}_i(\hat{\beta}_i)$ indicates the i th estimate of the standard errors of $\hat{\beta}$ based on the asymptotic standard errors and $\sigma(\hat{\beta})$ indicates the standard errors from the Monte Carlo replications.

Since in a simulation study we generate a data set from a given model, the methods that are usually used to assess whether a model fits to a data set, are not applicable. However, usual inferences about a parameter that are common in nonlinear regression books (Graybill and Iyer (1994), Seber and Wild (1989), Allen (1997)), can be done to evaluate the *quality* of an estimation procedure. For example hypothesis

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta \neq \beta_0 \end{cases}$$

can be tested using a z statistic $Z = \frac{\hat{\beta}-\beta_0}{\sqrt{v/R}}$. We can assess the quality of the test by calculating Type I error, Type II error, and the power. Using power (rejection rate when H_0 is not true) as a measure of the *quality* of the test, is very common in regression articles (See Wright (2003), and Schafgans (2004) as examples).

Constructing confidence intervals for the parameter is another common inference. We can construct a two-sided $1-\alpha$ confidence interval, by $\hat{\beta} \pm z_{\alpha/2}\sqrt{v/R}$. The rate of the coverage of the confidence interval and/or the average length of the interval have been used in articles to assess the quality of the inference about the parameter (See Fygenson and Ritov (1994) as an example).

Finally, graphical methods are useful to evaluate the asymptotic behavior of an estimator. As asymptotic theories demonstrates, we expect to see $\hat{\beta}$ converges to β, and confidence interval becomes smaller, as number of observation increases.

In the simulation studies in the past chapters, we evaluated the quality of the estimators, using different criteria, such as speed of convergence, bias, accuracy, root mean squared errors, and confidence intervals.

Chapter 5

Summary and Future Work

This thesis consists of four major contributions to the theory and method of nonlinear inference. The first contribution is to extend the SLS method to the Quasi-likelihood Variance Function (QVF) models. There we have established the strong consistency and asymptotic normality of the SLS estimators. We have derived the asymptotic covariance matrix of the most efficient SLS estimator. Since the optimum weighting matrix plays a very important role in the SLS estimator, we have found the explicit form of the matrix, presenting the asymptotic covariance matrix of the most efficient SLS estimator in homoscedastic linear regression models. We have shown some computational methods to deal with heteroscedasticity and we have applied the SLS method in some examples. The asymptotic confidence interval and testing hypothesis for the parameters of the QVF models based on SLS estimator, is a subject of future work.

The second contribution is theoretical and numerical comparisons between the SLS and GMM estimators. We have shown that SLS estimator is asymptotically more efficient than GMM estimator. The simulation studies show the superiority of the SLS comparing to GMM, even when it uses more equations. Similar to GMM, the higher order asymptotic properties of SLS can be investigated. More specifically, asymptotic bias of SLS is of interest.

The third contribution is to introduce the SLS in a censored model and we present the theories that show the estimator is strongly consistent and asymptotically normally distributed under some regularity conditions. We have also proposed a computationally simpler estimator which is consistent and asymptotically normal under the same regularity conditions. Finite sample behavior of the proposed estimators under both correctly and misspecified models have been investigated through Monte Carlo simulations. The simulation results show that the proposed estimator using optimal weighting matrix performs very similar to the maximum likelihood estimator, and the estimator with the identity weight is more robust against the misspecification. Improvement of the optimum SLS using a nonparametric weighting matrix is an extension of interest.

The fourth contribution is to extend the SLS method to generalized linear regression models with measurement error. We have estimated the parameters of a linear measurement error model using the SLS estimation in the second stage and we have shown that the parameter estimators are consistent and asymptotically normally distributed, when there is no assumption on the distribution of the error terms. Comparing to the more convenient two-stage method which uses OLS estimator in the second stage, we have proved that the two-stage SLS estimator is asymptotically more efficient. A simulation study shows an outstanding RMSE reduction in a moderate sample size. We have also presented the method of moments estimation for generalized linear measurement error models using the instrumental variable approach. We have assumed that the measurement error has a parametric distribution that is not necessarily normal, while the distributions of the unobserved covariates are nonparametric. We have also proposed simulation-based estimators for the situation where the closed forms of the moments are not available. We have shown that the proposed estimators are strongly consistent and asymptotically normally distributed under some regularity conditions. Finite sample performances of the estimators have been investigated through simulation studies. The extension of the method to the case that the measurement error has a nonparametric distribution of the error term will also be studied.

As one of the minimization methods, SLS has its own computational issues. We have discussed about various computational methods with their advantages and disadvantages. We have encountered some of the issues in nonlinear models and we have improved the method of computation. We are still looking for an efficient and fast computational method to estimate parameters. It is particularly important because those practical models tend to have a large number of parameters.

Appendix A

Proofs

Throughout the proofs, we use the following notations. For any matrix M, its Euclidean norm is denoted as $\|M\| = \sqrt{\text{trace}(M'M)}$, and vec$M$ denotes the column vector consisting of the stacked up columns of M. Further, $\otimes$ denotes the Kronecker product operator.

A.1 Proof of Theorem 2.1.3

We show that Assumptions 1 - 4 are sufficient for all conditions of Lemma 2.1.1. By Cauchy-Schwarz inequality and Assumptions 1 – 3 we have

$$E\left[\|W_1\| \sup_{\Omega} (Y_1 - f(X_1;\beta))^2\right] \leq 2E\,\|W_1\|\,Y_1^2 + 2E\,\|W_1\| \sup_{\Omega} f^2(X_1;\beta) < \infty$$

and

$$E\left[\|W_1\|\sup_{\Gamma}\left(Y_1^2 - f^2(X_1;\beta) - \sigma^2 g^2(X_1;\beta,\theta)\right)^2\right]$$
$$\leq 2E\|W_1\| Y_1^4 + 2E\|W_1\|\sup_{\Omega} f^4(X_1;\beta)$$
$$+2E\|W_1\|\sup_{\Sigma}\sigma^4\sup_{\Omega\times\Theta} g^4(X_1;\beta,\theta) < \infty,$$

which imply

$$E\sup_{\Gamma}\rho_1'(\gamma)W_1\rho_1(\gamma) \leq E\|W_1\|\sup_{\Gamma}\|\rho_1(\gamma)\|^2 < \infty. \quad \text{(A.1)}$$

It follows from the uniform law of large numbers (ULLN) (Jennrich 1969, Theorem 2) that $\frac{1}{n}Q_n(\gamma)$ converges almost surely ($a.s.$) to $Q(\gamma) = E\rho_1'(\gamma)W(X_1)\rho_1(\gamma)$ uniformly for all $\gamma \in \Gamma$. Since $\rho_1(\gamma) - \rho_1(\gamma_0)$ does not depend on Y_1, we have

$$E[\rho_1'(\gamma_0)W_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = E[E(\rho_1'(\gamma_0)|X_1)W_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = 0,$$

which implies $Q(\gamma) = Q(\gamma_0) + E[(\rho_1(\gamma) - \rho_1(\gamma_0))'W_1(\rho_1(\gamma) - \rho_1(\gamma_0))]$. It follows that $Q(\gamma) \geq Q(\gamma_0)$ and, by Assumption 4, equality holds if and only if $\gamma = \gamma_0$. Thus all conditions of Lemma 2.1.1 hold and, therefore, $\hat{\gamma}_{SLS} \xrightarrow{a.s.} \gamma_0$ follows.

A.2 Proof of Theorem 2.1.4

By Assumption 5 the first derivative $\partial Q_n(\gamma)/\partial\gamma$ exists and has a first-order Taylor expansion in Γ. Since $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, for sufficiently large n it holds with probability one

$$\frac{\partial Q_n(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'}(\hat{\gamma}_{SLS} - \gamma_0) = \frac{\partial Q_n(\hat{\gamma}_{SLS})}{\partial\gamma} = 0, \tag{A.2}$$

where $\|\tilde{\gamma}_n - \gamma_0\| \leq \|\hat{\gamma}_{SLS} - \gamma_0\|$. The first derivative of $Q_n(\gamma)$ in $(A.2)$ is given by

$$\frac{\partial Q_n(\gamma)}{\partial\gamma} = 2\sum_{i=1}^{n} \frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \rho_i(\gamma),$$

where

$$\frac{\partial\rho'(\gamma)}{\partial\gamma} = -\begin{pmatrix} \frac{\partial f(X;\beta)}{\partial\beta} & 2\sigma^2 g(X;\beta,\theta)\frac{\partial g(X;\beta,\theta)}{\partial\beta} + 2f(X;\beta)\frac{\partial f(X;\beta)}{\partial\beta} \\ 0 & 2\sigma^2 g(X;\beta,\theta)\frac{\partial g(X;\beta,\theta)}{\partial\theta} \\ 0 & g^2(X;\beta,\theta) \end{pmatrix}$$

The second derivative of $Q_n(\gamma)$ in $(A.2)$ is given by

$$\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'} = 2\sum_{i=1}^{n}\left[\frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \frac{\partial\rho_i(\gamma)}{\partial\gamma'} + (\rho_i'(\gamma) W_i \otimes I_{p+q+1})\frac{\partial \text{vec}(\partial\rho_i'(\gamma)/\partial\gamma)}{\partial\gamma'}\right],$$

where

$$\frac{\partial \text{vec}(\partial \rho_i'(\gamma)/\partial \gamma)}{\partial \gamma'} =$$

$$-\begin{pmatrix} a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ b & c & 2g(X_i;\beta,\theta)\frac{\partial g(X_i;\beta,\theta)}{\partial \beta} \\ d & e & 2g(X_i;\beta,\theta)\frac{\partial g(X_i;\beta,\theta)}{\partial \theta} \\ 2g(X_i;\beta,\theta)\frac{\partial g(X_i;\beta,\theta)}{\partial \beta'} & 2g(X_i;\beta,\theta)\frac{\partial g(X_i;\beta,\theta)}{\partial \theta'} & 0 \end{pmatrix}.$$

Here

$$a = \frac{\partial^2 f(X_i;\beta)}{\partial \beta \partial \beta'},$$

$$\begin{aligned} b &= 2\sigma^2 g(X_i;\beta,\theta)\frac{\partial^2 g(X_i;\beta,\theta)}{\partial \beta \partial \beta'} + 2\sigma^2 \frac{\partial g(X_i;\beta,\theta)}{\partial \beta}\frac{\partial g(X_i;\beta,\theta)}{\partial \beta'} \\ &+ 2f(X_i;\beta)\frac{\partial^2 f(X_i;\beta)}{\partial \beta \partial \beta'} + 2\frac{\partial f(X_i;\beta)}{\partial \beta}\frac{\partial f(X_i;\beta)}{\partial \beta'}, \end{aligned}$$

$$c = 2\sigma^2 g(X_i;\beta,\theta)\frac{\partial^2 g(X_i;\beta,\theta)}{\partial \beta \partial \theta'} + 2\sigma^2 \frac{\partial g(X_i;\beta,\theta)}{\partial \beta}\frac{\partial g(X_i;\beta,\theta)}{\partial \theta'},$$

$$d = 2\sigma^2 g(X_i;\beta,\theta)\frac{\partial^2 g(X_i;\beta,\theta)}{\partial \theta \partial \beta'} + 2\sigma^2 \frac{\partial g(X_i;\beta,\theta)}{\partial \theta}\frac{\partial g(X_i;\beta,\beta)}{\partial \beta'},$$

and

$$e = 2\sigma^2 g(X_i;\beta,\theta)\frac{\partial^2 g(X_i;\beta,\theta)}{\partial \theta \partial \theta'} + 2\sigma^2 \frac{\partial g(X_i;\beta,\theta)}{\partial \theta}\frac{\partial g(X_i;\beta,\beta)}{\partial \theta'}.$$

By Assumption 5 and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
& E \sup_{\Gamma} \left\| \frac{\partial \rho_1'(\gamma)}{\partial \gamma} W_1 \frac{\partial \rho_1(\gamma)}{\partial \gamma'} \right\| \leq E \left\| W_1 \right\| \sup_{\Gamma} \left\| \frac{\partial \rho_1'(\gamma)}{\partial \gamma} \right\|^2 \\
\leq \; & E \left\| W_1 \right\| \sup_{\Gamma} \left(\left\| \frac{\partial f(X_1;\beta)}{\partial \beta} \right\|^2 + 4\sigma^4 \left\| g(X_1;\beta,\theta) \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^2 \right. \\
+ \; & \left. 4\sigma^4 \left\| g(X_1;\beta,\theta) \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \right\|^2 + 4 \left\| f(X_1;\beta) \frac{\partial f(X_1;\beta)}{\partial \beta} \right\|^2 + \left\| g^2(X_1;\beta,\theta) \right\|^2 \right) \\
\leq \; & E \left\| W_1 \right\| \sup_{\Omega} \left\| \frac{\partial f(X_1;\beta)}{\partial \beta} \right\|^2 + E \left\| W_1 \right\| \sup_{\Omega \times \Theta} \left\| g^2(X_1;\beta,\theta) \right\|^2 \\
+ \; & 4 \sup_{\Sigma} \sigma^4 \left[E \left(\left\| W_1 \right\| \sup_{\Omega \times \Theta} \left\| g(X_1;\beta,\theta) \right\|^4 \right) E \left(\left\| W_1 \right\| \sup_{\Omega \times \Theta} \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^4 \right) \right]^{1/2} \\
+ \; & 4 \sup_{\Sigma} \sigma^4 \left[E \left(\left\| W_1 \right\| \sup_{\Omega \times \Theta} \left\| g(X_1;\beta,\theta) \right\|^4 \right) E \left(\left\| W_1 \right\| \sup_{\Omega \times \Theta} \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \right\|^4 \right) \right]^{1/2} \\
+ \; & 4 \left[E \left(\left\| W_1 \right\| \sup_{\Omega} \left\| f(X_1;\beta) \right\|^4 \right) E \left(\left\| W_1 \right\| \sup_{\Omega} \left\| \frac{\partial f(X_1;\beta)}{\partial \beta} \right\|^4 \right) \right]^{1/2} \\
< \; & \infty. \qquad \text{(A.3)}
\end{aligned}
$$

Similarly, because of (A.1) and

$$
\begin{aligned}
& E\left(\|W_1\| \sup_{\Gamma} \left\| \frac{\partial \text{vec}(\partial \rho_1'(\gamma)/\partial \gamma)}{\partial \gamma'} \right\|^2 \right) \\
\leq \; & E\,\|W_1\| \sup_{\Gamma} \left(\left\| \frac{\partial^2 f(X_1;\beta)}{\partial \beta \partial \beta'} \right\|^2 + 4\sigma^4 \left\| g(X_1;\beta,\theta) \frac{\partial^2 g(X_1;\beta,\theta)}{\partial \beta \partial \beta'} \right\|^2 \right. \\
+ \; & 4\sigma^4 \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^4 + 4 \left\| f(X_1;\beta) \frac{\partial^2 f(X_1;\beta)}{\partial \beta \partial \beta'} \right\|^2 + 4 \left\| \frac{\partial f(X_1;\beta)}{\partial \beta} \right\|^4 \\
+ \; & 8\sigma^4 \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^2 + 8\sigma^4 \left\| g(X_1;\beta,\theta) \frac{\partial^2 g(X_1;\beta,\theta)}{\partial \theta \partial \beta'} \right\|^2 \\
+ \; & 4\sigma^4 \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \right\|^4 + 4\sigma^4 \left\| g(X_1;\beta,\theta) \frac{\partial^2 g(X_1;\beta,\theta)}{\partial \theta \partial \theta'} \right\|^2 \\
+ \; & 8 \left. \left\| g(X_1;\beta,\theta) \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^2 + 8 \left\| g(X_1;\beta,\theta) \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \right\|^2 \right) \\
\leq \; & E\left(\|W_1\| \sup_{\Omega} \left\| \frac{\partial^2 f(X_1;\beta)}{\partial \beta \partial \beta'} \right\|^2 \right) + 4 \sup_{\Sigma} \sigma^4 E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^4 \right) \\
+ \; & 4 \sup_{\Sigma} \sigma^4 \left[E\left(\|W_1\| \sup_{\Omega \times \Theta} \|g(X_1;\beta,\theta)\|^4 \right) E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial^2 g(X_1;\beta,\theta)}{\partial \beta \partial \beta'} \right\|^4 \right) \right]^{1/2} \\
+ \; & 4 \left[E\left(\|W_1\| \sup_{\Omega} \|f(X_1;\beta)\|^4 \right) E\left(\|W_1\| \sup_{\Omega} \left\| \frac{\partial^2 f(X_1;\beta)}{\partial \beta \partial \beta'} \right\|^4 \right) \right]^{1/2} \\
+ \; & 4E\left(\|W_1\| \sup_{\Omega} \left\| \frac{\partial f(X_1;\beta)}{\partial \beta} \right\|^4 \right) + 4 \sup_{\Sigma} \sigma^4 E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \right\|^4 \right) \\
+ \; & 8 \sup_{\Sigma} \sigma^4 \left[E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \beta} \right\|^4 \right) E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial g(X_1;\beta,\theta)}{\partial \theta} \right\|^4 \right) \right]^{1/2} \\
+ \; & 8 \sup_{\Sigma} \sigma^4 \left[E\left(\|W_1\| \sup_{\Omega \times \Theta} \|g(X_1;\beta,\theta)\|^4 \right) E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial^2 g(X_1;\beta,\theta)}{\partial \theta \partial \beta'} \right\|^4 \right) \right]^{1/2} \\
+ \; & 4 \sup_{\Sigma} \sigma^4 \left[E\left(\|W_1\| \sup_{\Omega \times \Theta} \|g(X_1;\beta,\theta)\|^4 \right) E\left(\|W_1\| \sup_{\Omega \times \Theta} \left\| \frac{\partial^2 g(X_1;\beta,\theta)}{\partial \theta \partial \theta'} \right\|^4 \right) \right]^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad 8\left[E\left(\|W_1\|\sup_{\Omega\times\Theta}\|g(X_1;\beta,\theta)\|^4\right)E\left(\|W_1\|\sup_{\Omega\times\Theta}\left\|\frac{\partial g(X_1;\beta,\theta)}{\partial\beta}\right\|^4\right)\right]^{1/2} \\
&+ \quad 8\left[E\left(\|W_1\|\sup_{\Omega\times\Theta}\|g(X_1;\beta,\theta)\|^4\right)E\left(\|W_1\|\sup_{\Omega\times\Theta}\left\|\frac{\partial g(X_1;\beta,\theta)}{\partial\theta}\right\|^4\right)\right]^{1/2} \\
&< \quad \infty,
\end{aligned}
$$

we have

$$
\begin{aligned}
& E\sup_{\Gamma}\left\|(\rho_1'(\gamma)W_1\otimes I_{p+q+1})\,\frac{\partial\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\| \\
\leq \quad & (p+q+1)E\,\|W_1\|\sup_{\Gamma}\|\rho_1(\gamma)\|\left\|\frac{\partial\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\| \\
\leq \quad & (p+q+1)\left[E\left(\|W_1\|\sup_{\Gamma}\|\rho_1(\gamma)\|^2\right)E\left(\|W_1\|\sup_{\Gamma}\left\|\frac{\partial\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|^2\right)\right]^{1/2} \\
< \quad & \infty. \qquad \text{(A.4)}
\end{aligned}
$$

It follows from (A.3), (A.4) and the ULLN that

$$
\begin{aligned}
& \frac{1}{n}\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'} \xrightarrow{a.s.} \frac{\partial^2 Q(\gamma)}{\partial\gamma\partial\gamma'} \\
= \quad & 2E\left[\frac{\partial\rho_1'(\gamma)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma)}{\partial\gamma'} + (\rho_1'(\gamma)W_1\otimes I_{p+q+1})\frac{\partial\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right]
\end{aligned}
$$

uniformly for all $\gamma\in\Gamma$. Therefore by Lemma 2.1.2, we have

$$
\frac{1}{n}\frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'} \xrightarrow{a.s.} \frac{\partial^2 Q(\gamma_0)}{\partial\gamma\partial\gamma'} = 2B, \qquad \text{(A.5)}
$$

where the second equality holds, because

$$
\begin{aligned}
& E\left[(\rho_1'(\gamma_0)W_1\otimes I_{p+q+1})\frac{\partial\mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \\
= \quad & E\left[(E(\rho_1'(\gamma_0)|X_1)W_1\otimes I_{p+q+1})\frac{\partial\mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] = 0.
\end{aligned}
$$

Furthermore, since $\frac{\partial \rho_i'(\gamma)}{\partial \gamma} W_i \rho_i(\gamma)$ are *i.i.d.* with zero mean, the Central Limit Theorem (CLT) implies that

$$\frac{1}{\sqrt{n}} \frac{\partial Q_n(\gamma_0)}{\partial \gamma} \xrightarrow{L} N(\mathbf{0}, 4A), \tag{A.6}$$

where A is given in (2.3). It follows from $(A.2)$, $(A.5)$, $(A.6)$ and Assumption 6, that $\sqrt{n}(\hat{\gamma}_{SLS} - \gamma_0)$ converges in distribution to $N(\mathbf{0}, B^{-1}AB^{-1})$.

A.3 Proof of Theorem 2.5.4

First, assumption A7 and the Dominated Convergence Theorem imply that $m_1(X;\gamma)$, $m_2(X;\gamma)$ and therefore $Q_n(\gamma)$ are continuous in $\gamma \in \Gamma$. Let $Q(\gamma) = E\rho_1'(\gamma)W_1\rho_1(\gamma)$. Since by Hölder's inequality and assumptions A8 - A10

$$\begin{aligned} & E \|W_1\| \sup_\Gamma [Y_1 - m_1(X_1;\gamma)]^2 \\ \leq \; & 2E \|W_1\| Y_1^2 + 2E \|W_1\| \sup_\Gamma m_1^2(X_1;\gamma) \\ \leq \; & 2E \|W_1\| Y_1^2 + 2E \|W_1\| \int \sup_{\Omega \times \Phi} (X_1'\beta + \varepsilon)^2 f(\varepsilon;\phi) d\varepsilon \\ < \; & \infty \end{aligned}$$

and

$$
\begin{aligned}
& E\left\|W_1\right\| \sup_{\Gamma}[Y_1^2 - m_2(X_1;\gamma)]^2 \\
\leq\ & 2E\left\|W_1\right\| Y_1^4 + 2E\left\|W_1\right\| \sup_{\Gamma} m_2^2(X_1;\gamma) \\
\leq\ & 2E\left\|W_1\right\| Y_1^4 + 2E\left\|W_1\right\| \int \sup_{\Omega\times\Phi}(X_1'\beta+\varepsilon)^4 f(\varepsilon;\phi)d\varepsilon \\
<\ & \infty,
\end{aligned}
$$

we have

$$
\begin{aligned}
& E\sup_{\Gamma}\left\|\rho_1'(\gamma)W_1\rho_1(\gamma)\right\| \leq E\left\|W_1\right\| \sup_{\Gamma}\left\|\rho_1(\gamma)\right\|^2 \\
\leq\ & E\left\|W_1\right\| \sup_{\Gamma}[Y_1 - m_1(X_1;\gamma)]^2 + E\left\|W_1\right\| \sup_{\Gamma}[Y_1^2 - m_2(X_1;\gamma)]^2 \\
<\ & \infty. \qquad \text{(A.7)}
\end{aligned}
$$

It follows from Jenrich (1969, Theorem 2) that $\frac{1}{n}Q_n(\gamma)$ converges almost surely to $Q(\gamma)$ uniformly in $\gamma \in \Gamma$. Further, since

$$
E[\rho_1'(\gamma_0)W_1(\rho_1(\gamma)-\rho_1(\gamma_0))] = E[(E(\rho_1'(\gamma_0)|X_1)W_1(\rho_1(\gamma)-\rho_1(\gamma_0))] = 0,
$$

we have

$$
Q(\gamma) = Q(\gamma_0) + E[(\rho_1(\gamma)-\rho_1(\gamma_0))'W_1(\rho_1(\gamma)-\rho_1(\gamma_0))].
$$

It follows that $Q(\gamma) \geq Q(\gamma_0)$ and, by assumption A11, equality holds if and only if $\gamma = \gamma_0$. Thus all conditions of Lemma 2.1.1 hold and, therefore, $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$ follows.

A.4 Proof of Theorem 2.5.5

By assumption A12 the first derivative $\partial Q_n(\gamma)/\partial\gamma$ exists and has a first order Taylor expansion in a neighborhood $\Gamma_0 \subset \Gamma$ of γ_0. Since $\partial Q_n(\hat{\gamma}_n)/\partial\gamma = 0$ and $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, for sufficiently large n we have

$$0 = \frac{\partial Q_n(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'}(\hat{\gamma}_n - \gamma_0), \tag{A.8}$$

where $\|\tilde{\gamma}_n - \gamma_0\| \leq \|\hat{\gamma}_n - \gamma_0\|$. The first derivative of $Q_n(\gamma)$ in $(A.8)$ is given by

$$\frac{\partial Q_n(\gamma)}{\partial\gamma} = 2\sum_{i=1}^{n} \frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \rho_i(\gamma),$$

where

$$\frac{\partial\rho_i'(\gamma)}{\partial\gamma} = -\left(\frac{\partial m_1(X_i;\gamma)}{\partial\gamma}, \frac{\partial m_2(X_i;\gamma)}{\partial\gamma}\right)$$

and the first derivatives of $m_1(X_i;\gamma)$ and $m_2(X_i;\gamma)$ with respect to γ are given after assumption A13. Therefore, by the Central Limit Theorem we have

$$\frac{1}{\sqrt{n}}\frac{\partial Q_n(\gamma_0)}{\partial\gamma} \xrightarrow{L} N(0, 4A), \tag{A.9}$$

where

$$A = E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma} W_1 \rho_1(\gamma_0)\rho_1'(\gamma_0) W_1 \frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}\right].$$

The second derivative of $Q_n(\gamma)$ in $(A.8)$ is given by

$$\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'} = 2\sum_{i=1}^{n}\left[\frac{\partial\rho_i'(\gamma)}{\partial\gamma}W_i\frac{\partial\rho_i(\gamma)}{\partial\gamma'} + (\rho_i'(\gamma)W_i \otimes I_{p+q})\frac{\partial \text{vec}(\partial\rho_i'(\gamma)/\partial\gamma)}{\partial\gamma'}\right],$$

where

$$\frac{\partial \text{vec}(\partial\rho_i'(\gamma)/\partial\gamma)}{\partial\gamma'} = -\left(\frac{\partial^2 m_1(X;\gamma)}{\partial\gamma\partial\gamma'}, \frac{\partial^2 m_2(X;\gamma)}{\partial\gamma\partial\gamma'}\right)'.$$

The elements in $\partial^2 m_1(x;\gamma)/\partial\gamma\partial\gamma'$ are

$$\frac{\partial^2 m_1(x;\gamma)}{\partial\beta\partial\beta'} = xx'f(-x'\beta;\phi), \frac{\partial^2 m_1(x;\gamma)}{\partial\beta\partial\phi'} = x\int_{-x'\beta}^{\infty}\frac{\partial f(\varepsilon;\phi)}{\partial\phi'}d\varepsilon,$$

$$\frac{\partial^2 m_1(x;\gamma)}{\partial\phi\partial\phi'} = \int_{-x'\beta}^{\infty}(x'\beta+\varepsilon)\frac{\partial^2 f(\varepsilon;\phi)}{\partial\phi\partial\phi'}d\varepsilon,$$

and the elements in $\partial^2 m_2(x;\gamma)/\partial\gamma\partial\gamma'$ are

$$\frac{\partial^2 m_2(x;\gamma)}{\partial\beta\partial\beta'} = 2xx'\int_{-x'\beta}^{\infty}f(\varepsilon;\phi)d\varepsilon,$$

$$\frac{\partial^2 m_2(x;\gamma)}{\partial\beta\partial\phi'} = 2x\int_{-x'\beta}^{\infty}(x'\beta+\varepsilon)\frac{\partial f(\varepsilon;\phi)}{\partial\phi'}d\varepsilon,$$

$$\frac{\partial^2 m_2(x;\gamma)}{\partial\phi\partial\phi'} = \int_{-x'\beta}^{\infty}(x'\beta+\varepsilon)^2\frac{\partial^2 f(\varepsilon;\phi)}{\partial\phi\partial\phi'}d\varepsilon.$$

Analogous to the proof of Theorem 2.5.4, in the following we verify by assumption A12 and Jenrich (1969, Theorem 2) that $(1/n)\partial^2 Q_n(\gamma)/\partial\gamma\partial\gamma'$ converges almost surely to $\partial^2 Q(\gamma)/\partial\gamma\partial\gamma'$ uniformly in $\gamma \in \Gamma_0$. First, by assumption A12, we have

$$
\begin{aligned}
& E \sup_{\Gamma} \left\| \frac{\partial \rho_1'(\gamma)}{\partial \gamma} W_1 \frac{\partial \rho_1(\gamma)}{\partial \gamma'} \right\| \leq E \left\| W_1 \right\| \sup_{\Gamma} \left\| \frac{\partial \rho_1'(\gamma)}{\partial \gamma} \right\|^2 \\
\leq \; & E \left\| W_1 \right\| \sup_{\Gamma} \left(\left\| \frac{\partial m_1(X_1; \gamma)}{\partial \beta} \right\|^2 + \left\| \frac{\partial m_1(X_1; \gamma)}{\partial \phi} \right\|^2 \right. \\
+ \; & \left. \left\| \frac{\partial m_2(X_1; \gamma)}{\partial \beta} \right\|^2 + \left\| \frac{\partial m_2(X_1; \gamma)}{\partial \phi} \right\|^2 \right) \\
\leq \; & E \left\| W_1 \right\| X_1' X_1 \sup_{\Gamma} (\int_{-X_1'\beta}^{\infty} f(\varepsilon; \phi) d\varepsilon)^2 \\
+ \; & E \left\| W_1 \right\| \sup_{\Gamma} \left\| \int_{-X_1'\beta}^{\infty} (X_1'\beta + \varepsilon) \frac{\partial f(\varepsilon; \phi)}{\partial \phi} \right\|^2 d\varepsilon \\
+ \; & 8E \left\| W_1 \right\| X_1' X_1 \sup_{\Gamma} (X_1'\beta)^2 (\int_{-X_1'\beta}^{\infty} f(\varepsilon; \phi) d\varepsilon)^2 \\
+ \; & 8E \left\| W_1 \right\| X_1' X_1 \sup_{\Gamma} (\int_{-X_1'\beta}^{\infty} \varepsilon f(\varepsilon; \phi) d\varepsilon)^2 \\
+ \; & E \left\| W_1 \right\| \sup_{\Gamma} \left\| \int_{-X_1'\beta}^{\infty} (X_1'\beta + \varepsilon)^2 \frac{\partial f(\varepsilon; \phi)}{\partial \phi} \right\|^2 d\varepsilon < \infty
\end{aligned}
$$

and

$$
\begin{aligned}
& E \sup_{\Gamma} \left\| (\rho_1'(\gamma) W_1 \otimes I_{p+q}) \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma)/\partial \gamma)}{\partial \gamma'} \right\| \\
\leq \; & \sqrt{2(p+q)} E \left\| W_1 \right\| \sup_{\Gamma} \left\| \rho_1(\gamma) \right\| \left\| \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma)/\partial \gamma)}{\partial \gamma'} \right\| \\
\leq \; & \sqrt{2(p+q)} \left(E \left\| W_1 \right\| \sup_{\Gamma} \left\| \rho_1(\gamma) \right\|^2 E \left\| W_1 \right\| \sup_{\Gamma} \left\| \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma)/\partial \gamma)}{\partial \gamma'} \right\|^2 \right)^{1/2} \\
< \; & \infty,
\end{aligned}
$$

where the last inequality holds, because

$$E\left(\|W_1\| \sup_{\Gamma} \left\| \frac{\partial \text{vec}(\partial \rho_1'(\gamma)/\partial \gamma)}{\partial \gamma'} \right\| \right)^2 \quad < \quad \infty$$

and, by (A.7), $E\|W_1\| \sup_{\Gamma} \|\rho_1(\gamma)\|^2 < \infty$. Therefore by Lemma 2.1.2 we have

$$\begin{aligned} & \frac{1}{n}\frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'} \\ \xrightarrow{a.s.} \quad & 2E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'} + (\rho_1'(\gamma_0)W_1 \otimes I_{p+q})\frac{\partial \text{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \\ = \quad & 2B, \end{aligned} \tag{A.10}$$

where the second equality holds, because

$$
\begin{aligned}
& E\left(\|W_1\|\sup_\Gamma\left\|\frac{\partial \mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|\right)^2 \\
\leq\ & E\,\|W_1\|\sup_\Gamma\left(\left\|\frac{\partial^2 m_1(X_1;\gamma)}{\partial\beta\partial\beta'}\right\|^2+\left\|\frac{\partial^2 m_1(X_1;\gamma)}{\partial\phi\partial\phi'}\right\|^2\right. \\
+\ & 2\left\|\frac{\partial^2 m_1(X_1;\gamma)}{\partial\beta\partial\phi'}\right\|^2+\left\|\frac{\partial^2 m_2(X_1;\gamma)}{\partial\beta\partial\beta'}\right\|^2 \\
+\ & \left.2\left\|\frac{\partial^2 m_2(X_1;\gamma)}{\partial\beta\partial\phi'}\right\|^2+\left\|\frac{\partial^2 m_2(X_1;\gamma)}{\partial\phi\partial\phi'}\right\|^2\right) \\
\leq\ & E\,\|W_1\|\,\|X_1X_1'\|^2\sup_\Gamma f^2(-X_1'\beta;\phi) \\
+\ & 2E\,\|W_1\|\sup_\Gamma\left\|X_1\int_{-X_1'\beta}^\infty\frac{\partial f(\varepsilon;\phi)}{\partial\phi}d\varepsilon\right\|^2 \\
+\ & E\,\|W_1\|\sup_\Gamma\left\|\int_{-X_1'\beta}^\infty(X_1'\beta+\varepsilon)\frac{\partial^2 f(\varepsilon;\phi)}{\partial\phi\partial\phi'}d\varepsilon\right\|^2 \\
+\ & 4E\,\|W_1\|\,\|X_1X_1'\|\sup_\Gamma\left(\int_{-X_1'\beta}^\infty f(\varepsilon;\phi)d\varepsilon\right)^2 \\
+\ & 16E\,\|W_1\|\sup_\Gamma\left\|X_1\left(X_1'\beta\int_{-X_1'\beta}^\infty\frac{\partial f(\varepsilon;\phi)}{\partial\phi}d\varepsilon\right)\right\|^2 \\
+\ & 16E\,\|W_1\|\sup_\Gamma\left\|X_1\int_{-X_1'\beta}^\infty\varepsilon\frac{\partial f(\varepsilon;\phi)}{\partial\phi}d\varepsilon\right\|^2 \\
+\ & E\,\|W_1\|\sup_\Gamma\left\|\int_{-X_1'\beta}^\infty(X_1'\beta+\varepsilon)^2\frac{\partial^2 f(\varepsilon;\phi)}{\partial\phi\partial\phi'}d\varepsilon\right\|^2<\infty
\end{aligned}
$$

It follows then from $(A.8)-(A.10)$, Assumption A13 and the Slutsky Theorem, that $\sqrt{n}(\hat\gamma_n-\gamma_0)\xrightarrow{L}N(0,B^{-1}AB^{-1})$.

A.5 Proof of Theorem 3.2.4

To prove the consistency, we use the Uniform Law of Large Numbers (ULLN, Jennrich 1969, Theorem 2). The idea is to show that $E\sup_\gamma |Q_n(\gamma)| < \infty$. We prove the theorem for two cases.

(i). Assume that H is known, then by ASSUMPTIONS 15 and 16, and the Dominated Convergence Theory,

$$\begin{aligned}
& E\sup_\gamma |\rho_1'(\gamma)W_1\rho_1(\gamma)| \\
\leq\ & E\,\|W_1\| \sup_\gamma \|\rho_1(\gamma)\|^2 \\
\leq\ & 2E\,\|W_1\| \left(\|Y_1 Z_1\|^2 + Y_1^4\right) \\
+\ & 2E\,\|W_1\| \left(\int \sup_\gamma f(x'\beta) f_U(x - HV_1;\phi)(\|x\|+1)dx\right)^2 \\
+\ & 2E\,\|W_1\| \left(\int \sup_\gamma f^2(x'\beta) f_U(x - HV_1;\phi)dx\right)^2 \\
+\ & 2E\,\|W_1\| \left(\int \sup_\gamma g^2(x'\beta;\theta) f_U(x - HV_1;\phi)dx\right)^2 \\
<\ & \infty. \qquad \text{(A.11)}
\end{aligned}$$

It follows from ULLN that

$$\sup_\gamma \left| \frac{1}{n}\sum_{i=1}^n \rho_i'(\gamma) W_i \rho_i(\gamma) - Q(\gamma) \right| \xrightarrow{a.s.} 0,$$

where $Q(\gamma) = E\rho_1'(\gamma)W_1\rho_1(\gamma)$. Therefore,

$$\sup_{\gamma}\left|\frac{1}{n}Q_n(\gamma) - Q(\gamma)\right| \xrightarrow{a.s.} 0.$$

Now for the next step, we use Lemma 2.1.1 to show that $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$. Since $E(\rho_1(\gamma_0)|V_1) = 0$ and $\rho_1(\gamma) - \rho_1(\gamma_0)$ depends on V_1 only, we have

$$E[\rho_1'(\gamma_0)W_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = E[E(\rho_1'(\gamma_0)|V_1)W_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = 0,$$

which implies $Q(\gamma) = Q(\gamma_0) + E[(\rho_1(\gamma) - \rho_1(\gamma_0))'W_1(\rho_1(\gamma) - \rho_1(\gamma_0))]$. By Assumption 17, $Q(\gamma) \geq Q(\gamma_0)$ and equality holds if and only if $\gamma = \gamma_0$. Thus, $Q(\gamma)$ attains a unique minimum at $\gamma_0 \in \Gamma$, and it follows from the Lemma that $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$.

(ii). Assume that H is unknown. We know that $\hat{\rho}_i(\gamma)$ and hence $Q_n(\gamma)$ are continuously differentiable with respect to ψ. Therefore, for sufficiently large n, $Q_n(\gamma)$ has the first-order Taylor expansion about ψ_0:

$$Q_n(\gamma) = \sum_{i=1}^{n}\rho_i'(\gamma)W_i\rho_i(\gamma) + 2\sum_{i=1}^{n}\rho_i'(\gamma,\tilde{\psi})W_i\frac{\partial\rho_i(\gamma,\tilde{\psi})}{\partial\psi'}\left(\hat{\psi} - \psi_0\right), \tag{A.12}$$

where $\rho_i(\gamma) = Y_iT_i - m(HV_i;\gamma)$, $\rho_i(\gamma,\tilde{\psi}) = Y_iT_i - m(\tilde{H}V_i;\gamma)$, and $\tilde{\psi} = \text{vec}\tilde{H}$ satisfies $\left\|\tilde{\psi} - \psi_0\right\| \leq \left\|\hat{\psi} - \psi_0\right\|$. Moreover from A.11 we have

$$E \sup_{\gamma} |\rho_1'(\gamma) W_1 \rho_1(\gamma)| \;\; < \;\; \infty.$$

Therefore, it follows from ULLN that the first term on the right-hand side of (A.12) satisfies

$$\sup_{\gamma} \left| \frac{1}{n} \sum_{i=1}^{n} \rho_i'(\gamma) W_i \rho_i(\gamma) - Q(\gamma) \right| \xrightarrow{a.s.} 0,$$

where $Q(\gamma) = E\rho_1'(\gamma) D_1 \rho_1(\gamma)$. Similarly, by Cauchy-Schwarz inequality and ASSUMPTION 18

$$\begin{aligned}
& \left(E \sup_{\gamma,\psi} \left\| \rho_1'(\gamma,\psi) W_1 \frac{\partial \rho_1(\gamma,\psi)}{\partial \psi'} \right\| \right)^2 \\
\leq \;\; & E \|W_1\| \sup_{\gamma,\psi} \|\rho_1(\gamma,\psi)\|^2 \, E \|W_1\| \sup_{\gamma,\psi} \left\| \frac{\partial \rho_1(\gamma,\psi)}{\partial \psi'} \right\|^2 \\
\leq \;\; & pE \|W_1\| \sup_{\gamma,\psi} \|\rho_1(\gamma,\psi)\|^2 \, E \|W_1\| \, \|V_1\|^2 \\
& \left(\int \sup_{\gamma,\psi} \left\| f(x'\beta) \frac{\partial f_U(x - HV_1;\phi)}{\partial u'} (\|x\| + 1) \right\| dx \right)^2 \\
+ \;\; & pE \|W_1\| \sup_{\gamma,\psi} \|\rho_1(\gamma,\psi)\|^2 \, E \|W_1\| \, \|V_1\|^2 \\
& \left(\int \sup_{\gamma,\psi} \left\| f^2(x'\beta) \frac{\partial f_U(x - HV_1;\phi)}{\partial u'} \right\| dx \right)^2 \\
+ \;\; & pE \|W_1\| \sup_{\gamma,\psi} \|\rho_1(\gamma,\psi)\|^2 \, E \|W_1\| \, \|V_1\|^2 \\
& \left(\int \sup_{\gamma,\psi} \left\| g^2(x'\beta;\theta) \frac{\partial f_U(x - HV_1;\phi)}{\partial u'} \right\| dx \right)^2 < \infty,
\end{aligned}$$

then again by the ULLN we have

$$\sup_{\gamma,\psi}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho_i'(\gamma,\psi)W_i\frac{\partial\rho_i(\gamma,\psi)}{\partial\psi'}\right\| = O(1) \quad (a.s.)$$

and therefore, since $\hat{\psi} = \text{vec}\hat{H}$ is the least squares estimator of ψ_0,

$$\begin{aligned}
&\sup_{\gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho_i'(\gamma,\tilde{\psi})W_i\frac{\partial\rho_i(\gamma,\tilde{\psi})}{\partial\psi'}\left(\hat{\psi}-\psi_0\right)\right\| \\
\leq\ &\sup_{\gamma,\psi}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho_i'(\gamma,\psi)W_i\frac{\partial\rho_i(\gamma,\psi)}{\partial\psi'}\right\|\left\|\hat{\psi}-\psi_0\right\| \xrightarrow{a.s.} 0. \quad \text{(A.13)}
\end{aligned}$$

It follows from (A.12) - (A.13) that

$$\sup_{\gamma}\left|\frac{1}{n}Q_n(\gamma)-Q(\gamma)\right| \xrightarrow{a.s.} 0.$$

In the previous part of the theorem, we showed that $Q(\gamma)$ attains a unique minimum at $\gamma_0 \in \Gamma$. Thus, it follows from Lemma 2.1.1 that $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$.

A.6 Proof of Theorem 3.2.5

In the first step of the proof, using Lemma 2.1.2, we need to show that $E\sup_{\gamma}\left\|\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'}\right\| < \infty$. By the Dominated Convergence Theorem, the first derivative $\partial Q_n(\gamma)/\partial\gamma$ has the first-order Taylor expansion in the open neighborhood $\Gamma_0 \subset \Gamma$ of γ_0. Since $\partial Q_n(\hat{\gamma}_n)/\partial\gamma = 0$ and $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, for sufficiently large n we have

$$\frac{\partial Q_n(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_n(\tilde\gamma)}{\partial\gamma\partial\gamma'}(\hat\gamma_n - \gamma_0) = 0, \qquad \text{(A.14)}$$

where $\|\tilde\gamma - \gamma_0\| \le \|\hat\gamma_n - \gamma_0\|$. The first and the second derivative of $Q_n(\gamma)$ in (A.14) is given by

$$\frac{\partial Q_n(\gamma)}{\partial\gamma} = 2\sum_{i=1}^{n} \frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \rho_i(\gamma),$$

and

$$\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'} = 2\sum_{i=1}^{n}\left[\frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \frac{\partial\rho_i(\gamma)}{\partial\gamma'} + (\rho_i'(\gamma)W_i \otimes I_{p+k+1})\frac{\partial \mathrm{vec}(\partial\rho_i'(\gamma)/\partial\gamma)}{\partial\gamma'}\right],$$

respectively. Assumptions 15 – 18 imply that

$$E\sup_\gamma\left\|\frac{\partial\rho_1'(\gamma)}{\partial\gamma} W_1 \frac{\partial\rho_1(\gamma)}{\partial\gamma'}\right\| < \infty.$$

and

$$\left(E\sup_\gamma\left\|(\rho_1'(\gamma)W_1 \otimes I_{p+k+1})\frac{\partial \mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|\right)^2 < \infty.$$

Since $\partial\mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)/\partial\gamma'$ depends on V_1 only and therefore

$$\begin{aligned} & E\left[(\rho_1'(\gamma_0)W_1 \otimes I_{p+k+1})\,\frac{\partial \mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \\ = \; & E\left[(E(\rho_1'(\gamma_0)|V_1)W_1 \otimes I_{p+k+1})\,\frac{\partial \mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \\ = \; & 0, \end{aligned}$$

it follows from the ULLN and Lemma 2.1.2 that

$$\begin{aligned}
& \frac{1}{2n}\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'} \\
\xrightarrow{a.s.} \quad & E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'} + \left(\rho_1'(\gamma_0)W_1\otimes I_{p+k+1}\right)\frac{\partial \text{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \\
= \quad & B.
\end{aligned}$$

Now, we know that $\frac{\partial\rho_i'(\gamma)}{\partial\gamma}W_i\rho_i(\gamma), i = 1, 2, ..., n$ are *i.i.d.* with the mean vector zero and the variance covariance matrix A, where A is given in Theorem 3.2.5. Therefore, by Slutsky's Theorem, we have

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{L} N(0, B^{-1}AB^{-1}).$$

A.7 Proof of Theorem 3.2.6

Similar to Theorem 3.2.5, using Lemma 2.1.2, we can show that $E\sup_{\gamma,\psi}\left\|\frac{\partial^2 Q_n(\tilde{\gamma})}{\partial\gamma\partial\gamma'}\right\| < \infty$. By ASSUMPTIONS 16 – 18 and the Dominated Convergence Theorem, the first derivative $\partial Q_n(\gamma)/\partial\gamma$ exists and has the first-order Taylor expansion in the open neighborhood $\Gamma_0 \subset \Gamma$ of γ_0. Since $\partial Q_n(\hat{\gamma}_n)/\partial\gamma = 0$ and $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, for sufficiently large n we have

$$\frac{\partial Q_n(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_n(\tilde{\gamma})}{\partial\gamma\partial\gamma'}(\hat{\gamma}_n - \gamma_0) = 0, \qquad \text{(A.15)}$$

where $\|\tilde{\gamma} - \gamma_0\| \le \|\hat{\gamma}_n - \gamma_0\|$. The first derivative of $Q_n(\gamma)$ in (A.15) is given by

$$\frac{\partial Q_n(\gamma)}{\partial \gamma} = 2 \sum_{i=1}^{n} \frac{\partial \hat{\rho}_i'(\gamma)}{\partial \gamma} W_i \hat{\rho}_i(\gamma),$$

where $\partial \hat{\rho}_i(\gamma)/\partial \gamma'$ consists of the following nonzero elements

$$\begin{aligned}
\frac{\partial m_1(\hat{H}V_i;\gamma)}{\partial \beta'} &= \int \frac{\partial f(x'\beta)}{\partial \beta'} f_U(x - \hat{H}V_i;\phi)dx, \\
\frac{\partial m_2(\hat{H}V_i;\gamma)}{\partial \beta'} &= \int \frac{\partial g^2(x'\beta;\theta)}{\partial \beta'} f_U(x - \hat{H}V_i;\phi)dx \\
&+ \int \frac{\partial f^2(x'\beta)}{\partial \beta} f_U(x - \hat{H}V_i;\phi)dx, \\
\frac{\partial m_3(\hat{H}V_i;\gamma)}{\partial \beta'} &= \int x \frac{\partial f(x'\beta)}{\partial \beta'} f_U(x - \hat{H}V_i;\phi)dx,
\end{aligned}$$

$$\frac{\partial m_2(\hat{H}V_i;\gamma)}{\partial \theta'} = \int \frac{\partial g^2(x'\beta;\theta)}{\partial \theta'} f_U(x - \hat{H}V_i;\phi)dx,$$

$$\begin{aligned}
\frac{\partial m_1(\hat{H}V_i;\gamma)}{\partial \phi'} &= \int f(x'\beta) \frac{\partial f_U(x - \hat{H}V_i;\phi)}{\partial \phi'} dx, \\
\frac{\partial m_2(\hat{H}V_i;\gamma)}{\partial \phi'} &= \int f^2(x'\beta) \frac{\partial f_U(x - \hat{H}V_i;\phi)}{\partial \phi'} dx \\
&+ \int g^2(x'\beta;\theta) \frac{\partial f_U(x - \hat{H}V_i;\phi)}{\partial \phi'} dx, \\
\frac{\partial m_3(\hat{H}V_i;\gamma)}{\partial \phi'} &= \int x f(x'\beta) \frac{\partial f_U(x - \hat{H}V_i;\phi)}{\partial \phi'} dx.
\end{aligned}$$

The second derivative of $Q_n(\gamma)$ in (A.15) is

$$\frac{\partial^2 Q_n(\gamma)}{\partial \gamma \partial \gamma'} = 2 \sum_{i=1}^{n} \left[\frac{\partial \hat{\rho}_i'(\gamma)}{\partial \gamma} W_i \frac{\partial \hat{\rho}_i(\gamma)}{\partial \gamma'} + (\hat{\rho}_i'(\gamma) W_i \otimes I_{p+k+1}) \frac{\partial \mathrm{vec}(\partial \hat{\rho}_i'(\gamma)/\partial \gamma)}{\partial \gamma'} \right],$$

where $\partial \text{vec}(\partial \hat{\rho}_i'(\gamma)/\partial\gamma)/\partial\gamma'$ consists of the following nonzero elements

$$
\begin{aligned}
\frac{\partial^2 m_1(\hat{H}V_i;\gamma)}{\partial\beta\partial\beta'} &= \int \frac{\partial^2 f(x'\beta)}{\partial\beta\partial\beta'} f_U(x-\hat{H}V_i;\phi)dx, \\
\frac{\partial^2 m_2(\hat{H}V_i;\gamma)}{\partial\beta\partial\beta'} &= \int \frac{\partial^2 f^2(x'\beta)}{\partial\beta\partial\beta'} f_U(x-\hat{H}V_i;\phi)dx \\
&+ \int \frac{\partial^2 g^2(x'\beta;\theta)}{\partial\beta\partial\beta'} f_U(x-\hat{H}V_i;\phi)dx, \\
\frac{\partial^2 m_3(\hat{H}V_i;\gamma)}{\partial\beta\partial\beta'} &= \int x \otimes \frac{\partial^2 f(x'\beta)}{\partial\beta\partial\beta'} f_U(x-\hat{H}V_i;\phi)dx,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 m_2(\hat{H}V_i;\gamma)}{\partial\beta\partial\theta'} &= \int \frac{\partial^2 g^2(x'\beta;\theta)}{\partial\beta\partial\theta'} f_U(x-\hat{H}V_i;\phi)dx, \\
\frac{\partial^2 m_2(\hat{H}V_i;\gamma)}{\partial\theta\partial\theta'} &= \int \frac{\partial^2 g^2(x'\beta;\theta)}{\partial\theta\partial\theta'} f_U(x-\hat{H}V_i;\phi)dx,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 m_1(\hat{H}V_i;\gamma)}{\partial\phi\partial\beta'} &= \int \frac{\partial f_U(x-\hat{H}V_i;\phi)}{\partial\phi} \frac{\partial f(x'\beta)}{\partial\beta'} dx, \\
\frac{\partial^2 m_2(\hat{H}V_i;\gamma)}{\partial\phi\partial\beta'} &= \int \frac{\partial f_U(x-\hat{H}V_i;\phi)}{\partial\phi} \frac{\partial f^2(x'\beta)}{\partial\beta'} dx \\
&+ \int \frac{\partial f_U(x-\hat{H}V_i;\phi)}{\partial\phi} \frac{\partial g^2(x'\beta;\theta)}{\partial\beta'} dx, \\
\frac{\partial^2 m_3(\hat{H}V_i;\gamma)}{\partial\phi\partial\beta'} &= \int x \otimes \frac{\partial f_U(x-\hat{H}V_i;\phi)}{\partial\phi} \frac{\partial f(x'\beta)}{\partial\beta'} dx, \\
\frac{\partial^2 m_2(\hat{H}V_i;\gamma)}{\partial\phi\partial\theta'} &= \int \frac{\partial f_U(x-\hat{H}V_i;\phi)}{\partial\phi} \frac{\partial g^2(x'\beta;\theta)}{\partial\theta'} dx,
\end{aligned}
$$

and

$$\begin{aligned}
\frac{\partial^2 m_1(\hat{H}V_i;\gamma)}{\partial\phi\partial\phi'} &= \int \frac{\partial^2 f_U(x-\hat{H}V_i;\phi)}{\partial\phi\partial\phi'} f(x'\beta)dx, \\
\frac{\partial^2 m_2(\hat{H}V_i;\gamma)}{\partial\phi\partial\phi'} &= \int \frac{\partial^2 f_U(x-\hat{H}V_i;\phi)}{\partial\phi\partial\phi'} f^2(x'\beta)dx \\
&+ \int \frac{\partial^2 f_U(x-\hat{H}V_i;\phi)}{\partial\phi\partial\phi'} g^2(x'\beta;\theta)dx, \\
\frac{\partial^2 m_3(\hat{H}V_i;\gamma)}{\partial\phi\partial\phi'} &= \int x \otimes \frac{\partial^2 f_U(x-\hat{H}V_i;\phi)}{\partial\phi\partial\phi'} f(x'\beta)dx.
\end{aligned}$$

It follows from Assumptions 15 – 18 that

$$
\begin{aligned}
& E \sup_{\gamma,\psi} \left\| \frac{\partial \rho_1'(\gamma,\psi)}{\partial \gamma} W_1 \frac{\partial \rho_1(\gamma,\psi)}{\partial \gamma'} \right\| \\
\leq \;& E \left\| W_1 \right\| \sup_{\gamma,\psi} \left\| \frac{\partial \rho_1(\gamma,\psi)}{\partial \gamma'} \right\|^2 \\
= \;& E \left\| W_1 \right\| \sup_{\psi,\gamma} \left(\left\| \frac{\partial m_1(HV_i;\gamma)}{\partial \beta'} \right\|^2 + \left\| \frac{\partial m_2(HV_i;\gamma)}{\partial \beta'} \right\|^2 \left\| \frac{\partial m_3(HV_i;\gamma)}{\partial \beta'} \right\|^2 \right) \\
+ \;& E \left\| W_1 \right\| \sup_{\psi,\gamma} \left(\left\| \frac{\partial m_1(HV_i;\gamma)}{\partial \phi'} \right\|^2 + \left\| \frac{\partial m_2(HV_i;\gamma)}{\partial \phi'} \right\|^2 \left\| \frac{\partial m_3(HV_i;\gamma)}{\partial \phi'} \right\|^2 \right) \\
+ \;& E \left\| W_1 \right\| \sup_{\psi,\gamma} \left(\left\| \frac{\partial m_2(HV_i;\gamma)}{\partial \theta'} \right\| \right)^2 \\
\leq \;& E \left\| W_1 \right\| \left(\int \sup_{\psi,\gamma} \left\| \frac{\partial f(x'\beta)}{\partial \beta'} f_U(x - HV_1;\phi) \right\| (\|x\| + 1) dx \right)^2 \\
+ \;& E \left\| W_1 \right\| \left(\int \sup_{\psi,\gamma} \left\| \frac{\partial f^2(x'\beta)}{\partial \beta'} f_U(x - HV_1;\phi) \right\| dx \right. \\
+ \;& \left. \left\| \frac{\partial g^2(x'\beta;\theta)}{\partial \beta'} f_U(x - HV_1;\phi) \right\| dx \right)^2 \\
+ \;& E \left\| W_1 \right\| \left(\int \sup_{\psi,\gamma} \left\| \frac{\partial g^2(x'\beta;\theta)}{\partial \theta'} f_U(x - HV_1;\phi) \right\| dx \right)^2 \\
+ \;& E \left\| W_1 \right\| \left(\int \sup_{\psi,\gamma} \left\| \frac{\partial f_U(x - HV_i;\phi)}{\partial \phi'} f(x'\beta) dx \right\| (\|x\| + 1) dx \right)^2 \\
+ \;& E \left\| W_1 \right\| \left(\int \sup_{\psi,\gamma} \left\| \frac{\partial f_U(x - HV_i;\phi)}{\partial \phi'} (f^2(x'\beta) + g^2(x'\beta;\theta)) \right\| dx \right)^2 < \infty.
\end{aligned}
$$

Similarly, by ASSUMPTIONS 15 – 18 we have

$$
\begin{aligned}
& \left(E \sup_{\gamma,\psi} \left\| (\rho_1'(\gamma,\psi) W_1 \otimes I_{p+k+1}) \frac{\partial \text{vec}(\partial \rho_1'(\gamma,\psi)/\partial\gamma)}{\partial \gamma'} \right\| \right)^2 \\
\leq \; & (p+k+1) \left(E \|W_1\| \sup_{\gamma,\psi} \|\rho_1(\gamma,\psi)\| \left\| \frac{\partial \text{vec}(\partial \rho_1'(\gamma,\psi)/\partial\gamma)}{\partial \gamma'} \right\| \right)^2 \\
\leq \; & (p+k+1) E \|W_1\| \sup_{\gamma,\psi} \|\rho_1(\gamma,\psi)\|^2 E \|W_1\| \sup_{\gamma,\psi} \left\| \frac{\partial \text{vec}(\partial \rho_1'(\gamma,\psi)/\partial\gamma)}{\partial \gamma'} \right\|^2 \\
< \; & \infty.
\end{aligned}
$$

Since $\partial \text{vec}(\partial \rho_1'(\gamma_0)/\partial\gamma)/\partial\gamma'$ depends on V_1 only and therefore

$$
\begin{aligned}
& E \left[(\rho_1'(\gamma_0) W_1 \otimes I_{p+q}) \frac{\partial \text{vec}(\partial \rho_1'(\gamma_0)/\partial\gamma)}{\partial \gamma'} \right] \\
= \; & E \left[(E(\rho_1'(\gamma_0)|V_1) W_1 \otimes I_{p+k+1}) \frac{\partial \text{vec}(\partial \rho_1'(\gamma_0)/\partial\gamma)}{\partial \gamma'} \right] \\
= \; & 0,
\end{aligned}
$$

it follows from the ULLN and Lemma 2.1.2 that

$$
\begin{aligned}
& \frac{1}{2n} \frac{\partial^2 Q_n(\tilde{\gamma})}{\partial\gamma\partial\gamma'} \\
\xrightarrow{a.s.} \; & E \left[\frac{\partial \rho_1'(\gamma_0)}{\partial\gamma} W_1 \frac{\partial \rho_1(\gamma_0)}{\partial\gamma'} + (\rho_1'(\gamma_0) W_1 \otimes I_{p+k+1}) \frac{\partial \text{vec}(\partial \rho_1'(\gamma_0)/\partial\gamma)}{\partial \gamma'} \right] \\
= \; & B. \qquad \text{(A.16)}
\end{aligned}
$$

Similarly, for sufficiently large n, $\partial Q_n(\gamma_0)/\partial\gamma$, has the first-order Taylor expansion about ψ_0:

$$
\frac{\partial Q_n(\gamma_0)}{\partial\gamma} = 2 \sum_{i=1}^{n} \frac{\partial \rho_i'(\gamma_0)}{\partial\gamma} W_i \rho_i(\gamma_0) + \frac{\partial^2 \tilde{Q}_n(\gamma_0)}{\partial\gamma\partial\psi'} \left(\hat{\psi} - \psi_0 \right), \qquad \text{(A.17)}
$$

where

$$\frac{\partial^2 \tilde{Q}_n(\gamma_0)}{\partial\gamma\partial\psi'} =$$
$$2\sum_{i=1}^{n}\left[\frac{\partial\rho_i'(\gamma_0,\tilde{\psi})}{\partial\gamma}W_i\frac{\partial\rho_i(\gamma_0,\tilde{\psi})}{\partial\psi'} + \left(\rho_i'(\gamma_0,\tilde{\psi})W_i \otimes I_{p+k+1}\right)\frac{\partial \mathrm{vec}(\partial\rho_i'(\gamma_0,\tilde{\psi})/\partial\gamma)}{\partial\psi'}\right],$$

and $\tilde{\psi} = \mathrm{vec}\tilde{H}$ satisfies $\left\|\tilde{\psi}-\psi_0\right\| \leq \left\|\hat{\psi}-\psi_0\right\|$. Similar to (A.16), it can be shown that

$$\frac{1}{2n}\frac{\partial^2 \tilde{Q}_n(\gamma_0)}{\partial\gamma\partial\psi'} \xrightarrow{a.s.} E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\psi'}\right]. \tag{A.18}$$

By definition (3.30), $\hat{H}-H = \left[\sum_{i=1}^n (Z_i - HV_i)V_i'\right]\left(\sum_{i=1}^n V_iV_i'\right)^{-1}$, which can be written as

$$\hat{\psi}-\psi_0 = \mathrm{vec}(\hat{H}-H) = \left(\sum V_iV_i' \otimes I_p\right)^{-1}\sum_{i=1}^{n} V_i \otimes (Z_i - HV_i)$$

(Magnus and Neudecker (1988), p.30). Hence (A.17) can be written as

$$\frac{\partial Q_n(\gamma_0)}{\partial\gamma} = 2C_n\sum_{i=1}^{n} B_i,$$

where

$$C_n = \left(I_{p+k+1}, \frac{1}{2}\frac{\partial^2\tilde{Q}_n(\gamma_0)}{\partial\gamma\partial\psi'}\left(\sum_{i=1}^{n} V_iV_i'\otimes I_p\right)^{-1}\right)$$

and

$$B_i = \begin{pmatrix} \frac{\partial\rho_i'(\gamma_0)}{\partial\gamma}W_i\rho_i(\gamma_0) \\ \\ V_i\otimes(Z_i - HV_i)\end{pmatrix}.$$

Now, we know that $B_i, i = 1, 2, ..., n$ are *i.i.d.* with the mean vector zero and the variance covariance matrix A, where A is given in Theorem 3.2.6. On the other hand, by the Law of Large Numbers, $n\left(\sum V_i V_i' \otimes I_p\right)^{-1} \xrightarrow{a.s.} (EV_1V_1' \otimes I_p)^{-1}$, which together with (A.18) implies

$$C_n \xrightarrow{a.s.} \left(I_{p+k+1}, E\left[\frac{\partial \rho_1'(\gamma_0)}{\partial \gamma} W_1 \frac{\partial \rho_1(\gamma_0)}{\partial \psi'}\right] \left(EV_1V_1' \otimes I_p\right)^{-1} \right) = C,$$

where C is given in Theorem 3.2.6. Therefore, by the Central Limit Theorem, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_i \xrightarrow{L} N(0, A)$, where $A = E(B_1 B_1')$. Therefore, by Slutsky's Theorem, we have

$$\frac{1}{2\sqrt{n}} \frac{\partial Q_n(\gamma_0)}{\partial \gamma} \xrightarrow{L} N(0, CAC'). \tag{A.19}$$

Finally, the theorem follows from (A.15), (A.16) and (A.19).

A.8 Proof of Theorem 3.2.7

Since

$$
\begin{aligned}
& E\left(\sup_{\gamma}\|\rho_{1,S}(\gamma)\|\,|V_1, Y_1, Z_1\right) \\
\leq\ & \|Y_1T_1\| + E\left(\sup_{\gamma}\|m_S(HV_1;\gamma)\|\,|V_1\right) \\
\leq\ & \|Y_1T_1\| + \frac{1}{S}\sum_{j=1}^{S} E\left(\sup_{\gamma}\left\|\frac{f(x'_{ij}\beta)f_U(x_{ij}-HV_1;\phi)}{l(x_{ij})}\right\||V_1\right) \\
+\ & \frac{1}{S}\sum_{j=1}^{S} E\left(\sup_{\gamma}\left\|\frac{f^2(x'_{ij}\beta)f_U(x_{ij}-HV_1;\phi)}{l(x_{ij})}\right\||V_1\right) \\
+\ & \frac{1}{S}\sum_{j=1}^{S} E\left(\sup_{\gamma}\left\|\frac{g^2(x'_{ij}\beta;\theta)f_U(x_{ij}-HV_1;\phi)}{l(x_{ij})}\right\||V_1\right) \\
+\ & \frac{1}{S}\sum_{j=1}^{S} E\left(\sup_{\gamma}\left\|\frac{x_{ij}f(x'_{ij}\beta)f_U(x_{ij}-HV_1;\phi)}{l(x_{ij})}\right\||V_1\right) \\
\leq\ & \|Y_1T_1\| + \int\sup_{\gamma} f(x'\beta)f_U(x-HV_1;\phi)(\|x\|+1)dx \\
+\ & \int\sup_{\gamma} f^2(x'\beta)f_U(x-HV_1;\phi)dx \\
+\ & \int\sup_{\gamma} g^2(x'\beta;\theta)f_U(x-HV_1;\phi)dx,
\end{aligned}
$$

and similarly

$$
\begin{aligned}
& E\left(\sup_{\gamma} \|\rho_{1,2S}(\gamma)\| \,|V_1, Y_1, Z_1\right) \\
\leq\ & \|Y_1 T_1\| + \int \sup_{\gamma} f(x'\beta) f_U(x - HV_1; \phi)(\|x\| + 1) dx \\
+\ & \int \sup_{\gamma} f^2(x'\beta) f_U(x - HV_1; \phi) dx \\
+\ & \int \sup_{\gamma} g^2(x'\beta; \theta) f_U(x - HV_1; \phi) dx,
\end{aligned}
$$

and because $\rho_{1,S}(\gamma)$ and $\rho_{1,2S}(\gamma)$ are conditionally independent given (V_1, Y_1, Z_1), it follows from ASSUMPTIONS 15 – 16 that

$$
\begin{aligned}
& E \sup_{\gamma} \left|\rho'_{1,S}(\gamma) W_1 \rho_{1,2S}(\gamma)\right| \\
\leq\ & E\,\|W_1\|\, E\left(\sup_{\gamma} \|\rho_{1,S}(\gamma)\| \,|V_1, Y_1, Z_1\right) E\left(\sup_{\gamma} \|\rho_{1,2S}(\gamma)\| \,|V_1, Y_1, Z_1\right) \\
\leq\ & E\,\|W_1\| \Bigg(\|Y_1 T_1\| + \int \sup_{\gamma} f(x'\beta) f_U(x - HV_1; \phi)(\|x\| + 1) dx \\
+\ & \int \sup_{\gamma} f^2(x'\beta) f_U(x - HV_1; \phi) dx \\
+\ & \int \sup_{\gamma} g^2(x'\beta; \theta) f_U(x - HV_1; \phi) dx\Bigg)^2 \\
\leq\ & 2E\,\|W_1\| \left(Y_1^4 + \|Y_1 Z_1\|^2\right) \\
+\ & 2E\,\|W_1\| \Bigg(\int \sup_{\gamma} f(x'\beta) f_U(x - HV_1; \phi)(\|x\| + 1) dx \\
+\ & \int \sup_{\gamma} f^2(x'\beta) f_U(x - HV_1; \phi) dx \\
+\ & \int \sup_{\gamma} g^2(x'\beta; \theta) f_U(x - HV_1; \phi) dx\Bigg)^2 \\
<\ & \infty.
\end{aligned}
$$

Therefore by the ULLN we have

$$\sup_{\gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\rho'_{i,S}(\gamma)W_i\rho_{i,2S}(\gamma)-E\rho'_{1,S}(\gamma)W_1\rho_{1,2S}(\gamma)\right|\xrightarrow{a.s.}0,$$

where

$$\begin{aligned}&E\rho'_{1,S}(\gamma)W_1\rho_{1,2S}(\gamma)\\=\;&E\left[E\left(\rho'_{1,S}(\gamma)|V_1,Y_1,Z_1\right)W_1E\left(\rho'_{1,2S}(\gamma)|V_1,Y_1,Z_1\right)\right]\\=\;&E\rho'_1(\gamma)W_1\rho_1(\gamma)\\=\;&Q(\gamma).\end{aligned}$$

Therefore,

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

We showed in the proof of Theorem 3.2.4 that $Q(\gamma)$ attains a unique minimum at $\gamma_0\in\Gamma$. Therefore $\hat{\gamma}_{n,S}\xrightarrow{a.s.}\gamma_0$ follows from Lemma 2.1.1.

To prove the second part of the theorem, first, $\partial Q_{n,S}(\gamma)/\partial\gamma$ has the first-order Taylor expansion in an open neighborhood $\Gamma_0\subset\Gamma$ of γ_0:

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

where $\|\tilde{\gamma}-\gamma_0\|\le\|\hat{\gamma}_{n,S}-\gamma_0\|$. The first and the second derivative of $Q_{n,S}(\gamma)$ is given by

$$\frac{\partial Q_{n,S}(\gamma)}{\partial\gamma}=\sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,S}(\gamma)}{\partial\gamma}W_i\rho_{i,2S}(\gamma)+\frac{\partial\rho'_{i,2S}(\gamma)}{\partial\gamma}W_i\rho_{i,S}(\gamma)\right],$$

and

$$
\begin{aligned}
&\frac{\partial^2 Q_{n,S}(\gamma)}{\partial\gamma\partial\gamma'} \\
= &\sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,S}(\gamma)}{\partial\gamma}W_i\frac{\partial\rho_{i,2S}(\gamma)}{\partial\gamma'} + (\rho'_{i,2S}(\gamma)W_i\otimes I_{p+k+1})\frac{\partial\text{vec}(\partial\rho'_{i,S}(\gamma)/\partial\gamma)}{\partial\gamma'}\right] \\
+ &\sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,2S}(\gamma)}{\partial\gamma}W_i\frac{\partial\rho_{i,S}(\gamma)}{\partial\gamma'} + (\rho'_{i,S}(\gamma)W_i\otimes I_{p+k+1})\frac{\partial\text{vec}(\partial\rho'_{i,2S}(\gamma)/\partial\gamma)}{\partial\gamma'}\right],
\end{aligned}
$$

respectively. Similar to (A.16), we can show that $\frac{1}{n}\frac{\partial^2 Q_{n,S}(\gamma)}{\partial\gamma\partial\gamma'}$ converges $a.s.$ to

$$
\begin{aligned}
&E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'} + (\rho'_{1,2S}(\gamma_0)W_1\otimes I_{p+k+1})\frac{\partial\text{vec}(\partial\rho'_{1,S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] + \\
&E\left[\frac{\partial\rho'_{1,2S}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'} + (\rho'_{1,S}(\gamma_0)W_1\otimes I_{p+k+1})\frac{\partial\text{vec}(\partial\rho'_{1,2S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]
\end{aligned}
$$

uniformly for all $\gamma\in\Gamma$. Since

$$
E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right] = E\left[\frac{\partial\rho'_{1}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1}(\gamma_0)}{\partial\gamma'}\right] = B
$$

and

$$
E\left[(\rho'_{1,2S}(\gamma_0)W_1\otimes I_{p+k+1})\frac{\partial\text{vec}(\partial\rho'_{1,S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] = 0,
$$

we have

$$
\frac{1}{n}\frac{\partial^2 Q_{n,S}(\gamma)}{\partial\gamma\partial\gamma'} \xrightarrow{a.s.} 2B. \tag{A.20}
$$

Further, by the Central Limit Theorem we have

$$
\frac{1}{2\sqrt{n}}\frac{\partial Q_{n,S}(\gamma)}{\partial\gamma} \xrightarrow{L} N(0, A_S), \tag{A.21}
$$

where

$$
\begin{aligned}
A_S &= \frac{1}{4}E\left[\left(\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,2S}(\gamma_0) + \frac{\partial\rho'_{1,2S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,S}(\gamma_0)\right)\right. \\
&\times \left.\left(\rho'_{1,2S}(\gamma_0)W_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'} + \rho'_{1,S}(\gamma_0)W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right)\right] \\
&= \frac{1}{2}E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,2S}(\gamma_0)\rho'_{1,2S}(\gamma_0)W_1\frac{\partial\rho_{1,S}(\psi_0)}{\partial\gamma'}\right] \\
&+ \frac{1}{2}E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,2S}(\gamma_0)\rho'_{1,S}(\gamma_0)W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right],
\end{aligned}
$$

Finally, the second part of Theorem 3.2.7 follows from (A.20) and (A.21), and Slutsky's Theorem .

A.9 Proof of Theorem 3.2.8

The proof of the first part of Theorem 3.2.8 is similar to that for Theorem 3.2.4. First, for sufficiently large n, $Q_{n,S}(\gamma)$ has the first-order Taylor expansion about ψ_0:

$$
\begin{aligned}
Q_{n,S}(\gamma) &= \sum_{i=1}^{n}\rho'_{i,S}(\gamma)W_i\rho_{i,2S}(\gamma) \qquad \text{(A.22)} \\
&+ \sum_{i=1}^{n}\left[\rho'_{i,S}(\gamma,\tilde{\psi})W_i\frac{\partial\rho_{i,2S}(\gamma,\tilde{\psi})}{\partial\psi'} + \rho'_{i,2S}(\gamma,\tilde{\psi})W_i\frac{\partial\rho_{i,S}(\gamma,\tilde{\psi})}{\partial\psi'}\right]\left(\hat{\psi}-\psi_0\right),
\end{aligned}
$$

where $\rho_{i,S}(\gamma) = Y_iT_i - m_S(HV_i;\gamma)$, $\rho_{i,S}(\gamma,\tilde{\psi}) = Y_iT_i - m_S(\tilde{H}V_i;\gamma)$, $\left\|\tilde{\psi}-\psi_0\right\| \le \left\|\hat{\psi}-\psi_0\right\|$ and $\rho_{i,2S}(\gamma,\tilde{\psi})$ is given similarly. Further, because $\rho_{1,S}(\gamma)$ and $\rho_{1,2S}(\gamma)$ are conditionally independent given (V_1, Y_1, Z_1), it follows from ASSUMPTIONS 15 – 16 that

$$E \sup_{\gamma} \left| \rho'_{1,S}(\gamma) W_1 \rho_{1,2S}(\gamma) \right| < \infty.$$

Therefore by the ULLN the first term on the right-hand side of (A.22) satisfies

$$\sup_{\gamma} \left| \frac{1}{n} \sum_{i=1}^{n} \rho'_{i,S}(\gamma) W_i \rho_{i,2S}(\gamma) - E\rho'_{1,S}(\gamma) W_1 \rho_{1,2S}(\gamma) \right| \xrightarrow{a.s.} 0, \tag{A.23}$$

where

$$E\rho'_{1,S}(\gamma) W_1 \rho_{1,2S}(\gamma) = Q(\gamma).$$

The second term on the right-hand side of (A.22) satisfies

$$\begin{aligned}
& E \sup_{\gamma,\psi} \left\| \rho'_{1,S}(\gamma,\psi) W_1 \frac{\partial \rho_{1,2S}(\gamma,\psi)}{\partial \psi'} \right\| \\
\leq \; & E \left\| W_1 \right\| \sup_{\gamma,\psi} \left\| \rho'_{1,S}(\gamma,\psi) \right\| \left\| \frac{\partial \rho_{1,2S}(\gamma,\psi)}{\partial \psi'} \right\| \\
\leq \; & E \left\| W_1 \right\| E \left(\sup_{\gamma,\psi} \left\| \rho_{1,S}(\gamma,\psi) \right\| | V_1, Y_1, Z_1 \right) \\
& E \left(\sup_{\gamma,\psi} \left\| \frac{\partial \rho_{1,2S}(\gamma,\psi)}{\partial \psi'} \right\| | V_1, Y_1, Z_1 \right).
\end{aligned}$$

Since

$$
\begin{aligned}
& E\left(\sup_{\gamma,\psi}\left\|\frac{\partial\rho_{1,2S}(\gamma,\psi)}{\partial\psi'}\right\| |V_1, Y_1, Z_1\right) \\
\leq\ & \frac{1}{S}\sum_{j=S+1}^{2S} E\left(\sup_{\gamma,\psi}\left\|\frac{f(x_{1j}'\beta)}{l(x_{1j})}\frac{\partial f_U(x_{1j}-HV_1;\phi)}{\partial u'}(V_1\otimes I_p)'\right\| |V_1\right) \\
+\ & \frac{1}{S}\sum_{j=S+1}^{2S} E\left(\sup_{\gamma,\psi}\left\|\frac{f^2(x_{1j}'\beta)}{l(x_{1j})}\frac{\partial f_U(x_{1j}-HV_1;\phi)}{\partial u'}(V_1\otimes I_p)'\right\| |V_1\right) \\
+\ & \frac{1}{S}\sum_{j=S+1}^{2S} E\left(\sup_{\gamma,\psi}\left\|\frac{g^2(x_{1j}'\beta;\theta)}{l(x_{1j})}\frac{\partial f_U(x_{1j}-HV_1;\phi)}{\partial u'}(V_1\otimes I_p)'\right\| |V_1\right) \\
+\ & \frac{1}{S}\sum_{j=S+1}^{2S} E\left(\sup_{\gamma,\psi}\left\|\frac{x_{1j}f(x_{1j}'\beta)}{l(x_{1j})}\frac{\partial f_U(x_{1j}-HV_1;\phi)}{\partial u'}(V_1\otimes I_p)'\right\| |V_1\right) \\
\leq\ & \sqrt{p}\,\|V_1\|\int\sup_{\gamma,\psi}\left\|f(x'\beta)\frac{\partial f_U(x-HV_1;\phi)}{\partial u'}\right\|(\|x\|+1)dx \\
+\ & \sqrt{p}\,\|V_1\|\int\sup_{\gamma,\psi}\left\|f^2(x'\beta)\frac{\partial f_U(x-HV_1;\phi)}{\partial u'}\right\|dx \\
+\ & \sqrt{p}\,\|V_1\|\int\sup_{\gamma,\psi}\left\|g^2(x'\beta;\theta)\frac{\partial f_U(x-HV_1;\phi)}{\partial u'}\right\|dx,
\end{aligned}
$$

by Cauchy-Schwarz inequality and ASSUMPTIONS 15 – 18,

$$
\left(E\sup_{\gamma,\psi}\left\|\rho_{1,S}'(\gamma,\psi)W_1\frac{\partial\rho_{1,2S}(\gamma,\psi)}{\partial\psi'}\right\|\right)^2<\infty.
$$

Thus by the ULLN we have

$$
\begin{aligned}
& \sup_{\gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho_{i,2S}'(\gamma,\tilde{\psi})W_i\frac{\partial\rho_{i,S}(\gamma,\tilde{\psi})}{\partial\psi'}\left(\hat{\psi}-\psi_0\right)\right\| \qquad \text{(A.24)} \\
\leq\ & \sup_{\gamma,\psi}\left\|\frac{1}{n}\sum_{i=1}^{n}\rho_{i,2S}'(\gamma,\psi)W_i\frac{\partial\rho_{i,S}(\gamma,\psi)}{\partial\gamma'}\right\|\left\|\left(\hat{\psi}-\psi_0\right)\right\|\xrightarrow{a.s.}0.
\end{aligned}
$$

It follows from (A.22) - (A.24) that

$$\sup_{\gamma} \left| \frac{1}{n} Q_{n,S}(\gamma) - Q(\gamma) \right| \xrightarrow{a.s.} 0. \tag{A.25}$$

It has been shown in the proof of THEOREM 3.2.4 that $Q(\gamma)$ attains a unique minimum at $\gamma_0 \in \Gamma$. Therefore $\hat{\gamma}_{n,S} \xrightarrow{a.s.} \gamma_0$ follows from Lemma 2.1.1.

The proof of the second part of Theorem 3.2.8 is analogous to that of Theorem 3.2.6. First, by ASSUMPTION 18, $\partial Q_{n,S}(\gamma)/\partial\gamma$ has the first-order Taylor expansion in an open neighborhood $\Gamma_0 \subset \Gamma$ of γ_0:

$$\frac{\partial Q_{n,S}(\gamma_0)}{\partial \gamma} + \frac{\partial^2 Q_{n,S}(\tilde{\gamma})}{\partial \gamma \partial \gamma'}(\hat{\gamma}_{n,S} - \gamma_0) = 0, \tag{A.26}$$

where $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma}_{n,S} - \gamma_0\|$. The first derivative of $Q_{n,S}(\gamma)$ in $(A.26)$ is given by

$$\frac{\partial Q_{n,S}(\gamma)}{\partial \gamma} = \sum_{i=1}^{n} \left[\frac{\partial \hat{\rho}'_{i,S}(\gamma)}{\partial \gamma} W_i \hat{\rho}_{i,2S}(\gamma) + \frac{\partial \hat{\rho}'_{i,2S}(\gamma)}{\partial \gamma} W_i \hat{\rho}_{i,S}(\gamma) \right],$$

and second derivative in $(A.26)$ is given by

$$\begin{aligned} & \frac{\partial^2 Q_{n,S}(\gamma)}{\partial \gamma \partial \gamma'} = \\ & \sum_{i=1}^{n} \left[\frac{\partial \hat{\rho}'_{i,S}(\gamma)}{\partial \gamma} W_i \frac{\partial \hat{\rho}_{i,2S}(\gamma)}{\partial \gamma'} + (\hat{\rho}'_{i,2S}(\gamma) W_i \otimes I_{p+k+1}) \frac{\partial \mathrm{vec}(\partial \hat{\rho}'_{i,S}(\gamma)/\partial \gamma)}{\partial \gamma'} \right] \\ + \; & \sum_{i=1}^{n} \left[\frac{\partial \hat{\rho}'_{i,2S}(\gamma)}{\partial \gamma} W_i \frac{\partial \hat{\rho}_{i,S}(\gamma)}{\partial \gamma'} + (\hat{\rho}'_{i,S}(\gamma) W_i \otimes I_{p+k+1}) \frac{\partial \mathrm{vec}(\partial \hat{\rho}'_{i,2S}(\gamma)/\partial \gamma)}{\partial \gamma'} \right]. \end{aligned}$$

Completely analogous to (A.16), we can show that $\frac{1}{n}\frac{\partial^2 Q_{n,S}(\gamma)}{\partial\gamma\partial\gamma'}$ converges $a.s.$ to

$$E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}+(\rho'_{1,2S}(\gamma_0)W_1\otimes I_{p+k+1})\frac{\partial \text{vec}(\partial\rho'_{1,S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]+$$
$$E\left[\frac{\partial\rho'_{1,2S}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'}+(\rho'_{1,S}(\gamma_0)W_1\otimes I_{p+k+1})\frac{\partial \text{vec}(\partial\rho'_{1,2S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]$$

uniformly for all $\gamma\in\Gamma$. Since

$$\begin{aligned}
& E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right]\\
= & E\left[E\left(\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}|V_1\right)W_1E\left(\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}|V_1\right)\right]\\
= & E\left[\frac{\partial\rho'_1(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}\right]=B
\end{aligned}$$

and

$$\begin{aligned}
& E\left[(\rho'_{1,2S}(\gamma_0)W_1\otimes I_{p+k+1})\frac{\partial \text{vec}(\partial\rho'_{1,S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]\\
= & E\left[\left(E(\rho'_{1,2S}(\gamma_0)|V_1)W_1\otimes I_{p+k+1}\right)\frac{\partial \text{vec}(\partial\rho'_{1,S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]=0,
\end{aligned}$$

we have

$$\frac{1}{n}\frac{\partial^2 Q_{n,S}(\tilde\gamma)}{\partial\gamma\partial\gamma'}\xrightarrow{a.s.}2B. \tag{A.27}$$

Again, by Assumption 18, $\partial Q_{n,S}(\gamma_0)/\partial\gamma$ has the first-order Taylor expansion about ψ_0:

$$\begin{aligned}
\frac{\partial Q_{n,S}(\gamma_0)}{\partial\gamma} = & \sum_{i=1}^n\left[\frac{\partial\rho'_{i,S}(\gamma_0)}{\partial\gamma}W_i\rho_{i,2S}(\gamma_0)+\frac{\partial\rho'_{i,2S}(\gamma_0)}{\partial\gamma}W_i\rho_{i,S}(\gamma_0)\right] \quad \text{(A.28)}\\
+ & \frac{\partial^2\tilde Q_{n,S}(\gamma_0)}{\partial\gamma\partial\psi'}\left(\hat\psi-\psi_0\right),
\end{aligned}$$

where

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

$$= \sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,S}(\gamma_0,\tilde{\psi})}{\partial\gamma}W_i\frac{\partial\rho_{i,2S}(\gamma_0,\tilde{\psi})}{\partial\psi'} + \left(\rho'_{i,2S}(\gamma_0,\tilde{\psi})W_i\otimes I_{p+k+1}\right)\frac{\partial \text{vec}(\partial\rho'_{i,S}(\gamma_0,\tilde{\psi})/\partial\gamma)}{\partial\psi'}\right]$$

$$+ \sum_{i=1}^{n}\left[\frac{\partial\rho'_{i,2S}(\gamma_0,\tilde{\psi})}{\partial\gamma}W_i\frac{\partial\rho_{i,S}(\gamma_0,\tilde{\psi})}{\partial\psi'} + \left(\rho'_{i,S}(\gamma_0,\tilde{\psi})W_i\otimes I_{p+k+1}\right)\frac{\partial \text{vec}(\partial\rho'_{i,2S}(\gamma_0,\tilde{\psi})/\partial\gamma)}{\partial\psi'}\right],$$

and $\left\|\tilde{\psi}-\psi_0\right\| \le \left\|\hat{\psi}-\psi_0\right\|$. Now rewrite (A.28) as

$$\frac{\partial Q_{n,S}(\gamma_0)}{\partial\gamma} = 2C_{n,S}\sum_{i=1}^{n}B_{i,S},$$

where

$$C_{n,S} = \left(I_{p+k+1}, \frac{1}{2}\frac{\partial^2\tilde{Q}_{n,S}(\gamma_0)}{\partial\gamma\partial\psi'}\left(\sum_{i=1}^{n}V_iV_i'\otimes I_p\right)^{-1}\right)$$

and

$$B_{i,S} = \frac{1}{2}\begin{pmatrix} \frac{\partial\rho'_{i,S}(\gamma_0)}{\partial\gamma}W_i\rho_{i,2S}(\gamma_0) + \frac{\partial\rho'_{i,2S}(\gamma_0)}{\partial\gamma}W_i\rho_{i,S}(\gamma_0) \\ 2V_i\otimes(Z_i - HV_i)\end{pmatrix}.$$

Then, analogous to (A.18), we can show that

$$\frac{1}{n}\frac{\partial^2\tilde{Q}_{n,S}(\gamma_0)}{\partial\gamma\partial\psi'} \xrightarrow{a.s.} 2E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\psi'}\right]$$

and hence

$$C_{n,S} \xrightarrow{a.s.} \left(I_{p+k+1}, E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\psi'}\right](EV_1V_1'\otimes I_p)^{-1}\right) = C. \quad \text{(A.29)}$$

Further, by the Central Limit Theorem we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}B_{i,S} \xrightarrow{L} N(0, A_S), \quad \text{(A.30)}$$

where

$$A_S = EB_{1,S}B'_{1,S} = \begin{pmatrix} A_{S,11} & A'_{S,21} \\ A_{S,21} & A_{S,22} \end{pmatrix},$$

$$\begin{aligned} A_{S,11} &= \frac{1}{4}E\left[\left(\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,2S}(\gamma_0) + \frac{\partial\rho'_{1,2S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,S}(\gamma_0)\right)\times\right. \\ &\quad \left.\left(\rho'_{1,2S}(\gamma_0)W_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'} + \rho'_{1,S}(\gamma_0)W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right)\right] \\ &= \frac{1}{2}E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,2S}(\gamma_0)\rho'_{1,2S}(\gamma_0)W_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'}\right] \\ &+ \frac{1}{2}E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}W_1\rho_{1,2S}(\gamma_0)\rho'_{1,S}(\gamma_0)W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right], \end{aligned}$$

$$\begin{aligned} & A_{S,21} \\ = & \frac{1}{2}E\left[(V_1\otimes(Z_1-HV_1))\left(\rho'_{1,2S}(\gamma_0)W_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'} + \rho'_{1,S}(\gamma_0)W_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right)\right] \\ = & E\left[(V_1\otimes(Z_1-HV_1))\,\rho'_1(\gamma_0)W_1\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}\right] = A_{21} \end{aligned}$$

and

$$\begin{aligned} A_{S,22} &= E\left[(V_1\otimes(Z_1-HV_1))\,(V_1\otimes(Z_1-HV_1))'\right] \\ &= E\left[V_1V'_1\otimes(Z_1-HV_1)\,(Z_1-HV_1)'\right] \\ &= A_{22}. \end{aligned}$$

It follows from (A.29) and (A.30) that

$$\frac{1}{2\sqrt{n}}\frac{\partial Q_{n,S}(\gamma_0)}{\partial\gamma} \xrightarrow{L} N(0, CA_SC'). \qquad \text{(A.31)}$$

Finally, the second part of Theorem 3.2.8 follows from (A.26), (A.27) and (A.31).

Bibliography

[1] Abarin, T. and Wang, L. (2009). Second-order least squares estimation of censored regression models. *Journal of Statistical Planning and Inference* 139, 125 – 135.

[2] Abarin, T. and Wang, L. (2008). Instrumental variable approach to covariate measurement error in generalized linear models. *Submitted.*

[3] Abarin, T. and Wang, L. (2006). Comparison of GMM with second-order least squares estimation in nonlinear models. *Far East Journal of Theoretical Statistics* 20 (2), 179 – 196.

[4] Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6, 251–262.

[5] Aitkin, M. and Rocci, R. (2002). A general maximum likelihood anal-

ysis of measurement error in generalized linear models. *Statistics and Computing*, 12, 163 – 174.

[6] Allen, Michael P. (1997). *Understanding regression analysis.* Plenum Press. New York and London.

[7] Amemiya, T. (1973). Regression analysyis when the dependent variable is truncated normal. *Econometrica* 41, 997 – 1016.

[8] Anderson, T. G. and Sorenson, B. E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 14, 328 – 352.

[9] Austin, P. C., Ghali, W.A., and Tu, J. V. (2003). A comparison of several regression models for analyzing cost of CABG surgery. *Statistics in Medicine* 22, 2799 – 2816.

[10] Bianco, A. M., Garcia Ben, M. and Johai, V. J. (2005). Robust estimation for linear regression with asymmetric errors. *Canadian Journal of Statistics* 33, 511 – 528.

[11] Buzas, J.S. and Stefanski L.A. (1996). Instrumental variable estimation in generalized linear measurement error models. *Journal of American Statistical Association*, 91, 999 – 1006.

[12] Carroll, R.J. and Ruppert, D. (1988). *Transformation and weighting in Regression*. Chapman & Hall, London.

[13] Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*, Chapman & Hall, London.

[14] Diaz-Emparanza, I. (2000). *Is a small Monte Carlo analysis a good analysis? Checking the size, power and consistency of a simulation-based test.* EconWPA in its series Econometrics 0004005.

[15] Finster, M. P. (1987). An analysis of five simulation methods for determining the number of replications in a complex monte carlo study. *Statistics and Probability Letters* 5, 353 – 360.

[16] Franks, P. W., Ekelund, U., Brage, S., Wong, M., and Wareham, N. J. (2004). Does the association of habitual physical activity with the metabolic syndrome differ by level of cardiorespiratory fitness? *Diabetes care*, 27,5, 1187 – 1193.

[17] Fuller, A. W. (1987). *Measurement error models*. Wiley, New York.

[18] Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *The Annals of Statistics* 22, 732 – 746.

[19] Graybill, F. A., Iyer, H. K. (1994). *Regression analysis: concepts and applications*. Duxbury Press.

[20] Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029 – 1054.

[21] Heckman, J. J. and MaCurdy, T. E. (1986). Labor Econometrics. In Z. Griliches and M. D. Intriligator eds., *Handbook of Econometrics III*, North-Holland, Amsterdam.

[22] Huber, P. J. (2004). *Robust Statistics*, Wiley, New York.

[23] Jarque, C. M., Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review* 55, 163 – 172.

[24] Jennrich, R. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics* 40, 633 – 643.

[25] Kiechl, S., Schett, G., Wenning, G., Redlich, K., Oberhollenzer, M., Mayr, A., Santer, P., Smolen, J., Poewe, W., and Willeit, J. (2004).

Osteoprotegerin is a risk factor for progressive atherosclerosis and cardiovascular disease. *Circulation*, 109, 2175 – 2180.

[26] Killingsworth, M. R. and Heckman, J. J. (1986). Female labor supply: A survey. In O. Ashenfelter and R. Layard eds., *Handbook of Labor Economics I*, North-Holland, Amsterdam.

[27] Koenker, R. and Machado, J. A. F. (1999). GMM inference when the number of moments conditions is large. *Journal of Econometrics* 93, 327 – 344.

[28] Kukush, A. and Schneeweiss H. (2005). Comparing different estimators in a nonlinear measurement error model. *Mathematical methods of Statistics*, 14, 1, 53 – 79.

[29] Lange, k (1999). *Numerical analysis for statisticians*. Springer- Verlag, New york.

[30] Li, H. (2005). *A Simulation study of the second-order least squares estimators for nonlinear mixed effects models*. Master's thesis, University of Manitoba.

[31] Li, T. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65, 139 – 165.

[32] Li, T. and Hsiao, C. (2004). Robust estimation of generalized linear models with measurement errors. *Journal of Econometrics*, 118, 51 – 65.

[33] Magnus, J. R. and Neudecker, H. (1988). *Matrix differential calculus with application in statistics and econometrics.* Wiley, New York.

[34] Marazzi, A. and Yohai, V. J. (2004). Adaptive truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference* 122, 271 – 291.

[35] Mátyás, L. (1999). *Generalized method of moments estimation.* Cambridge University Press.

[36] McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models.* Chapman and Hall: London.

[37] Mooney, Christopher Z. (1997). *Monte Carlo simulation.* Sage publication.

[38] Mordecai, A. (2003). *Nonlinear programming: Analysis and methods.* Dover publishing.

[39] Nakamura, T. (1990). Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika*, 77, 1, 127 – 137.

[40] Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303 – 325.

[41] Powell, J. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* 54, 1435 – 1460.

[42] Prichard, D. J., Downie, J. and Bacon, D. W. (1977). Further consideration of heteroscedasticity in fitting kinetic models. *Technometrics*, 19, 227 – 236.

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

[44] Robert, C. P. and Cassella, G. (2004). *Monte Carlo statistical methods.* Springer Texts in Statistics. 2nd edition.

[45] Ross, Sheldon M. (1997). *Simulation*, 2nd edition. Academic Press.

[46] Roy, S. and Banerjee T. (2006). A flexible model for generalized linear regression with measurement error. *AISM*, 58, 153 – 169.

[47] Schafgans, Marcia M. A. (2004). Finite sample properties for the semi-parametric estimation of the intercept of a censored regression model. *Statistica Neerlandica* 58, 35 – 56.

[48] Schafer, D. W. (2001). Semiparametric maximum likelihood for measurement error model regression. *Biometrics*, 57, 1, 53 – 61.

[49] Schennach, M. S. (2004). Estimation of nonlinear models with measurement error. *Econometrica*, 72, 1, 33 – 75.

[50] Schennach, M. S. (2007). Instrumental variable estimation of nonlinear errors-in-variables models, *Econometrica*, 75, 201 – 239.

[51] Seber, G. A. F., Wild, C. J. (1989). *Nonlinear regression.* Wiley series in probability and mathematical statistics.

[52] Stefanski L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics*, 13, 4, 1335 – 1351.

[53] Stefanski L. A. and Carroll, R. J. (1991). Deconvolution-Based score tests in measurement error models. *The Annals of Statistics*, 19, 1, 249 – 259.

[54] Stimmer, K. (2003). Modeling gene expression measurement error: a quasi-likelihood approach. *BMC Bioinformatics*, 4, 1 – 10.

[55] Stock, J. H. and Wright J. H. (2000). GMM with weak identification. *Econometrica* 68, 1055 – 1096.

[56] Taupin, M. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *The Annals of Statistics*, 29, 1, 66 – 93.

[57] Thoresen, M. and Laake, P. (2003). The use of replicates in logistic measurement error modeling. *Scandinavian Journal of Statistics*, 30, 625 – 636.

[58] Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26, 24 – 36.

[59] van der Vaart, AW, A. (2000). *Asymptotic Statistics*. Cambridge University Press.

[60] Wang, L. (1998). Estimation of censored linear errors-in-variables models. *Journal of Econometrics* 84, 383 – 400.

[61] Wang, L. (2003). Estimation of nonlinear Berkson-type measurement error models. *Statistica Sinica*13, 1201 – 1210.

[62] Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics* 32, 2559 – 2579.

[63] Wang, L. and Hsiao, C. (1995). A simulated semiparametric estimation of nonlinear errors-in-variables models. *Working Paper*, Department of Economics, University of Southern California.

[64] Wang, L. and Hsiao, C. (2007). Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Working Paper*, Department of Statistics, University of Manitoba.

[65] Wang, L. and Leblanc A. (2007). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics*, DOI: 10.1007/s10463-007-0139-z.

[66] Williams, M. (1997). A regression technique accounting for heteroscedastic and asymmetric errors. *Journal of Agricultural, Biological, and Environmental Statistics* 2, 108 – 129.

[67] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25 – 51.

[68] Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. The MIT Press.

[69] Wright, J. H. (2003). Detecting lack of identification in GMM. *Econometric Theory* 19, 322 – 330.