Poisson regression with Laplace measurement error

by

Shengnan Chen

M.S., North China University of Technology, 2016

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

In this dissertation, novel estimation procedures are proposed for a class of Poisson linear regression when the covariate is contaminated with Laplace measurement error.

This dissertation contains two research projects. In the first project, we propose a weighted least squares estimation procedure that incorporates the first two conditional moments of the response variable given the observed surrogate, and the weight function is intentionally chosen to avoid the complexity caused by the random denominator and to increase the estimation efficiency. To solve for the conditional moments, a Tweedie-type formula for the conditional expectation of the likelihood function given the observed surrogate has been adopted. Instead of assuming the distribution of the unobserved covariate is known, we assume that the distribution of that latent variable is unknown. Large sample properties of the proposed estimator, including the consistency and the asymptotic normality, are discussed. The finite sample performance of the proposed estimation procedure is evaluated by simulation studies, showing that the proposed estimator is more efficient than the existing ones.

In the second project, we propose a corrected maximum likelihood estimation procedure based upon the Tweedie-type formula. Two situations, the distribution of the latent variable is known as well as unknown, are considered. Large sample properties of the proposed estimator are discussed, and simulation study shows that the estimator is more efficient than the existing estimation procedures. Besides, further simulation studies are also conducted to compare our proposed two estimation procedures. And sensitivity analysis has been done to examine the robustness of our methods in real data.

Although the discussion is conducted for univariate cases, the proposed estimation procedure can be readily extended to the multivariate cases by using multivariate Tweedie-type formulae.

Poisson regression with Laplace measurement error

by

Shengnan Chen

M.S., North China University of Technology, 2016

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Dr. Weixing Song

Copyright

Abstract

In this dissertation, novel estimation procedures are proposed for a class of Poisson linear regression when the covariate is contaminated with Laplace measurement error.

This dissertation contains two research projects. In the first project, we propose a weighted least squares estimation procedure that incorporates the first two conditional moments of the response variable given the observed surrogate, and the weight function is intentionally chosen to avoid the complexity caused by the random denominator and to increase the estimation efficiency. To solve for the conditional moments, a Tweedie-type formula for the conditional expectation of the likelihood function given the observed surrogate has been adopted. Instead of assuming the distribution of the unobserved covariate is known, we assume that the distribution of that latent variable is unknown. Large sample properties of the proposed estimator, including the consistency and the asymptotic normality, are discussed. The finite sample performance of the proposed estimation procedure is evaluated by simulation studies, showing that the proposed estimator is more efficient than the existing ones.

In the second project, we propose a corrected maximum likelihood estimation procedure based upon the Tweedie-type formula. Two situations, the distribution of the latent variable is known as well as unknown, are considered. Large sample properties of the proposed estimator are discussed, and simulation study shows that the estimator is more efficient than the existing estimation procedures. Besides, further simulation studies are also conducted to compare our proposed two estimation procedures. And sensitivity analysis has been done to examine the robustness of our methods in real data.

Although the discussion is conducted for univariate cases, the proposed estimation procedure can be readily extended to the multivariate cases by using multivariate Tweedie-type formulae.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Throughout my Ph.D. study, I have received a great deal of support and assistance.

First and foremost I would like to express my sincere gratitude and appreciation for my esteemed major advisor Dr. Weixing Song. He admitted me into this Ph.D. program in 2016, have been teaching me since 2017, and superivising me since 2018. I would not have been able to finish my dissertation as well as Ph.D. study without his valuable guidance.

Also, I would like to thank Dr. Gyuhyeong Goh, Dr. Wei-Wen Hsu and Dr. Bala Natarajan for being willing to be on my dissertation committee. To me they are excellent teachers, wise and respectful seniors as well as genial people. I am very grateful that they allow me to go this far.

Besides, special thanks go to Bonnie Messmer, Jo Blackburn and Alisha Dean for their daily support, which makes my experience in our department a even better one.

Thank everyone in the department who has ever delivered his or her kindness to me. Thank all my fellow classmates. It's been my pleasure and honor to take classes as well as the qualify exam with you guys. I will always remember the time that we supported each other to face the challenges that came one after another. Thank K-State for having me here for an unforgettable six-year Ph.D. study. Thank Department of Statistics for offering me graduate teaching assistantships, for providing me with an opportunity to strengthen my knowledge in statistics and also a transition before I step into the industry.

Last but not the least, thank my parents and younger sister for their unconditional love and support, for offering me the opportunity of learning to be a daughter as well as an elder sister, for being my harbor as well as motivation.

Life goes on. The future is full of uncertainties as well as possibilities. Wish everyone all the best!

Chapter 1

Introduction

Regression model is a common model to explore the relationship between a response variable Y and some independent variables X. In many studies researchers are interested in the relationship

$$Y = m(X; \beta) + \epsilon,$$

where β is the unknown regression parameter and ϵ is the random error.

However, in real applications, the true independent variable X sometimes cannot be observed directly. One could only observe some surrogates for X instead. In other words, the independent variable is measured with error. The statistical models with error-contaminated variables are called measurement error models or errors-in-variables models.

A variety of methods have been proposed to estimate the parameters in the generalized linear regressions models when the measurement error follows normal distribution. However, few works have been done when the measurement error follows other distributions, such as the Laplace distribution. As a representative example of ordinary-smooth distributions, it's very common to see the Laplace errors in the real practice. The research interest in this dissertation is to propose novel estimation procedures for the parameter in Poisson linear regression model when the covariates are contaminated with the Laplace measurement error. We will first consider the weighted least square estimation procedure, then the corrected maximum likelihood estimation procedure.

In this chapter, we will introduce some background knowledge that is related with our two estimation procedures.

1.1 Measurement Error Model

In the measurement error literature, there are two kinds of measurement errors, the classical error and the Berkson error. The main difference between these two error structures lies in how X and its surrogate Z are related. The classical error model specifies $Z = X + U$, where $E[U|X] = 0$, U is independent of X, and therefore $E[Z|X] = X$. While the Berkson error model specifies $X = Z + U$ where $E[U|Z] = 0$, U is independent of Z, and as a consequence $E(X|Z) = Z$. Therefore, $\text{Var}(Z) > \text{Var}(X)$ for the classical errors and $\text{Var}(X) > \text{Var}(Z)$ for the Berkson errors. While these two types of errors are both of researcher's interest, the classical error is more common in the literature. In our research, we are also interested in the situation when the measurement error follows the classical error structure.

1.2 Tweedie's Formula and Tweedie-Type Formula

As disclosed in Efron (2011), the Tweedie's formula is named after Maurice Kenneth Tweedie and it was first discussed in Robbins (1956). Assuming that X is p-dimensional random vector, and $U \sim N_P(\mathbf{0}, \Sigma_u)$, $Z = X + U$, the Tweedie's formula could be expressed as

$$E(X|Z) = Z + \Sigma_u \frac{g'(Z)}{g(Z)},$$

where $g(\cdot)$ is the density function of Z.

However, notice that this Tweedie's formula only focuses on $E(X|Z)$ under the normal measurement error. When the measurement error U follows Laplace distribution, $U \sim ML_p(0, \Sigma_u)$, Shi and Song (2015) proposed a Tweedie-type formula

$$E[m(X)|Z] = \frac{1}{g(Z)} \left[\int m(x) f(Z-x) g(x) dx - \frac{1}{2} \sum_{j,l=1}^{p} \sigma_{jl} \int m(x) f(Z-x) \frac{\partial^2 g(x)}{\partial x_j \partial x_l} dx \right],$$

where $m(X)$ is a measurable function of X, $f(\cdot)$ is the density function of U and σ_{jl} is the (j,l)th element of Σ.

In Chapter 2 and Chapter 3, we will adopt this Tweedie-Type formula to calculate $E[m(X)|Z]$ for a given function $m(X)$ for both the proposed weighted least squares estimation and the corrected maximum likelihood estimation procedure.

1.3 Existing Estimation Procedures

There have been different kinds of bias-correction estimation procedures proposed in literature for estimating the parameters in nonlinear models with measurement error. An extensive discussion on this topic can be found in Carroll et al. (2006).

1.3.1 Regression Calibration

Suppose that the mean of Y given X can be modeled by $E(Y|X) = m_Y(X;\beta)$ for some unknown parameter β. X is not known due to the presence of measurement error and we can observe Z which is related to X. The regression calibration method is based on replacing the unobserved X by the regression of X on Z, $\mu_X(Z,\gamma)$, depending on parameters γ, which are estimated by $\hat{\gamma}$. And use $\mu_X(Z,\hat{\gamma})$ to obtain the parameter estimate using a stantard analysis. After replacement, regression calibration is dealing with the following approximate model

$$E(Y|Z) \approx m_Y(\mu_X(Z,\gamma),\beta)$$

However, notice that this is only an approximate model for the observed data. This method only produce approximately consistent estimators. In cases when $E(X|Z)$ is in a complicated form or when the regression model is highly nonlinear, simply replacing X with $E(X|Z)$ may not generate a satisfying estimate.

In Chapter 3, we are making some adjustment to the traditional regression calibration method. Instead of replacing X with $E(X|Z)$ and using the approximate model, we try to calculate $E(Y|Z)$ directly, when the measurement error U follows Laplace distribution, using the Tweedie-type formula mentioned in the previous subsection.

1.3.2 Score Function Methods

There are two kinds of score function methods, conditional score method and corrected score method.

Stefanski and Carroll (1987) discussed the conditional score method in the generalized linear normal measurement error model. Given a covariate $X = x$, Y has the density function

$$h_Y(y;\theta,x) = \exp\left\{\frac{y(\alpha+\beta^T x) - b(\alpha+\beta^T x)}{a(\phi)} + c(y,\phi)\right\},$$

which is that of a generalized linear model in canonical form in McCullagh and Nelder (1989). And $\theta^T = (\alpha, \beta^T, \phi)$ is the unknown parameters to be estimated. In Stefanski and Carroll (1987)'s paper, they introduce Δ, a complete and sufficient statistic for measurement error u, that has the form of

$$\Delta = \Delta(Y, X, \theta) = X + Y\Omega\beta,$$

where $\Omega = \bar{\Omega}/a(\phi)$ and $\bar{\Omega}$ is the covariance matrix of u. Based on the conditional distribution $Y|\Delta$, they are able to derive unbiased estimating equations for θ that are independent of u. For Poisson regression, the conditional distribution has the form

$$\text{pr}_\theta(Y = k|\Delta = \delta) = \frac{(k!)^{-1}\exp\{k(\alpha+\beta^T\delta) - \frac{1}{2}k^2\beta^T\Omega\beta\}}{\sum(j!)^{-1}\exp\{j(\alpha+\beta^T\delta) - \frac{1}{2}j^2\beta^T\Omega\alpha\}},$$

where the sum is over $j = 0, ..., \infty$. However, it could be noticed that this unbiased conditional score function has no closed form.

Front a different perspectiv, Nakamura (1990) proposed corrected score function method for the generalized linear models with normal measurement error. Let F be an open convex subset of a parameter space including β, he defines a function $l^*(\beta, X, Y)$ to be called a corrected log likelihood if

$$E^*\{l^*(\beta, X, Y)\} = l(\beta, Z, Y),$$

for any β in F, where $l(\beta, Z, Y)$ is the log likelihood function of β given Y and the observable variable Z. Meanwhile, $U^*(\beta, X, Y) = \partial l^*(\beta, X, Y)/\partial\beta$ is called a corrected score function, and the value β^* such that $U^*(\beta, X, Y) = 0$ is called a corrected estimate. For Poisson regression, $l^*(\beta, X, Y) = 0$ has the form

$$l^*(\beta) = \sum_{i=1}^{n}\{Y_i Z_i^{'}\beta - \ln Y_i! - \exp(Z_i^{'}\beta - 0.5\beta^{'}\Sigma_u\beta)\}.$$

Unfortunately, this corrected log likelihood function is not bounded in β.

To deal with the situation when the distribution of measurement error U is unknown, Guo and Li (2002) constructed a new type of consistent corrected score estimator for the parameter in Poisson regression model. The newly adjusted log-likelihood function is

$$Q_{new}(\beta) = \sum_{i=1}^{n}[Y_i Z_i\beta - \ln Y_i!] - E_X[\exp(X\beta)].$$

However, to calculate this estimator, we need to know $E_X[\exp(X\beta)]$, which often requires some knowledge of the density function of X. While in real applications, the density function of X is usually unknown.

Note that the majority of the methods above are all under the situation that measurement error follows normal distribution. While there are few literature when the measurement error follows Laplace distribution. To compare our proposed method with an existing method, we found that when the measurement error follows a Laplace distribution with zero mean and unknown variance, Hong and Tamer (2003) provided a modified method of the moment estimator to estimate the parameters in general nonlinear models. They showed that if $U \sim Laplace(0, \sigma_u^2)$, then

$$E\exp(X\beta) = E\left(1 - \frac{\sigma_u^2\beta^2}{2}\right)\exp(Z\beta).$$

And combining the results of Guo and Li (2002) and Hong and Tamer (2003), when the measurement error U follows a Laplace distribution with mean 0 and variance σ^2, there exists an estimation procedure of β by maximizing the following function

$$Q_{com}(\beta) = \sum_{i=1}^{n} \left[Y_i Z_i \beta - \log Y! - \left(1 - \frac{\sigma^2 \beta^2}{2}\right) \exp(Z_i \beta) \right].$$

However, note that the function $Q_{com}(\beta)$ is indeed unbounded in β.

1.3.3 Likelihood and Quasi-likelihood Methods

To conduct a likelihood analysis in measurement error models. One must specify a parametric model in case of X being observable, then choose the error structure, classical or Berkson. In the classical measurement error setting, we specify a model for the unobserved X given the observed covariate Z. Then the likelihood function can be constructed.

However, in some cases, the exact distribution cannot be identified for the data, therefore, the corresponding likelihood function is not available. A well known substitute for the likelihood method is the so called quasi-likelihood method, which is also known as quasi-likelihood and variance function (QVF) method. Proposed by Wedderburn (1974), quasi-likelihood estimation assumes only a mean-variance relationship rather than a specific distribution for Y_i. More specifically, instead of assuming a distributional type for Y_i, it assumes only

$$\text{Var}(Y_i) = v(\mu_i),$$

where $\mu_i = E(Y_i)$ and v is some chosen variance function. And It could be showed that the equations that determine quasi-likelihood estimates are the same as the likelihood equations for GLMs. Thus, to implement the quasi-likelihood method, we only have to know the mean and variance function of the response instead of the entire distribution. And to extend this method to measurement error setups, we would need to compute the mean and variance functions of the response given the observed covariates Z, or

$$E(Y|Z) = E\{m_Y(\cdot)|Z\}, \quad \text{Var}(Y|Z) = \sigma^2 E\{g^2(\cdot)|Z\} + \text{Var}\{m_Y(\cdot)|Z\},$$

which define a variance function model. If the functional forms of the mean and variance functions are known, then the parameters in the corresponding statistical models can be estimated by the common fitting algorithms. See Agresti (2002) and Carroll et al. (2006) for detailed discussions on the likelihood and quasi-likelihood methods.

In Chapter 3, we will derive the probability mass function of $(Y|Z)$ and the conditional log-likelihood function of $P(Y|Z)$ using the Tweedie-Type formula developed in Shi and Song (2015) and apply the maximum-likelihood estimation procedure. Large sample properties, including the consistency and the asymptotic normality of the resulting estimator, under both cases, the density function of X being known and unknown, will be thoroughly investigated.

1.3.4 Minimum Distance Estimation

For a class of nonlinear regression models with the Berkson measurement error structure, Wang (2004) proposed a minimum distance estimator for the regression parameters. The method is based on the first two conditional moments of the response variable given the observed covariate, and the estimator is the minimizer of the target function

$$Q_n(\gamma) = \sum_{i=1}^{n} \rho'(Y_i, Z_i; \gamma) W(Z_i) \rho(Y_i, Z_i; \gamma),$$

where $W(Z_i)$ is a weighting matrix which may depend on Z_i and

$$\rho(Y_i, Z_i; \gamma) = (Y_i - E(Y_i|Z_i),\ Y_i^2 - E(Y_i^2|Z_i))'.$$

This minimum distance estimation is indeed a weighted least square estimation procedure. However, note that under the Berkson error structure, the calculations of the conditional expectations $E(Y|Z)$ and $E(Y^2|Z)$ are quite straightforward. While this is not the case for the classical measurement error. In Chapter 2, for the Poisson regression model with Laplace measurement error, we will calculate these two conditional expectations based on the Tweedie-type formulae developed in Shi and Song (2015), then construct a weighted least squares estimation procedures similar to the procedures discussed in Wang (2004). Large sample properties of the weighted least squares estimator will be investigated.

1.4 Kernel Density Estimation

Kernel density estimation procedure is a very popular nonparametric smoothing technique for estimating the density function of a random variable. To be specific, suppose $X_1, X_2, \ldots, X_n$ is a sample from a univariate population X, then the commonly used Rosenblatt-Parzen kernel density estimator of density function f takes the form of

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where K is called a kernel density function, satisfying $K(x) \geq 0$, $\int K(x)dx = 1$ and is usually taken to be symmetric function such as the standard normal density function. h, a sequence of positive integers tending to 0 as the sample size n goes to infinity, is called the bandwidth. If we denote $K_h(u) = h^{-1}K(u/h)$, then the kernel density estimator can also be rewritten as $\hat{f}(x;h) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i)$. See Wand and Jones (1995) for an extensive discussions on the kernel smoothing for both the univariate and the multivariate data. The Rosenblatt-Parzen kernel density estimator will be adopted in our construction of the estimation procedure.

In Chapter 2 and Chapter 3, we will apply this density estimation procedure in estimating the density function of Z when the distribution of the true variable X is unknown.

Chapter 2

Weighted Least Squares Estimation in Poisson Regression with Laplace Measurement Error

2.1 Introduction

As a typical example of the generalized linear model, Poisson regression model is widely used for fitting the count data. The intensity function of the Poisson regression is often modeled as a linear function of the covariates. To be specific, let Y be a nonnegative integer, in Poisson regression, for a given covariate $X = x$, possibly multidimensional, Y assumes a Poisson distribution with intensity function $\lambda(x) = \exp(x^T\beta)$. Statistical inferences on Poisson regression are abundant in literature. An extensive discussion on the Poisson regression, as well as other related regression models on count data, can be found in Cameron and Trivedi (2013). However, the covariate X, sometimes a part of X, cannot be measured precisely due to potential measurement instruments imperfection or human errors. However, another covariate, denoted as Z, can be observed, which is related to X via the additive structure $Z = X + U$, where X and U are often assumed to be independent. The random variable U is called the measurement error. Another structure in measurement error literature for describing the relationship between Z and X is $X = Z + U$ with Z and U being independent, which is called Berkson error. The different natures of these two structures leads to very different theoretical developments. Throughout this chapter, we assume that Z and X follows the classical measurement error structure. Another important assumption made in the classical measurement error literature is that U is nondifferential, that is, the conditional distribution of Y given X, Z is the same as the conditional distribution of Y given X. In other words, given the true variable X, knowing the variable Z does not provide any further information on the distribution of Y. We will also adopt the nondifferential condition throughout the chapter. A comprehensive discussion on the measurement error modeling can be found in Fuller (2009), Cheng and Ness (2010), Buonaccorsi (2010) and Carroll et al. (2006).

It is well known that the naive estimation procedures, or simply replacing X with the data from Z in a standard estimation procedure often result in biased estimate. For Poisson regression, note that the mean and variance function of the response variable Y given X are all equal to $\exp(X^T\beta)$. However, the conditional expectation $E(Y|Z)$ is less than $\text{Var}(Y|Z)$. This over dispersion effect was first observed by Guo and Li (2002). The research on reducing or removing the effect of the measurement error in the estimation of β has attracted more and more attentions from both theoretical and applied statisticians. Some classical estimation procedures, such as the regression calibration, simulation extrapolation and instrumental variable methods have been applied to Poisson regression models. Examples include the conditional score estimator proposed in Stefanski and Carroll (1987), the corrected score estimator proposed in Stefanski (1989) and Nakamura (1990). However, the existing methods either have complicated algorithms, or the target functions to be maximized are not bounded, or the measurement errors to be assumed are normal. Although Guo and Li (2002) relaxed the normality assumption and constructed the exact corrected log-likelihood, but the computation of such likelihood function involves the estimation of $E[X\exp(X\beta)]$ which is not straightforward if the distribution of X is unknown.

In this chapter, we will propose a simple estimation procedure based on the first and second conditional moments of Y given the surrogate covariate Z. The proposed estimation procedure is a weighted least square methods which is similar to the minimum distance procedure discussed in Wang (2004) in the context of nonlinear regression with the Berkson measurement error. Since the conditional expectations of Y given Z are more complicated in the classical measurement error cases, the derivation of the estimation procedure, and the discussion of the statistical properties of the resulting estimators are more challenging than the minimum distance procedure discussed in Wang (2004).

To avoid notational complexity from multivariate covariate X, we will limit our discussion to the univariate cases, the estimation procedure can be readily extended to the multivariate covariates scenarios.

This chapter is organized as follows. The proposed weighted least square estimation procedures will be stated in Section 2.2. Large sample properties of the proposed estimator, including the consistency and the asymptotic normality, will be thoroughly discussed in Section 2.3. Simulation studies will be conducted in Section 2.4 and the proofs of the main results will be deferred to Section 2.5.

2.2 Weighted Least Squares Estimation Procedure

In the Berkson measurement error setup, Wang (2004) proposed a minimum distance estimation procedure for the nonlinear regression model based on the first and second conditional moments of the response variable given the surrogates. Under the nondifferential condition, we could obtain the first conditional moment of Y given Z as

$$E(Y|Z) = E[E(Y|Z, X)|Z] = E[E(Y|X)|Z] = E(\exp(X\beta_0)|Z),$$

where β_0 is the true value of β, and the second conditional moment as

$$E(Y^2|Z) = E[E(Y^2|Z, X)|Z] = E[E(Y^2|X)|Z] = E(\exp(X\beta_0)|Z) + E(\exp(2X\beta_0)|Z).$$

Unlike the Berkson measurement error, the above two conditional moments do not have a simple form even if the distribution of X is known in the classical measurement error setup.

For a twice differentiable function $m(x)$, Shi and Song (2015) showed that if the density function $g(Z)$ of Z is also twice continuously differentiable, then under the assumption of the Laplace measurement error $(0, \sigma_u^2)$, there holds the following Tweedie-type formula

$$\begin{aligned} E[m(X)|Z] = m(Z) \ &+ \frac{1}{g(Z)} \int_z^\infty \left[m'(x) - \frac{\sigma_u m''(x)}{2\sqrt{2}} \right] g(x) \exp\left(\frac{z-x}{\sigma_u/\sqrt{2}} \right) dx \\ &- \frac{1}{g(Z)} \int_{-\infty}^z \left[m'(x) + \frac{\sigma_u m''(x)}{2\sqrt{2}} \right] g(x) \exp\left(\frac{x-z}{\sigma_u/\sqrt{2}} \right) dx \end{aligned}$$

In the Poisson regression with a univariate X, the regression function $E(Y|X) = \exp(X\beta)$. Let $m(X) = \exp(X\beta)$, then according to the above Tweedie-type formula, we have

$$\begin{aligned} E[\exp(X\beta)|Z] = \exp(Z\beta) \ &+ \frac{1}{g(Z)} \left(\beta - \frac{\sigma_u}{2\sqrt{2}} \beta^2 \right) \exp\left(\frac{\sqrt{2}Z}{\sigma_u} \right) \int_Z^\infty g(x) \exp[(\beta - \sqrt{2}/\sigma_u)x] dx \\ &- \frac{1}{g(Z)} \left(\beta + \frac{\sigma_u}{2\sqrt{2}} \beta^2 \right) \exp\left(-\frac{\sqrt{2}Z}{\sigma_u} \right) \int_{-\infty}^Z g(x) \exp[(\beta + \sqrt{2}/\sigma_u)x] dx. \end{aligned}$$

Similarly, let $m(X) = \exp(2X\beta)$, we have

$$\begin{aligned} E[\exp(2X\beta)|Z] = \exp(2Z\beta) \ &+ \frac{1}{g(Z)} \left(2\beta - \frac{2\sigma_u}{\sqrt{2}} \beta^2 \right) \exp\left(\frac{\sqrt{2}Z}{\sigma_u} \right) \int_Z^\infty g(x) \exp[(2\beta - \sqrt{2}/\sigma_u)x] dx \\ &- \frac{1}{g(Z)} \left(2\beta + \frac{2\sigma_u}{\sqrt{2}} \beta^2 \right) \exp\left(-\frac{\sqrt{2}Z}{\sigma_u} \right) \int_{-\infty}^Z g(x) \exp[(2\beta + \sqrt{2}/\sigma_u)x] dx. \end{aligned}$$

If the density function $g(z)$ of Z is known, then similar to the minimum distance estimator proposed in Wang (2004), β can be estimated by $\tilde{\beta}_n = \arg\min \tilde{L}_n(\beta)$, where the target function $\tilde{L}_n(\beta)$ has the following form

$$\frac{1}{n} \sum_{i=1}^n [(Y_i - E(\exp(X_i\beta)|Z_i))^2 + (Y_i^2 - E(\exp(X_i\beta)|Z_i) - E(\exp(2X_i\beta)|Z_i))^2] W(Z_i),$$

where $W(Z_i)$ is the weight function which is nonnegative and chosen to improve the efficiency of the estimator.

To proceed, we denote

$$
\begin{aligned}
T(Z,\beta) = &\left(\beta - \frac{\sigma_u}{2\sqrt{2}}\beta^2\right)\exp\left(\frac{\sqrt{2}Z}{\sigma_u}\right)\int_Z^\infty g(x)\exp[(\beta-\sqrt{2}/\sigma_u)x]dx \\
&-\left(\beta + \frac{\sigma_u}{2\sqrt{2}}\beta^2\right)\exp\left(-\frac{\sqrt{2}Z}{\sigma_u}\right)\int_{-\infty}^Z g(x)\exp[(\beta+\sqrt{2}/\sigma_u)x]dx.
\end{aligned}
$$

Then the target function $L_n(\beta)$ can be written as

$$
\begin{aligned}
\tilde{L}_n(\beta) = &\frac{1}{n}\sum_{i=1}^n \left\{\left[Y_i - \exp(Z_i\beta) - \frac{T(Z_i,\beta)}{g(Z_i)}\right]^2\right. \\
&\left. + \left[Y_i^2 - \exp(2Z_i\beta) - \frac{T(Z_i,2\beta)}{g(Z_i)} - \exp(Z_i,\beta) - \frac{T(Z_i,\beta)}{g(Z_i)}\right]^2\right\} W(Z_i). \qquad (2.1)
\end{aligned}
$$

Note that in practice the density function g is rarely known, so the function $L_n(\beta)$ cannot be used directly. In this case, replacing the unknown g function with some nonparametric estimators, for example, the Rosenblatt-Parzen kernel density estimator, becomes a natural choice. However, such replacement will lead to a random denominator in $L_n(\beta)$ which makes the large sample theory development more difficult. To avoid this potential technical challenge, we can modify the function $L_n(\beta)$ by choosing the weight function to be the form of $g^2(z)W(z)$. As a result, $\tilde{L}_n(\beta)$ in (2.1) becomes

$$
\begin{aligned}
L_n(\beta) = &\frac{1}{n}\sum_{i=1}^n\{[Y_i g(Z_i) - \exp(Z_i\beta)g(Z_i) - T(Z_i,\beta)]^2 \\
&+ [Y_i^2 g(Z_i) - \exp(2Z_i\beta)g(Z_i) - T(Z_i,2\beta) - \exp(Z_i\beta)g(Z_i) - T(Z_i,\beta)]^2\}W(Z_i).
\end{aligned}
$$

For each $i = 1, 2, \ldots, n$, suppose Z is independent of Z_i and has the density function $g(z)$, then

$$
\int_{Z_i}^\infty g(x)\exp[(\beta-\sqrt{2}/\sigma_u)x]dx = E[\exp[(\beta-\sqrt{2}/\sigma_u)Z]I_{[Z_i,\infty)}(Z)|Z_i],
$$

$$\int_{-\infty}^{Z_i} g(x)[(\beta+\sqrt{2}/\sigma_u)x]dx = E[\exp[(\beta+\sqrt{2}/\sigma_u)Z]I_{(-\infty,Z_i)}(Z)|Z_i],$$

which indicates that the two integrals can be estimated by the empirical sample analogues

$$\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]} \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^{n}\exp[(\beta+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j<Z_i]}.$$

Thus, $T(Z_i,\beta)$ can be estimated by

$$\begin{aligned}\hat{T}(Z_i,\beta) = &\left(\beta-\frac{\sigma_u}{2\sqrt{2}}\beta^2\right)\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right)\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]} \\ &-\left(\beta+\frac{\sigma_u}{2\sqrt{2}}\beta^2\right)\exp\left(-\frac{\sqrt{2}Z_i}{\sigma_u}\right)\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j<Z_i]}.\end{aligned}$$

Therefore, in the cases of unknown $g(z)$, the proposed estimator of β is defined as

$$\hat{\beta}_n = \text{argmin}_\beta \hat{L}_n(\beta), \tag{2.2}$$

where

$$\begin{aligned}\hat{L}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\{&[Y_i\hat{g}(Z_i)-\exp(Z_i\beta)\hat{g}(Z_i)-\hat{T}(Z_i,\beta)]^2 \\ &+[Y_i^2\hat{g}(Z_i)-\exp(2Z_i\beta)\hat{g}(Z_i)-\hat{T}(Z_i,2\beta)-\exp(Z_i\beta)\hat{g}(Z_i)-\hat{T}(Z_i,\beta)]^2\}W(Z_i),\end{aligned} \tag{2.3}$$

where $\hat{g}(z)$ is the classic kernel density estimator $\hat{g}(z) = (nh)^{-1}\sum_{i=1}^{n} K\left((z-Z_i)/h\right)$ with K being a kernel density function and h being a bandwidth which is a sequence of positive integers indexed by the sample size n.

2.3 Large Sample Results

In this section, large sample properties of the estimator defined in (2.2), including the consistency and the asymptotic normality will be discussed. Recall that $\hat{\beta}_n$ is the minimizer of

$\hat{L}_n(\beta)$ defined in (2.3), so $\hat{\beta}_n$ is the solution of the equation $\dot{\hat{L}}_n(\beta) = 0$, where $\dot{\hat{L}}_n(\beta)$ is the derivative of $\hat{L}_n(\beta)$ with respect to β, which has the following form

$$
\begin{aligned}
\dot{\hat{L}}_n(\beta) = \frac{2}{n}\sum_{i=1}^{n}\{&[Y_i\hat{g}(Z_i) - \exp(Z_i\beta)\hat{g}(Z_i) - \widehat{T}(Z_i,\beta)][-Z_i\exp(Z_i\beta)\hat{g}(Z_i) - \dot{\widehat{T}}(Z_i,\beta)] \\
&+ [Y_i^2\hat{g}(Z_i) - \exp(Z_i\beta)\hat{g}(Z_i) - \exp(2Z_i\beta)\hat{g}(Z_i) - \widehat{T}(Z_i,\beta) - \widehat{T}(Z_i,2\beta)]\cdot \\
&[-Z_i\exp(Z_i\beta)\hat{g}(Z_i) - 2Z_i\exp(2Z_i\beta)\hat{g}(Z_i) - \dot{\widehat{T}}(Z_i,\beta) - \dot{\widehat{T}}(Z_i,2\beta)]\}W(Z_i),
\end{aligned}
$$

where

$$
\begin{aligned}
\dot{\widehat{T}}(Z_i,\beta) = &\left(1 - \frac{\sigma_u}{\sqrt{2}}\beta\right)\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right)\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta - \sqrt{2}/\sigma_u)Z_j]I_{[Z_j\ge Z_i]} \\
&+ \left(\beta - \frac{\sigma_u}{2\sqrt{2}}\beta^2\right)\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right)\frac{1}{n}\sum_{j=1}^{n}Z_j\exp[(\beta - \sqrt{2}/\sigma_u)Z_j]I_{[Z_j\ge Z_i]} \\
&- \left(1 + \frac{\sigma_u}{\sqrt{2}}\beta\right)\exp\left(-\frac{\sqrt{2}Z_i}{\sigma_u}\right)\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta + \sqrt{2}/\sigma_u)Z_j]I_{[Z_j< Z_i]} \\
&- \left(\beta + \frac{\sigma_u}{2\sqrt{2}}\beta^2\right)\exp\left(-\frac{\sqrt{2}Z_i}{\sigma_u}\right)\frac{1}{n}\sum_{j=1}^{n}Z_j\exp[(\beta + \sqrt{2}/\sigma_u)Z_j]I_{[Z_j< Z_i]}.
\end{aligned}
$$

The validity of the large sample results for the estimator $\hat{\beta}_n$ relies on the following technical assumptions.

(C1). The parameter space of β is a closed interval $\Theta = [\underline{\beta}, \bar{\beta}]$ in $\mathbb{R}$, and the true value β_0 is an interior point of Θ.

(C2). The density function $g(z)$ of Z is twice continuously differentiable.

(C3). The latent variable X satisfies $E\exp[(\beta^* + \sqrt{2}/\sigma_u)X] < \infty$, and $E\exp[-(\beta_* + \sqrt{2}/\sigma_u)X] < \infty$.

(C4). $EY(|Z|+1)\exp(Z\beta)W(Z)$ and $EYW(Z)\exp(\sqrt{2}Z/\sigma_u)$ are finite for all $\beta \in \Theta$.

(C5). Ee_{11}^2, Ee_{21}^2, Ee_{31}^2 and Ee_{41}^2 are all finite, where e_{11}, e_{21}, e_{31} and e_{41} are defined in (2.8), (2.9),(2.12), and (2.13), respectively.

The following theorems summarize the consistency and the asymptotic normality of $\hat{\beta}_n$.

Theorem 1. *Under conditions (C1)-(C4), $\hat{\beta}_n \to \beta_0$ in probability as $n \to \infty$.*

Theorem 2. *In addition to the conditions in Theorem 1, we further assume that (C5) holds, then $\sqrt{n}(\hat{\beta}_n - \beta_0) \Longrightarrow N(0, \sigma^2)$ as $n \to \infty$, where $\sigma^2 = E(e_{11} + e_{21} + e_{31} + e_{41})^2/\ddot{L}^2(\beta_0)$, $\ddot{L}(\beta) = \ddot{L}_1(\beta) + \ddot{L}_2(\beta)$ are defined in (2.4) and (2.5), respectively, and e_{11}, e_{21}, e_{31} and e_{41} are defined in (2.8), (2.9),(2.12), and (2.13), respectively.*

Although the asymptotic variance σ^2 has a complicated form, it can be well explained by four different sources of variabilities. In specific, e_{11} is the irreducible deviations of Y and Y^2 from $E(Y|Z)$ and $E(Y^2|Z)$, respectively, e_{21} reflects the variability from the estimation of the density function of Z, e_{31} represents the variability from the estimation of $T(Z, \beta_0)$, and e_{41} is caused by the estimation of both $T(Z, \beta_0)$ and $T(Z, 2\beta_0)$.

2.4 Simulation Studies

In this section, we conduct a simulation study to evaluate the performance of the proposed estimation procedure. The response variable Y is generated from a Poisson distribution with mean $\exp(X\beta)$, where $X \sim N(0, 1)$. The true value of β is chosen to be 1. We further contaminate the X variable with an additive independent Laplace measurement error U with mean 0 and variance σ_u^2, that is, $Z = X + U$. To see the effects of sample sizes and σ_u^2 on the estimates, we choose $n = 200, 300, 500$ and 1000, and $\sigma_u^2 = 0.5, 0.25$ and 0.1. For comparison purpose, the simulation is repeated 500 times for each setup. Then the means, variances, biases, and mean squares errors are computed, as well as the boxplots from 500 estimates.

The kernel density estimator is used to estimate the density function of Z, and bandwidth is chosen to be $h = an^{-1/5}$, where a is a positive constant to control the smoothness. By doing so, we can evaluate the influence of the bandwidth on the performance of the proposed estimator. In the simulation study, we choose $a = 0.8, 1$ and 1.2. The weight function $W(z) = 1$ is used in the simulation studies for the sake of simplicity.

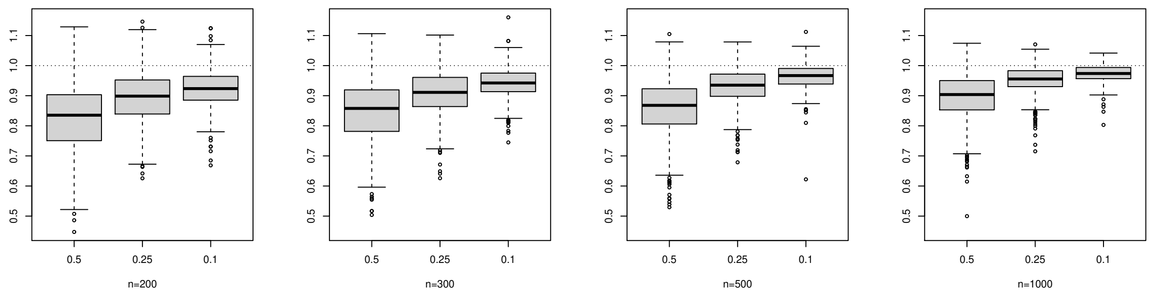

Figure 2.1: *Boxplots for* $h = 0.8n^{-1/5}$

Figure 2.1 are the boxplots of the 500 estimates of β for $h = 0.8n^{-1/5}$. It clearly shows that the estimates become better when the sample size gets bigger and the variance of the measurement error gets smaller. It is also noted that the estimated values tend to be smaller than the true values of the parameter. These observations are further confirmed by the biases and MSEs reported in Table 2.1.

	n=200			n=300			n=500			n=1000		
σ_u^2	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1
Mean	0.823	0.896	0.925	0.847	0.910	0.943	0.859	0.931	0.964	0.893	0.952	0.974
Bias	-0.177	-0.104	-0.075	-0.153	-0.090	-0.057	-0.141	-0.069	-0.036	-0.107	-0.048	-0.026
Variance	0.013	0.007	0.004	0.011	0.006	0.002	0.009	0.003	0.002	0.006	0.002	0.001
MSE	0.045	0.018	0.009	0.034	0.014	0.006	0.029	0.008	0.003	0.018	0.004	0.002

Table 2.1: *Mean, Bias, Variance and MSE of* $\hat{\beta}$ *when* $h = 0.8n^{-1/5}$

The simulation results for $h = n^{-1/5}$ and $1.2n^{-1/5}$, summarized in Figure 2.2, 2.3, Table 2.2 and 2.3, are similar to those reported in Figure 2.1 and Table 2.1.

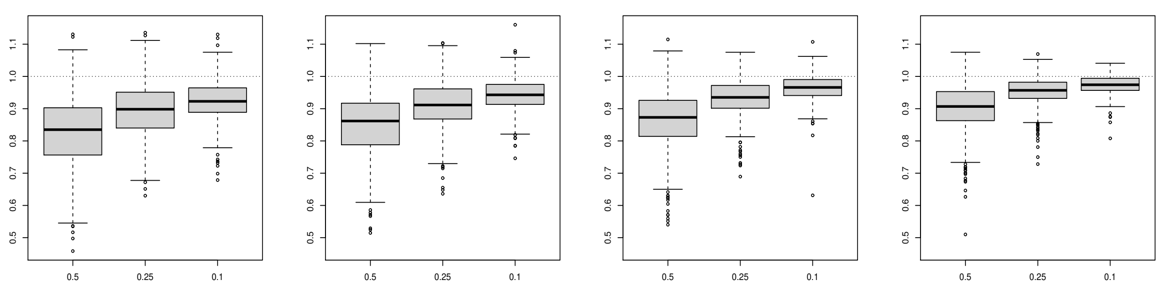

Figure 2.2: *Boxplots for* $h = n^{-1/5}$

	n=200			n=300			n=500			n=1000		
σ_u^2	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1
Mean	0.826	0.896	0.925	0.851	0.911	0.943	0.864	0.933	0.964	0.898	0.954	0.974
Bias	-0.174	-0.104	-0.075	-0.149	-0.089	-0.057	-0.136	-0.067	-0.036	-0.102	-0.046	-0.026
Variance	0.013	0.007	0.004	0.010	0.005	0.002	0.008	0.003	0.002	0.006	0.002	0.001
MSE	0.043	0.018	0.009	0.032	0.013	0.006	0.027	0.008	0.003	0.016	0.004	0.001

Table 2.2: *Mean, Bias, Variance and MSE of* $\hat{\beta}$ *when* $h = n^{-1/5}$

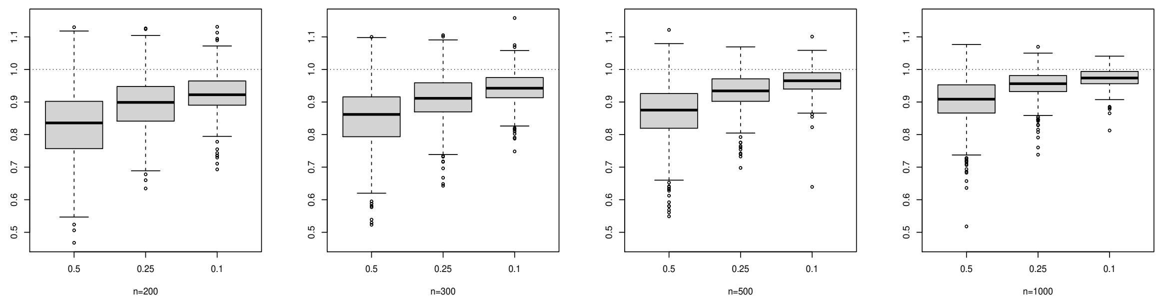

Figure 2.3: *Boxplots for* $h = 1.2n^{-1/5}$

	n=200			n=300			n=500			n=1000		
σ_u^2	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1
Mean	0.827	0.896	0.926	0.853	0.911	0.942	0.867	0.933	0.964	0.901	0.954	0.974
Bias	-0.173	-0.104	-0.074	-0.147	-0.089	-0.058	-0.133	-0.067	-0.036	-0.099	-0.046	-0.026
Variance	0.012	0.007	0.004	0.010	0.005	0.002	0.008	0.003	0.002	0.005	0.002	0.001
MSE	0.042	0.017	0.009	0.031	0.013	0.006	0.026	0.007	0.003	0.015	0.004	0.001

Table 2.3: *Mean, Bias, Variance and MSE of* $\hat{\beta}$ *when* $h = 1.2n^{-1/5}$

2.5 Appendix: Proofs of Main Results

Denote $L_1(\beta) = E\{[Y - E(\exp(X\beta)|Z)]^2 g^2(Z)W(Z)\}$, and $L_2(\beta) = E\{[Y^2 - E(\exp(X\beta)|Z) - E(\exp(2X\beta)|Z)]^2 g^2(Z)W(Z)\}$, we have $L(\beta) = L_1(\beta) + L_2(\beta)$. Then $\dot{L}(\beta)$, the derivative of $L(\beta)$, can be expressed as

$$
\begin{aligned}
\dot{L}(\beta) = \frac{\partial}{\partial \beta} \Bigg\{ E \Bigg[\left(Y - \exp(Z\beta) - \frac{T(Z,\beta)}{g(Z)} \right)^2 \\
+ \left(Y^2 - \exp(2Z\beta) - \frac{T(Z,2\beta)}{g(Z)} - \exp(Z\beta) - \frac{T(Z,\beta)}{g(Z)} \right)^2 \Bigg] g^2(Z)W(Z) \Bigg\} \\
= 2E\{[Yg(Z) - \exp(Z\beta)g(Z) - T(Z,\beta)][-Z\exp(Z\beta)g(Z) - \dot{T}(Z,\beta)] \\
+ [Y^2 g(Z) - \exp(Z\beta)g(Z) - \exp(2Z\beta)g(Z) - T(Z,\beta) - T(Z,2\beta)] \cdot
\end{aligned}
$$

$$[-Z\exp(Z\beta)g(Z) - 2Z\exp(2Z\beta)g(Z) - \dot{T}(Z,\beta) - \dot{T}(Z,2\beta)]\}W(Z).$$

Before we prove the consistency of $\hat{\beta}_n$, we shall show that $L(\beta)$ has a unique minimizer β_0. Recall that β_0 is the true value of β.

By taking the second derivative of $L_1(\beta)$ and $L_2(\beta)$ with respect to β, we see that

$$\ddot{L}_1(\beta) = 2E[E(\exp(X\beta)X^2|Z)g^2(Z)W(Z)], \tag{2.4}$$

$$\ddot{L}_2(\beta) = 2E[E(\exp(X\beta)X^2 + 4\exp(2X\beta)X^2|Z)g^2(Z)W(Z)]. \tag{2.5}$$

It is easy to see that they are strictly bigger than 0. Thus, both $L_1(\beta)$ and $L_2(\beta)$ are strictly convex in β. This implies that $L(\beta)$ is also strictly convex in β. Therefore, the minimizer of $L(\beta)$, if exists, must be unique. Now, let us show the minimizer of $L_1(\beta)$ and $L_2(\beta)$ is the true parameter value β_0. To see this, note that

$$\begin{aligned}
\dot{L}_1(\beta) = & -2E\{[Y - E(\exp(X\beta)|Z)]E(X\exp(X\beta)|Z)g^2(Z)W(Z)\} \\
= & -2E\left[E\{[Y - E(\exp(X\beta)|Z)]E(X\exp(X\beta)|Z)g^2(Z)W(Z)|Z\}\right] \\
= & -2E\left[\{E(Y|Z) - E(\exp(X\beta)|Z)\}E(X\exp(X\beta)|Z)g^2(Z)W(Z)\right] \\
= & -2E\left[\{E(\exp(X\beta_0)|Z) - E(\exp(X\beta)|Z)\}E(X\exp(X\beta)|Z)g^2(Z)W(Z)\right].
\end{aligned}$$

So, $\beta = \beta_0$ is a solution of $\dot{L}_1(\beta) = 0$. We also have

$$\begin{aligned}
\dot{L}_2(\beta) = & -2E\{[Y^2 - E(\exp(X\beta)|Z) - E(\exp(2X\beta)|Z)] \\
& [E(X\exp(X\beta)|Z) + 2E(X\exp(2X\beta)|Z)]g^2(Z)W(Z)\} \\
= & -2E\{[E(Y^2|Z) - E(\exp(X\beta)|Z) - E(\exp(2X\beta)|Z)] \\
& [E(X\exp(X\beta)|Z) + 2E(X\exp(2X\beta)|Z)]g^2(Z)W(Z)\} \\
= & -2E\{[E(\exp(X\beta_0)|Z) + E(\exp(2X\beta_0)|Z) - E(\exp(X\beta)|Z) - E(\exp(2X\beta)|Z)] \\
& [E(X\exp(X\beta)|Z) + 2E(X\exp(2X\beta)|Z)]g^2(Z)W(Z)\}.
\end{aligned}$$

Clearly, $\beta = \beta_0$ is a solution of $\dot{L}_2(\beta) = 0$. Thus, $\beta = \beta_0$ is the solution of $\dot{L}(\beta) = 0$, this implies, together with the previous discussion on the uniqueness of the minimizer, that β_0 is the unique minimizer of $L(\beta)$.

To show the consistency of $\hat{\beta}_n$, we need the following lemmas.

Lemma 3. *If the kernel function K has integrable characteristic function, $h \to 0$, $nh^2 \to \infty$, $g(Z)$ is uniformly continuous, then $\sup_z |\hat{g}(z) - g(z)| = o_p(1)$; If we further assume that the density function $g(z)$ is twice differentiable, and its second derivative is bounded, then by choosing the kernel function K to be supported on $[-1, 1]$, $h = n^{-1/5}(\log n)^{1/6}$, we have*

$$\sup_z |\hat{g}(z) - g(z)| = O(n^{-2/5}(\log n)^{1/3}), \quad a.s.$$

as $n \to \infty$.

The first statement in Lemma 3 is a well known result in kernel density estimation literature, while the second statement is from Chen (1983).

Lemma 4. *Let $\{\xi_i\}$ be random variables with finite second moment, $\{(\xi_i, X_i)\}, i = 1, \ldots, n$ are i.i.d.. Then the following iterative logarithm law holds,*

$$\sup_{x \in \mathbb{R}} |\frac{1}{n} \sum_{i=1}^{n} \xi_i I_{[X_i \leq x]} - E\xi_1 I_{[X_1 \leq x]}| = O\left(\sqrt{\frac{\log \log n}{n}}\right), \quad a.s.$$

Proof. First, assume ξ to be nonnegative. Let $\mu = E\xi$, assumed to be positive and finite. Note that $\mu = 0$ would mean $\xi = 0$, a.s. and the LIL is trivially true. Let $G(x) = E(\xi I(X \leq x))$. Clearly, $0 < G(\infty) = \mu < \infty$.

Let

$$G_n(x) := n^{-1} \sum_{i=1}^{n} \xi_i I(X_i \leq x).$$

By the LIL for iid r.v.'s., for each $-\infty \leq x \leq \infty$,

$$|G_n(x) - G(x)| = O(\sqrt{(\log \log(n)/n)}), \quad \text{a.s.}$$

To obtain this uniformly in x, take a partition of the interval $[-\infty, \infty]$ $x_0 = -\infty < x_1 < \cdots < x_K < x_{K+1} = \infty$ such that

$$\max_{1 \leq j \leq K+1} [G(x_j) - G(x_{j-1})] \leq \sqrt{n \log\log(n)}.$$

This can be always possible since G is non decreasing right continuous and $0 < G(\infty) = \mu < \infty$. Now

$$\sup_{-\infty \leq x \leq \infty} |G_n(x) - G(x)| = \max_{1 \leq j \leq K+1} \sup_{x_{j-1} < x \leq x_j} |G_n(x) - G(x)|$$

Now use the monotonicity of G_n and G to obtain for $x_{j-1} < x \leq x_j$,

$$G_n(x_{j-1}) - G(x_{j-1}) + G(x_{j-1}) - G(x_j) \leq G_n(x) - G(x) \leq G_n(x_j) - G(x_j) + G(x_j) - G(x_{j-1}).$$

Hence

$$\begin{aligned} \sup_{-\infty \leq x \leq \infty} |G_n(x) - G(x)| &\leq 2 \max_{1 \leq j \leq K+1} |G_n(x_j) - G(x_j)| + max_{1 \leq j \leq K+1} |G(x_j) - G(x_{j-1})| \\ &= O(\sqrt{n \log\log(n)}), \quad \text{a.s.} \end{aligned}$$

To conclude the proof, for general ξ, we can write $\xi_i = \xi_i^+ - \xi_i^-$ and apply the triangle inequality. □

Now let us prove the consistency of $\hat{\beta}_n$.

Proof of Theorem 1. To show the consistency of $\hat{\beta}_n$, it suffices to show that

$$\sup_{\beta \in \Theta} |\dot{\hat{L}}_n(\beta) - \dot{L}(\beta)| = o_p(1). \tag{2.6}$$

In fact, if (2.6) is true, then $\dot{\hat{L}}_n(\hat{\beta}_n) - \dot{L}(\hat{\beta}_n) = o_p(1)$, which further implies $\dot{L}(\hat{\beta}_n) = o_p(1)$. Since β_0 is the unique solution of $\dot{L}(\beta) = 0$, we could conclude that $\hat{\beta}_n \to \beta_0$ in probability.

Note that $\dot{\hat{L}}_n(\beta) - \dot{L}(\beta) = \dot{\hat{L}}_n(\beta) - \dot{L}_n(\beta) + \dot{L}_n(\beta) - \dot{L}(\beta)$. So to show (2.6), it suffices to verify that

$$\sup_{\beta\in\Theta}|\dot{\hat{L}}_n(\beta)-\dot{L}_n(\beta)|=o_p(1),\quad \sup_{\beta\in\Theta}|\dot{L}_n(\beta)-\dot{L}(\beta)|=o_p(1). \tag{2.7}$$

By assuming the kernel function K and the density function of Z satisfy the conditions stated in Lemma 3, we could have $\sup_{1\le i\le n}|\hat{g}(Z_i)-g(Z_i)|=o_p(1)$. Also, if we let

$$\xi_j=\exp\left[\left(\beta-\frac{\sqrt{2}}{\sigma_u}\right)Z_j\right]\quad \text{or}\quad \exp\left[\left(\beta+\frac{\sqrt{2}}{\sigma_u}\right)Z_j\right],$$

then by Lemma 4, for each $\beta\in\Theta$, we have

$$\sup_z\left|\frac{1}{n}\sum_{j=1}^n\xi_j I_{[Z_j\le z]}-E\xi_1 I_{[Z_1\le z]}\right|=o_p(1).$$

In fact, the $o_p(1)$ is uniform for $\beta\in\Theta$, which is guaranteed by assuming that

$$E\exp\left[\left(\beta^*+\frac{\sqrt{2}}{\sigma_u}\right)Z\right]+E\exp\left[-\left(\beta_*+\frac{\sqrt{2}}{\sigma_u}\right)Z\right]<\infty,$$

where $\beta^*=\max(|\bar{\beta}|,|\underline{\beta}|)$, $\beta_*=\min(|\bar{\beta}|,|\underline{\beta}|)$, and Theorem 16(a) in Ferguson (2017).

Now let's show $\sup_{\beta\in\Theta}|\dot{\hat{L}}_n(\beta)-\dot{L}_n(\beta)|=o_p(1)$. Note that

$$\begin{aligned}
\dot{L}_n(\beta) &= \frac{2}{n}\sum_{i=1}^n\{[Y_ig(Z_i)-\exp(Z_i\beta)g(Z_i)-T(Z_i,\beta)][-Z_i\exp(Z_i\beta)g(Z_i)-\dot{T}(Z_i,\beta)]\\
&\quad +[Y_i^2g(Z_i)-\exp(Z_i\beta)g(Z_i)-\exp(2Z_i\beta)g(Z_i)-T(Z_i,\beta)-T(Z_i,2\beta)]\cdot\\
&\quad [-Z_i\exp(Z_i\beta)g(Z_i)-2Z_i\exp(2Z_i\beta)g(Z_i)-\dot{T}(Z_i,\beta)-\dot{T}(Z_i,2\beta)]\}W(Z_i)\\
&= -\frac{2}{n}\sum_{i=1}^n Y_iZ_i\exp(Z_i\beta)g^2(Z_i)W(Z_i)-\frac{2}{n}\sum_{i=1}^n Y_ig(Z_i)\dot{T}(Z_i,\beta)W(Z_i)\\
&\quad +\frac{2}{n}\sum_{i=1}^n Z_i\exp(2Z_i\beta)g^2(Z_i)W(Z_i)+\frac{2}{n}\sum_{i=1}^n\exp(Z_i\beta)g(Z_i)\dot{T}(Z_i,\beta)W(Z_i)\\
&\quad +\frac{2}{n}\sum_{i=1}^n Z_i\exp(Z_i\beta)T(Z_i,\beta)g(Z_i)W(Z_i)+\frac{2}{n}\sum_{i=1}^n T(Z_i,\beta)\dot{T}(Z_i,\beta)W(Z_i)\\
&\quad -\frac{2}{n}\sum_{i=1}^n 2Y_i^2Z_i\exp(2Z_i\beta)g^2(Z_i)W(Z_i)-\frac{2}{n}\sum_{i=1}^n Y_i^2g(Z_i)\dot{T}(Z_i,2\beta)W(Z_i)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{2}{n}\sum_{i=1}^{n} Y_i^2 Z_i \exp(Z_i\beta) g^2(Z_i) W(Z_i) - \frac{2}{n}\sum_{i=1}^{n} Y_i^2 g(Z_i)\dot{T}(Z_i,\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} 2Z_i \exp(4Z_i\beta) g^2(Z_i) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} \exp(2Z_i\beta) g(Z_i)\dot{T}(Z_i,2\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} Z_i \exp(3Z_i\beta) g^2(Z_i) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} \exp(2Z_i\beta) g(Z_i)\dot{T}(Z_i,\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} 2Z_i \exp(2Z_i\beta) g(Z_i) T(Z_i,2\beta) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} T(Z_i,2\beta)\dot{T}(Z_i,2\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} Z_i \exp(Z_i\beta) g(Z_i) T(Z_i,2\beta) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} T(Z_i,2\beta)\dot{T}(Z_i,\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} 2Z_i \exp(3Z_i\beta) g^2(Z_i) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} \exp(Z_i\beta) g(Z_i)\dot{T}(Z_i,2\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} Z_i \exp(2Z_i\beta) g^2(Z_i) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} \exp(Z_i\beta) g(Z_i)\dot{T}(Z_i,\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} 2Z_i \exp(2Z_i\beta) g(Z_i) T(Z_i,\beta) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} T(Z_i,\beta)\dot{T}(Z_i,2\beta)W(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} Z_i \exp(Z_i\beta) g(Z_i) T(Z_i,\beta) W(Z_i) + \frac{2}{n}\sum_{i=1}^{n} T(Z_i,\beta)\dot{T}(Z_i,\beta)W(Z_i).
\end{aligned}
$$

$\dot{\hat{L}}_n(\beta)$ has a similar expression by simply replacing g, T by $\hat{g}$, $\hat{T}$.

Assuming $E[Y\exp(Z\beta)W(Z)|Z|] < \infty$ and $E[Y\exp(Z\beta)W(Z)g(Z)|Z|] < \infty$, the absolute difference between the first term of $\dot{\hat{L}}_n(\beta)$ and the first term of $\dot{L}_n(\beta)$ could be expressed as

$$
\begin{aligned}
&\left|\frac{2}{n}\sum_{i=1}^{n} Y_i Z_i \exp(Z_i\beta) g^2(Z_i) W(Z_i) - \frac{2}{n}\sum_{i=1}^{n} Y_i Z_i \exp(Z_i\beta)\hat{g}^2(Z_i)W(Z_i)\right|\\
=&\left|\frac{2}{n}\sum_{i=1}^{n} Y_i Z_i \exp(Z_i\beta)\hat{g}^2(Z_i) W(Z_i) - \frac{2}{n}\sum_{i=1}^{n} Y_i Z_i \exp(Z_i\beta) g^2(Z_i)W(Z_i)\right|\\
=&\left|\frac{2}{n}\sum_{i=1}^{n} Y_i Z_i \exp(Z_i\beta) W(Z_i)[\hat{g}^2(Z_i) - g^2(Z_i)]\right|\\
=&\left|\frac{2}{n}\sum_{i=1}^{n} Y_i Z_i \exp(Z_i\beta) W(Z_i)[(\hat{g}(Z_i) - g(Z_i))^2 + 2(\hat{g}(Z_i) - g(Z_i))g(Z_i)]\right|
\end{aligned}
$$

$$\leq \sup |\hat{g}(Z) - g(Z)|^2 \frac{2}{n} \sum_{i=1}^{n} Y_i \exp(Z_i\beta) W(Z_i)|Z_i|$$
$$+2 \sup |\hat{g}(Z) - g(Z)| \frac{2}{n} \sum_{i=1}^{n} Y_i |Z_i| \exp(Z_i\beta) W(Z_i) g(Z_i) = o_p(1).$$

Next, note that the second term of $\dot{\hat{L}}_n(\beta)$ is

$$-\frac{2}{n}\sum_{i=1}^{n} Y_i \hat{g}(Z_i)\dot{\hat{T}}(Z_i,\beta)W(Z_i) = -\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i) + g(Z_i)]W(Z_i)\cdot$$
$$\left\{\frac{1}{n}(1-\sigma_u\beta/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)\cdot\right.$$
$$\sum_{j=1}^{n}[((\beta-\sqrt{2}/\sigma_u)Z_j)I_{[Z_j\geq Z_i]} - E\exp((\beta-\sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]} + E\exp((\beta-\sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}]$$
$$+\frac{1}{n}(\beta-\sigma_u\beta^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)\cdot$$
$$\sum_{j=1}^{n}[Z_j\exp((\beta-\sqrt{2}/\sigma_u)Z_j)I_{[Z_j\geq Z_i]} - EZ\exp((\beta-\sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}$$
$$+EZ\exp((\beta-\sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}] - \frac{1}{n}(1+\sigma_u\beta/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)\cdot$$
$$\sum_{j=1}^{n}[\exp((\beta+\sqrt{2}/\sigma_u)Z_j)I_{[Z_j< Z_i]} - E\exp((\beta+\sqrt{2}/\sigma_u)Z)I_{[Z< Z_i]} + E\exp((\beta+\sqrt{2}/\sigma_u)Z)I_{[Z< Z_i]}]$$
$$-\frac{1}{n}(\beta+\sigma_u\beta^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)\cdot$$
$$\sum_{j=1}^{n}[Z_j\exp((\beta+\sqrt{2}/\sigma_u)Z_j)I_{[Z_j< Z_i]} - EZ\exp((\beta+\sqrt{2}/\sigma_u)Z)I_{[Z< Z_i]}$$
$$\left.+EZ\exp((\beta+\sqrt{2}/\sigma_u)Z)I_{[Z< Z_i]}]\right\}.$$

Denote

$$A_{n1}(Z_i) = \frac{1}{n}\sum_{j=1}^{n}\exp((\beta-\sqrt{2}/\sigma_u)Z_j)I_{[Z_j\geq Z_i]} - E\exp((\beta-\sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]},$$
$$A_{n2}(Z_i) = \frac{1}{n}\sum_{j=1}^{n}Z_j\exp((\beta-\sqrt{2}/\sigma_u)Z_j)I_{[Z_j\geq Z_i]} - EZ\exp((\beta-\sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]},$$

$$A_{n3}(Z_i) = \frac{1}{n}\sum_{j=1}^{n} \exp((\beta + \sqrt{2}/\sigma_u)Z_j)I_{[Z_j<Z_i]} - E\exp((\beta + \sqrt{2}/\sigma_u)Z)I_{[Z<Z_i]},$$

$$A_{n4}(Z_i) = \frac{1}{n}\sum_{j=1}^{n} Z_j\exp((\beta + \sqrt{2}/\sigma_u)Z_j)I_{[Z_j<Z_i]} - EZ\exp((\beta + \sqrt{2}/\sigma_u)Z)I_{[Z<Z_i]},$$

then the second term of $\dot{\hat{L}}_n(\beta)$ is equal to

$$\begin{aligned}
= &-\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i)]W(Z_i)(1 - \sigma_u\beta/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)A_{n1}(Z_i)\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i)]W(Z_i)(1 - \sigma_u\beta/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)E\exp((\beta - \sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(1 - \sigma_u\beta/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)A_{n1}(Z_i)\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(1 - \sigma_u\beta/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)E\exp((\beta - \sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i)]W(Z_i)(\beta - \sigma_u\beta^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)A_{n2}(Z_i)\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i)]W(Z_i)(\beta - \sigma_u\beta^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)EZ\exp((\beta - \sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(\beta - \sigma_u\beta^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)A_{n2}(Z_i)\\
&-\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(\beta - \sigma_u\beta^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)EZ\exp((\beta - \sqrt{2}/\sigma_u)Z)I_{[Z\geq Z_i]}\\
&+\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i)]W(Z_i)(1 + \sigma_u\beta/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)A_{n3}(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i) - g(Z_i)]W(Z_i)(1 + \sigma_u\beta/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)E\exp((\beta + \sqrt{2}/\sigma_u)Z)I_{[Z<Z_i]}\\
&+\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(1 + \sigma_u\beta/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)A_{n3}(Z_i)\\
&+\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(1 + \sigma_u\beta/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)E\exp((\beta + \sqrt{2}/\sigma_u)Z)I_{[Z<Z_i]}
\end{aligned}$$

$$+\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i)-g(Z_i)]W(Z_i)(\beta+\sigma_u\beta^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)A_{n4}(Z_i)$$
$$+\frac{2}{n}\sum_{i=1}^{n} Y_i[\hat{g}(Z_i)-g(Z_i)]W(Z_i)(\beta+\sigma_u\beta^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)EZ\exp((\beta+\sqrt{2}/\sigma_u)Z)I_{[Z<Z_i]}$$
$$+\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(\beta+\sigma_u\beta^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)A_{n4}(Z_i)$$
$$+\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)W(Z_i)(\beta+\sigma_u\beta^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)EZ\exp((\beta+\sqrt{2}/\sigma_u)Z)I_{[Z<Z_i]}.$$

By assuming that $EYW(Z)\exp\left(\sqrt{2}Z/\sigma_u\right)<\infty$, the first term in the above expression is $o_p(1)$. In fact, it is bounded above by

$$\frac{1}{n}\sum_{i=1}^{n} Y_iW(Z_i)\left|1-\frac{\sigma_u}{\sqrt{2}}\beta\right|\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right)\sup_{1\le i\le n}|A_{n1}(Z_i)|\sup_{1\le i\le n}|\hat{g}(Z_i)-g(Z_i)|=o_p(1).$$

Similarly, by imposing conditions such as above, we could show all other terms involving $\hat{g}-g$, $A_{nj}, j=1,2,3,4$ are of the order of $o_p(1)$. The terms not involving these terms are $-\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)\dot{T}(Z_i,\beta)W(Z_i)$. Hence

$$\sup_{\beta}\left|\frac{2}{n}\sum_{i=1}^{n} Y_i\hat{g}(Z_i)\dot{\hat{T}}(Z_i,\beta)W(Z_i)-\frac{2}{n}\sum_{i=1}^{n} Y_i g(Z_i)\dot{T}(Z_i,\beta)W(Z_i)\right|=o_p(1).$$

Using similar arguments, we can show that $\sup_{\beta\in\Theta}|\dot{\hat{L}}_n(\beta)-\dot{L}_n(\beta)|=o_p(1)$. In the meanwhile, the second claim in (2.7), i.e. $\sup_{\beta\in\Theta}|\dot{L}_n(\beta)-\dot{L}(\beta)|=o_p(1)$, could be justified by using law of large numbers for each $\beta\in\Theta$, the continuity of $\dot{L}_n(\beta)-\dot{L}(\beta)$ as a function of $\beta\in\Theta$ and the compactness of Θ. □

To show the asymptotic normality of $\hat{\beta}_n$, we need the following lemma whose proof can be found in Shi and Song (2015).

Lemma 5. *Assume that $\mu(x)$ is a continuous function, and the density function g of Z is* *twice differentiable with bounded second derivative. Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mu(Z_i)[\hat{g}_n(Z_i)-g(Z_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\mu(Z_i)g(Z_i)-E\mu(Z)g(Z)] + o_p(1).$$

Proof of Theorem 2. Let $\hat{\beta}_n$ be the solution of $\dot{\hat{L}}_n(\beta)=0$. By Taylor expansion,

$$0 = \dot{\hat{L}}_n(\hat{\beta}_n) = \dot{\hat{L}}_n(\beta_0) + \ddot{\hat{L}}_n(\tilde{\beta}_n)(\hat{\beta}_n-\beta_0).$$

where $\tilde{\beta}_n$ is between $\hat{\beta}_n$ and β_0. Denote $T_{1i}=T(Z_i,\beta_0)$, $T_{2i}=T(Z_i,2\beta_0)$, $W_i=W(Z_i)$, $\xi_i = Y_i-\exp(Z_i\beta_0)$, $\eta_i = Y_i^2-\exp(2Z_i\beta_0)-\exp(Z_i\beta_0)$, $V_i = 2Z_i\exp(2Z_i\beta_0)+Z_i\exp(Z_i\beta_0)$, $S_i = Z_i\exp(Z_i\beta_0)$. We have

$$\begin{aligned}
\dot{\hat{L}}_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\Big[(Y_i-\exp(Z_i\beta_0))\,\hat{g}_i-\hat{T}(Z_i,\beta_0)\Big]\cdot\Big[-Z_i\exp(Z_i\beta_0)\hat{g}_i-\dot{\hat{T}}(Z_i,\beta_0)\Big]\\
&+\Big[(Y_i^2-\exp(2Z_i\beta_0)-\exp(Z_i\beta_0))\hat{g}_i-\hat{T}(Z_i,2\beta_0)-\hat{T}(Z_i,\beta_0)\Big]\cdot\\
&\Big[-(2Z_i\exp(2Z_i\beta_0)+Z_i\exp(Z_i\beta_0))\,\hat{g}_i-\dot{\hat{T}}(Z_i,2\beta_0)-\dot{\hat{T}}(Z_i,\beta_0)\Big]\Big\}W(Z_i)\\
= -\frac{1}{n}\sum_{i=1}^{n}\Big\{&\Big[(Y_i-\exp(Z_i\beta_0))\,\hat{g}_i-\hat{T}_{1i}\Big]\cdot\Big[Z_i\exp(Z_i\beta_0)\hat{g}_i+\dot{\hat{T}}_{1i}\Big]\\
&+\Big[(Y_i^2-\exp(2Z_i\beta_0)-\exp(Z_i\beta_0))\hat{g}_i-\hat{T}_{2i}-\hat{T}_{1i}\Big]\cdot\\
&\Big[(2Z_i\exp(2Z_i\beta_0)+Z_i\exp(Z_i\beta_0))\,\hat{g}_i+\dot{\hat{T}}_{2i}+\dot{\hat{T}}_{1i}\Big]\Big\}W_i\\
= -\frac{1}{n}\sum_{i=1}^{n}\Big\{&(\xi_i\hat{g}_i-\hat{T}_{1i})(S_i\hat{g}_i+\dot{\hat{T}}_{1i})+(\eta_i\hat{g}_i-\hat{T}_{2i}-\hat{T}_{1i})(V_i\hat{g}_i+\dot{\hat{T}}_{2i}+\dot{\hat{T}}_{1i})\Big\}W_i.
\end{aligned}$$

Adding and subtracting g_i, T_{1i}, $\dot{T}_{1i}$, T_{2i}, $\dot{T}_{2i}$ from $\hat{g}_i$, $\hat{T}_{1i}$, $\dot{\hat{T}}_{1i}$, $\hat{T}_{2i}$, $\dot{\hat{T}}_{2i}$, respectively, also, define $\Delta g_i = \hat{g}(Z_i)-g(Z_i)$, $\Delta T_{1i} = \hat{T}(Z_i,\beta_0)-T(Z_i,\beta_0)$, $\Delta T_{2i} = \hat{T}(Z_i,2\beta_0)-T(Z_i,2\beta_0)$, and $\Delta\dot{T}_{1i} = \dot{\hat{T}}(Z_i,\beta_0)-\dot{T}(Z_i,\beta_0)$, $\Delta\dot{T}_{2i} = \dot{\hat{T}}(Z_i,2\beta_0)-\dot{T}(Z_i,2\beta_0)$, we can rewrite $\dot{\hat{L}}_n(\beta_0)$

as

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}[(\xi_i g_i - T_{1i})(S_i g_i + \dot{T}_{1i}) + (\eta_i g_i - T_{2i} - T_{1i})(V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})]W_i \\
&-\frac{1}{n}\sum_{i=1}^{n}[\xi_i g_i - T_{1i}][S_i \Delta g_i + \Delta\dot{T}_{1i}]W_i - \frac{1}{n}\sum_{i=1}^{n}[\xi_i \Delta g_i - \Delta T_{1i}][S_i g_i + \dot{T}_{1i}]W_i \\
&-\frac{1}{n}\sum_{i=1}^{n}(\eta_i g_i - T_{2i} - T_{1i})[V_i \Delta g_i + \Delta\dot{T}_{2i} + \Delta\dot{T}_{1i}]W_i - \frac{1}{n}\sum_{i=1}^{n}[\eta_i \Delta g_i - \Delta T_{2i} - \Delta T_{1i}][V_i g_i + \dot{T}_{2i} + \dot{T}_{1i}]W_i \\
&-\frac{1}{n}\sum_{i=1}^{n}[\xi_i \Delta g_i - \Delta T_{1i}][S_i \Delta g_i + \Delta\dot{T}_{1i}]W_i - \frac{1}{n}\sum_{i=1}^{n}[\eta_i \Delta g_i - \Delta T_{2i} - \Delta T_{1i}][V_i \Delta g_i + \Delta\dot{T}_{2i} + \Delta\dot{T}_{1i}]W_i \\
&= \\
&-\frac{1}{n}\sum_{i=1}^{n}[(\xi_i g_i - T_{1i})(S_i g_i + \dot{T}_{1i}) + (\eta_i g_i - T_{2i} - T_{1i})(V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})]W_i \\
&-\frac{1}{n}\sum_{i=1}^{n}[(\xi_i g_i - T_{1i})S_i + (S_i g_i + \dot{T}_{1i})\xi_i + (\eta_i g_i - T_{2i} - T_{1i})V_i + (V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})\eta_i]W_i \Delta g_i \\
&+\frac{1}{n}\sum_{i=1}^{n}[(S_i g_i + \dot{T}_{1i}) + (V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})]W_i \Delta T_{1i} + \frac{1}{n}\sum_{i=1}^{n}(V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})W_i \Delta T_{2i} \\
&-\frac{1}{n}\sum_{i=1}^{n}[(\xi_i g_i - T_{1i}) + (\eta_i g_i - T_{2i} - T_{1i})]W_i \Delta\dot{T}_{1i} - \frac{1}{n}\sum_{i=1}^{n}(\eta_i g_i - T_{2i} - T_{1i})W_i \Delta\dot{T}_{2i} \\
&-\frac{1}{n}\sum_{i=1}^{n}[\xi_i \Delta g_i - \Delta T_{1i}][S_i \Delta g_i + \Delta\dot{T}_{1i}]W_i - \frac{1}{n}\sum_{i=1}^{n}[\eta_i \Delta g_i - \Delta T_{2i} - \Delta T_{1i}][V_i \Delta g_i + \Delta\dot{T}_{2i} + \Delta\dot{T}_{1i}]W_i.
\end{aligned}
$$

For brevity, we denote the eight terms on the right hand side of the above expression as S_{nj}, $j = 1, 2, \ldots, 8$. Recall that the true parameter β_0 is the solution of $\dot{L}(\beta) = 0$. Then since

$$L(\beta) = E[(Y - E(\exp(X\beta)|Z))^2 + (Y^2 - E(\exp(X\beta)|Z) - E(\exp(2X\beta)|Z))^2]g^2(Z)W(Z),$$

we have

$$
\begin{aligned}
\dot{L}(\beta_0) = &-2E[(Y - E(\exp(X\beta_0)|Z) \cdot E(X\exp(X\beta_0)|Z)) \\
&+ (Y^2 - E(\exp(X\beta_0|Z) - E(\exp(2X\beta_0)|Z)) \cdot (E(X\exp(X\beta_0)|Z) + 2E(X\exp(2X\beta_0)|Z)))] \cdot \\
&\quad g^2(Z)W(Z).
\end{aligned}
$$

Also since

$$Y - E(\exp(X\beta_0)|Z) = \xi - \frac{T_1}{g(Z)}, \quad E(X\exp(X\beta_0)|Z) = Z\exp(Z\beta_0) + \frac{\dot{T}_1}{g} = S + \frac{\dot{T}_1}{g},$$

and

$$Y^2 - E(\exp(X\beta_0)|Z) - E(\exp(2X\beta_0)|Z) = \eta - \frac{T_2}{g} - \frac{T_1}{g},$$

$$E(X\exp(X\beta_0)|Z) + 2E(X\exp(2X\beta_0)|Z) = V + \frac{\dot{T}_2}{g} + \frac{\dot{T}_1}{g},$$

we have

$$E\left[\left(\xi - \frac{T_1}{g}\right)\left(S + \frac{\dot{T}_1}{g}\right) + \left(\eta - \frac{T_2}{g} - \frac{T_1}{g}\right)\left(V + \frac{\dot{T}_2}{g} + \frac{\dot{T}_1}{g}\right)\right] g^2 W = 0,$$

or $E\left[(\xi g - T_1)\left(Sg + \dot{T}_1\right) + (\eta g - T_2 - T_1)\left(Vg + \dot{T}_2 + \dot{T}_1\right)\right] W = 0$. For each $i = 1, 2, \ldots, n$, denote

$$e_{1i} = [(\xi_i g_i - T_{1i})(S_i g_i + \dot{T}_{1i}) + (\eta_i g_i - T_{2i} - T_{1i})(V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})]W_i, \tag{2.8}$$

then S_{n1} can be written as $n^{-1/2}\sum_{i=1}^n e_{1i}$. In fact, one can easily verify that

$$e_{1i} = \left\{[Y_i - E(Y_i|Z_i)]\frac{\partial E(Y_i|Z_i)}{\partial \beta} + [Y_i^2 - E(Y_i^2|Z_i)]\frac{\partial E(Y_i^2|Z_i)}{\partial \beta}\right\} g^2(Z_i)W(Z_i)$$

with $\beta = \beta_0$. By CLT, we have $\sqrt{n}S_{n1} \Longrightarrow N(0, \tau^2)$,
where $\tau^2 = E\left[(\xi g - T_1)(Sg + \dot{T}_1) + (\eta g - T_2 - T_1)(Vg + \dot{T}_2 + \dot{T}_1)\right]^2 W$ provided τ^2 exists. For S_{n2}, denote the coefficients of $W_i \Delta g_i$ as μ_{2i}, $S_{n2} = n^{-1}\sum_{i=1}^n \mu_{2i} W_i(\hat{g}_i - g_i)$. And if we further denote, for each i,

$$e_{2i} = \mu_{2i} W_i g_i - E\mu_2 W g, \tag{2.9}$$

then by Lemma 5, $\sqrt{n}S_{n2} = n^{-1/2}\sum_{i=1}^n e_{2i} + o_p(1)$. Further computation shows that

$$\begin{aligned}
e_{2i} =& W(Z_i)g(Z_i)\left([Y_i - E(Y_i|Z_i)]\frac{\partial \exp(Z_i\beta_0)}{\partial\beta}\right) \\
&+ W(Z_i)g(Z_i)\left([Y_i^2 - E(Y_i^2|Z_i)]\frac{\partial[\exp(Z_i\beta_0)+\exp(2Z_i\beta_0)]}{\partial\beta}\right) \\
&- \left(W(Z_i)g(Z_i)[Y_i - \exp(Z_i\beta_0)]\frac{\partial E(Y_i|Z_i)}{\partial\beta} - EW(Z)g(Z)\left([Y - \exp(Z\beta_0)]\frac{\partial E(Y|Z)}{\partial\beta}\right)\right) \\
&- \left(W(Z_i)g(Z_i)\left([Y_i^2 - \exp(Z_i\beta_0) - \exp(2Z_i\beta_0)]\frac{\partial E(Y_i^2|Z_i)}{\partial\beta}\right)\right. \\
&\quad \left. - EW(Z)g(Z)\left([Y^2 - \exp(Z\beta_0) - \exp(2Z\beta_0)]\frac{\partial E(Y^2|Z)}{\partial\beta}\right)\right).
\end{aligned}$$

While for S_{n3}, if we denote the coefficients of $W_i\Delta T_{1i}$ in S_{n3} as μ_{3i}, we have $\sqrt{n}S_{n3} = n^{-1}\sum_{i=1}^n \mu_{3i}W_i\Delta T_{1i}$ which can be written as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \mu_{3i}W_i \cdot \\
&\left\{\left(\beta_0 - \frac{\sigma_u^2}{2\sqrt{2}}\beta_0^2\right)\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right)\cdot\right. \\
&\frac{1}{n}\sum_{j=1}^n\left[\exp((\beta_0 - \sqrt{2}/\sigma_u)Z_j)I_{[Z_j\geq Z_i]} - \int_{Z_i}^{\infty}\exp((\beta_0-\sqrt{2}/\sigma_u)x)g(x)dx\right] \\
&-\left(\beta_0 + \frac{\sigma_u^2}{2\sqrt{2}}\beta_0^2\right)\exp\left(\frac{-\sqrt{2}Z_i}{\sigma_u}\right)\cdot \\
&\left.\frac{1}{n}\sum_{j=1}^n\left[\exp((\beta_0 + \sqrt{2}/\sigma_u)Z_j)I_{[Z_j< Z_i]} - \int_{-\infty}^{Z_i}\exp((\beta_0+\sqrt{2}/\sigma_u)x)g(x)dx\right]\right\}.
\end{aligned}$$

Note that $\mu_{3i} = -(S_ig_i + \dot{T}_{1i}) - (V_ig_i + \dot{T}_{2i} + \dot{T}_{1i})$. Let us further denote

$$\tilde{W}_i = \mu_{3i}W_i\left(\beta_0 - \frac{\sigma_u^2}{2\sqrt{2}}\beta_0^2\right)\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right),$$

then the first term in $\sqrt{n}S_{n3}$ can be written as $\sqrt{n}S_{n31} = n^{-3/2}\sum_{i=1}^n\sum_{j=1}^n \tilde{W}_iV_{ij}$, where

$$V_{ij} = \exp((\beta_0 - \sqrt{2}/\sigma_u)Z_j)I_{[Z_j\geq Z_i]} - \int_{Z_i}^{\infty}\exp((\beta_0-\sqrt{2}/\sigma_u)x)g(x)dx.$$

Denote $U_{ij} = (\tilde{W}_i V_{ij} + \tilde{W}_j V_{ji})/2$, then

$$\begin{aligned}\sqrt{n} S_{n31} &= \frac{1}{n\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n V_{ij} \tilde{W}_i \frac{1}{n\sqrt{n}} \sum_{i=1}^n \tilde{W}_i V_{ii} + \frac{1}{n\sqrt{n}} \sum_{i \neq j} \tilde{W}_i V_{ij} \\ &= \frac{1}{n\sqrt{n}} \sum_{i=1}^n \tilde{W}_i V_{ii} + \frac{n-1}{\sqrt{n}} \frac{2}{n(n-1)} \sum_{i<j} U_{ij} = \frac{1}{n\sqrt{n}} \sum_{i=1}^n \tilde{W}_i V_{ii} + \frac{n-1}{\sqrt{n}} U_n.\end{aligned}$$

Assume that $E\tilde{W}_i V_{ii} < \infty$, then the first term above is $o_p(1)$. And it is easy to see that U_n is a U-statistic. If $E[U_{12}]^2 < \infty$, and $Var[E(U_{12}|Z_1)] > 0$, then

$$\sqrt{n}[U_n - E(U_{12})] = \frac{1}{\sqrt{n}} \sum_{i=1}^n [E(U_{ij}|Z_i) - E(U_{12})] + o_p(1).$$

Note that

$$\begin{aligned}E[\tilde{W}_1 V_{12}|Z_1] = E\Big[&[-(S_1 g_1 + \dot{T}_{11}) - (V_1 g_1 + \dot{T}_{21} + \dot{T}_{11})] W_1 (\beta_0 - \sigma_u^2/(2\sqrt{2})\beta_0^2) \exp(\sqrt{2} Z_1/\sigma_u) \cdot \\ &[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z_2) I_{[Z_2 \geq Z_1]} - \int_{Z_1}^\infty \exp((\beta_0 - \sqrt{2}/\sigma_u)x) g(x) dx] | Z_1\Big] = 0,\end{aligned}$$

and

$$\begin{aligned}E[\tilde{W}_2 V_{21}|Z_1] &= E\Big[[-(S_2 g_2 + \dot{T}_{12}) - (V_2 g_2 + \dot{T}_{22} + \dot{T}_{12})] W_2 (\beta_0 - \sigma_u^2 \beta_0^2/(2\sqrt{2})) \exp(\sqrt{2} Z_2/\sigma_u) \cdot \\ &\quad \Big[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z_1) I_{[Z_2 < Z_1]} - \int_{Z_2}^\infty \exp((\beta_0 - \sqrt{2}/\sigma_u)x) g(x) dx\Big] | Z_1\Big] \\ &= \exp((\beta_0 - \sqrt{2}/\sigma_u) Z_1) \int \tilde{W}(v) I_{[v < Z_1]} g(v) dv - \int_{-\infty}^\infty \int_v^\infty \exp((\beta_0 - \sqrt{2}/\sigma_u)x) g(x) \tilde{W}(v) g(v) dx dv \\ &= \exp((\beta_0 - \sqrt{2}/\sigma_u) Z_1) \int_{-\infty}^{Z_1} \tilde{W}(v) g(v) dv - \int_{-\infty}^\infty \int_{-\infty}^\infty \exp((\beta_0 - \sqrt{2}/\sigma_u)x) I_{[x > Z_2]} g(x) \tilde{W}(v) g(v) dv dx \\ &= \exp((\beta_0 - \sqrt{2}/\sigma_u) Z_1) \int_{-\infty}^{Z_1} \tilde{W}(v) g(v) dv - \int_{-\infty}^\infty \int_{-\infty}^x \exp((\beta_0 - \sqrt{2}/\sigma_u)x) \tilde{W}(v) g(v) g(x) dv dx.\end{aligned}$$

Thus

$$E[U_{12}|Z_1] = \frac{1}{2}\left[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z_1) \int_{-\infty}^{Z_1} \tilde{W}(v) g(v) dv - E\left[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z_1) \int_{-\infty}^{Z_1} \tilde{W}(x) g(x) dx\right]\right],$$

and

$$\sqrt{n}(U_n - EU_{12}) = \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \left[(\exp((\beta_0 - \sqrt{2}/\sigma_u) Z_i) \int_{-\infty}^{Z_i} \tilde{W}(x) g(x) dx - \right.$$
$$\left. E\left[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z) \int_{-\infty}^{Z} \tilde{W}(x) g(x) dx \right] - EU_{12} \right] + o_p(1)$$

And since $EU_{12} = 0$, $\sqrt{n}U_n$ can be written as $n^{-1/2} \sum_{i=1}^{n} e_{31i} + o_p(1)$, where

$$e_{31i} = \frac{1}{2} \left[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z_i) \int_{-\infty}^{Z_i} \tilde{W}(x) g(x) dx - E\left[\exp((\beta_0 - \sqrt{2}/\sigma_u) Z) \int_{-\infty}^{Z} \tilde{W}(x) g(x) dx \right] \right] \tag{2.10}$$

and an $o_p(1)$ term. Similarly, if we denote

$$\hat{W}_i = \left(\beta_0 + \frac{\sigma_u^2}{2\sqrt{2}} \beta_0^2 \right) \mu_{3i} W_i \exp\left(\frac{-\sqrt{2} Z_i}{\sigma_u} \right),$$

and $\hat{V}_{ij} = \exp((\beta_0 + \sqrt{2}/\sigma_u) Z_j) I_{[Z_j < Z_i]} - \int_{-\infty}^{Z_i} \exp((\beta_0 + \sqrt{2}/\sigma_u) x) g(x) dx$, the second term in $\sqrt{n} S_{n3}$ can be written as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{W}_i \frac{1}{n} \sum_{j=1}^{n} \hat{V}_{ij} = \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \left[\exp((\beta_0 + \sqrt{2}/\sigma_u) Z_i) \int_{Z_i}^{\infty} \hat{W}(x) g(x) dx \right.$$
$$\left. - E\left[\exp((\beta_0 + \sqrt{2}/\sigma_u) Z) \int_{Z}^{\infty} \hat{W}(x) g(x) dx \right] \right] + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_{32i} + o_p(1),$$

where, for each i,

$$e_{32i} = \frac{1}{2} \left[\exp((\beta_0 + \sqrt{2}/\sigma_u) Z_i) \int_{Z_i}^{\infty} \hat{W}(x) g(x) dx - E\left[\exp((\beta_0 + \sqrt{2}/\sigma_u) Z) \int_{Z}^{\infty} \hat{W}(x) g(x) dx \right] \right]. \tag{2.11}$$

Hence for S_{n3}, we have $\sqrt{n} S_{n3} = n^{-1/2} \sum_{i=1}^{n} e_{3i} + o_p(1)$, where

$$e_{3i} = e_{31i} + e_{32i}, \tag{2.12}$$

and e_{31i}, e_{32i} are defined in (2.10), (2.11), respectively. Let $\nu_3(z) = -g(Z)[\partial E_\beta(Y|Z)/\partial\beta + \partial E_\beta(Y^2|Z)/\partial\beta]$,

$$L_1(Z,\beta_0) = \frac{1}{2}\exp(\beta_0 Z)\left(\beta_0 - \frac{\sigma_u^2}{2\sqrt{2}}\beta_0^2\right), \quad L_2(Z,\beta_0) = \frac{1}{2}\exp(\beta_0 Z)\left(\beta_0 + \frac{\sigma_u^2}{2\sqrt{2}}\beta_0^2\right).$$

We can rewrite $e_{31i} + e_{32i} = \xi_{3i} - E\xi_{3i}$, where

$$\begin{aligned}\xi_{3i} =& L_1(Z_i,\beta_0)E\left[\nu_3(Z)w(Z)\exp(-\sqrt{2}|Z-Z_i|/\sigma_u)I(Z\le Z_i)\Big|Z_i\right] \\ &+ L_2(Z_i,\beta_0)E\left[\nu_3(Z)w(Z)\exp(-\sqrt{2}|Z-Z_i|/\sigma_u)I(Z> Z_i)\Big|Z_i\right]\end{aligned}$$

Now let's consider S_{n4}. If we denote $\mu_{4i} = (V_i g_i + \dot{T}_{2i} + \dot{T}_{1i})W_i$, then $\sqrt{n}S_{n4}$ could be written as

$$\begin{aligned}\sqrt{n}S_{n4} =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mu_{4i}\Big\{(2\beta_0 - 2\sigma_u^2\beta_0^2/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)\cdot \\ & \frac{1}{n}\sum_{j=1}^{n}\left[\exp((2\beta_0-\sqrt{2}/\sigma_u)Z_j)I_{[Z_j\ge Z_i]} - \int_{Z_i}^{\infty}\exp((2\beta_0-\sqrt{2}/\sigma_u)x)g(x)dx\right] \\ & - (2\beta_0 + 2\sigma_u^2\beta_0^2/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)\cdot \\ & \frac{1}{n}\sum_{j=1}^{n}\left[\exp((2\beta_0+\sqrt{2}/\sigma_u)Z_j)I_{[Z_j< Z_i]} - \int_{-\infty}^{Z_i}\exp((2\beta_0+\sqrt{2}/\sigma_u)x)g(x)dx\right]\Big\}.\end{aligned}$$

Denote

$$\tilde{W}_i = \left(2\beta_0 - \frac{2\sigma_u^2}{\sqrt{2}}\beta_0^2\right)\mu_{4i}\exp\left(\frac{\sqrt{2}Z_i}{\sigma_u}\right), \quad \hat{W}_i = \left(2\beta_0 + \frac{2\sigma_u^2}{\sqrt{2}}\beta_0^2\right)\mu_{4i}\exp\left(-\frac{\sqrt{2}Z_i}{\sigma_u}\right),$$

then $\sqrt{n}S_{n4}$ can be written as $n^{-1/2}\sum_{i=1}^{n} e_{4i} + o_p(1)$, where, for each i,

$$\begin{aligned}e_{4i} =& \frac{1}{2}\Big\{\Big[\exp((2\beta_0-\sqrt{2}/\sigma_u)Z_i)\int_{-\infty}^{Z_i}\tilde{W}(x)g(x)dx - E\left[\exp((2\beta_0-\sqrt{2}/\sigma_u)Z)\int_{-\infty}^{Z}\tilde{W}(x)g(x)dx\right]\Big] + \\ & \Big[\exp((2\beta_0+\sqrt{2}/\sigma_u)Z_i)\int_{Z_i}^{\infty}\hat{W}(x)g(x)dx - E\left[\exp((2\beta_0+\sqrt{2}/\sigma_u)Z)\int_{Z}^{\infty}\hat{W}(x)g(x)dx\right]\Big]\Big\}.\end{aligned} \tag{2.13}$$

In fact, after some algebra, we can rewrite e_{4i} as $\xi_{4i} - E\xi_{4i}$, where

$$\begin{aligned}
\xi_{4i} = & -L_1(Z_i,\beta_0)E\left[\dot{K}_1(Z)g(Z)W(Z)\exp(-\sqrt{2}|Z-Z_i|/\sigma_u)I(Z \le Z_i)\Big|Z_i\right] \\
& - L_2(Z_i,\beta_0)E\left[\dot{K}_1(Z)g(Z)W(Z)\exp(-\sqrt{2}|Z-Z_i|/\sigma_u)I(Z > Z_i)\Big|Z_i\right] \\
& - [L_1(Z_i,\beta_0) - L_1(2\beta_0, Z_i)]E\left[\dot{K}_1(Z)g(Z)W(Z)\exp(-\sqrt{2}|Z-Z_i|/\sigma_u)I(Z \le Z_i)\Big|Z_i\right] \\
& - [L_2(Z_i,\beta_0) - L_2(2\beta_0, Z_i)]E\left[\dot{K}_1(Z)g(Z)W(Z)\exp(-\sqrt{2}|Z-Z_i|/\sigma_u)I(Z > Z_i)\Big|Z_i\right]
\end{aligned}$$

Next, let's consider S_{n5}. Let $\mu_{5i} = [(\xi_i g_i - T_{1i}) + (\eta_i g_i - T_{2i} - T_{1i})]W_i$, then

$$\begin{aligned}
\sqrt{n}S_{n5} = & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mu_{5i}\Bigg\{(1-\sigma_u\beta_0/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j \ge Z_i]} \\
& +(\beta_0-\sigma_u\beta_0^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)\frac{1}{n}\sum_{j=1}^{n}Z_j\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j \ge Z_i]} \\
& -(1+\sigma_u\beta_0/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)\frac{1}{n}\sum_{j=1}^{n}\exp[(\beta_0+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j < Z_i]} \\
& -(\beta_0+\sigma_u\beta_0^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)\frac{1}{n}\sum_{j=1}^{n}Z_j\exp[(\beta_0+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j < Z_i]} \\
& -(1-\sigma_u\beta_0/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)\int_{Z_i}^{\infty}g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx \\
& -(\beta_0-\sigma_u\beta_0^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)\int_{Z_i}^{\infty}g(x)x\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx \\
& +(1+\sigma_u\beta_0\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)\int_{-\infty}^{Z_i}g(x)\exp[(\beta_0+\sqrt{2}\sigma_u)x]dx \\
& +(\beta_0+\sigma_u\beta_0^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)\int_{-\infty}^{Z_i}g(x)x\exp[(\beta_0+\sqrt{2}/\sigma_u)x]dx\Bigg\}.
\end{aligned}$$

Note that

$$E[\xi g - T|Z] = E\left[\left(\xi - \frac{T}{g}\right)g|Z\right] = g(Z)E[(Y - E(\exp(\beta X)|Z))] = 0, \qquad (2.14)$$

now let's consider

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i g_i - T_{1i})W_i\Big\{(1-\sigma_u\beta_0/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)\cdot \\
&\quad\frac{1}{n}\sum_{j=1}^{n}\Big[\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]} - \int_{Z_i}^{\infty} g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\Big]\Big\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i g_i - T_{1i})W_i\Big\{(\beta_0-\sigma_u\beta_0^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)\cdot \\
&\quad\frac{1}{n}\sum_{j=1}^{n}\Big[Z_j\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]} - \int_{Z_i}^{\infty} g(x)x\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\Big]\Big\} \qquad (2.15) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i g_i - T_{1i})W_i\Big\{(1+\sigma_u\beta_0/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)\cdot \\
&\quad\frac{1}{n}\sum_{j=1}^{n}\Big[\exp[(\beta_0+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j< Z_i]} - \int_{-\infty}^{Z_i} g(x)\exp[(\beta_0+\sqrt{2}/\sigma_u)x]dx\Big]\Big\} \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i g_i - T_{1i})W_i\Big\{(\beta_0+\sigma_u\beta_0^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)\cdot \\
&\quad\frac{1}{n}\sum_{j=1}^{n}\Big[Z_j\exp[(\beta_0+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j< Z_i]} - \int_{-\infty}^{Z_i} g(x)x\exp[(\beta_0+\sqrt{2}/\sigma_u)x]dx\Big]\Big\}.
\end{aligned}
$$

Denote

$$
\begin{aligned}
\zeta_i &= (1-\sigma_u\beta_0/\sqrt{2})(\xi_i g_i - T_i)W_i\exp(\sqrt{2}Z_i/\sigma_u), \\
V_{ij} &= \exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]} - \int_{Z_i}^{\infty} g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx.
\end{aligned}
$$

Then the first term in (2.15) can be written as $n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_i V_{ij}$. Rearrange it as

$$
\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\zeta_i V_{ii} + \frac{1}{n\sqrt{n}}\sum_{i\neq j}\zeta_i V_{ij}. \qquad (2.16)
$$

The first term in (2.16), under the assumption $E|\zeta_i V_{ii}| < \infty$, is $o_p(1)$. For the second term, note that

$$
E\left(\frac{1}{n\sqrt{n}}\sum_{i\neq j}\zeta_i V_{ij}\right) = \frac{n(n-1)}{n\sqrt{n}}E\zeta_1 V_{12} = 0,
$$

and

$$E\left(\frac{1}{n\sqrt{n}}\sum_{i\neq j}\zeta_i V_{ij}\right)^2 = \frac{1}{n^3}\sum_{i\neq j}E\zeta_i^2V_{ij}^2 + \frac{1}{n^3}\sum_{i\neq j}\sum_{\substack{k\neq l,\\(i,j)\neq(k,l)}}E\zeta_i\zeta_kV_{ij}V_{kl}.$$

By assuming that $E\zeta_1^2V_{12}^2 < \infty$, $n^{-3}\sum_{i\neq j}E\zeta_i^2V_{ij}^2 = n^{-2}(n-1)E\zeta_1^2V_{12}^2 \to 0$. For the cross-over terms, we consider several cases.

(i). $i = k$, $j \neq l$. Without loss of generality, let $i = k = 1$, $j = 2$, $l = 3$,

$$\begin{aligned}E\zeta_1^2V_{12}V_{13} = E\zeta_1^2&\left[\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_2]I_{[Z_2\geq Z_1]} - \int_{Z_1}^{\infty}g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\right]\cdot\\&\left[\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_3]I_{[Z_3\geq Z_1]} - \int_{Z_1}^{\infty}g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\right] = 0.\end{aligned}$$

(ii). $i \neq k$, $j = l$. Without loss of generality, let $i = 1$, $k = 2$, $j = l = 3$,

$$\begin{aligned}E\zeta_1\zeta_2V_{13}V_{23} = E\zeta_1^2&\left[\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_3]I_{[Z_3\geq Z_1]} - \int_{Z_1}^{\infty}g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\right]\cdot\\&\left[\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_3]I_{[Z_3\geq Z_2]} - \int_{Z_2}^{\infty}g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\right]\\&= E[\zeta_1\zeta_2E[V_{13}V_{23}|Z_1,Z_2]] = E[E[V_{13}V_{23}|Z_1,Z_2]\cdot E[\zeta_1\zeta_2|Z_1,Z_2]] = 0.\end{aligned}$$

by (2.14).

(iii). $i \neq k$, $j \neq l$. Without loss of generality, let $i = l = 1$, $k = j = 2$, $E\zeta_1\zeta_2V_{12}V_{21} = 0$ by (2.14).

(iv). i, j, k, l are all different. Without loss of generality, let $i = 1$, $j = 2$, $k = 3$, $l = 4$, $E\zeta_1\zeta_3V_{12}V_{34} = 0$.

So, the first term in (2.15) is $o_p(1)$. Similarly, the second, third and fourth term in (2.15) are also $o_p(1)$. Therefore, for $\sqrt{n}S_{n5}$, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mu}_{5i}W_i\Big\{(1-\sigma_u\beta_0/\sqrt{2})\exp(\sqrt{2}Z_i/\sigma_u)\cdot \\
&\frac{1}{n}\sum_{j=1}^{n}\left[\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]}-\int_{Z_i}^{\infty}g(x)\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\right]\Big\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mu}_{5i}W_i\Big\{(\beta_0-\sigma_u\beta_0^2/(2\sqrt{2}))\exp(\sqrt{2}Z_i/\sigma_u)\cdot \\
&\frac{1}{n}\sum_{j=1}^{n}\left[Z_j\exp[(\beta_0-\sqrt{2}/\sigma_u)Z_j]I_{[Z_j\geq Z_i]}-\int_{Z_i}^{\infty}g(x)x\exp[(\beta_0-\sqrt{2}/\sigma_u)x]dx\right]\Big\} \qquad (2.17)\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mu}_{5i}W_i\Big\{(1+\sigma_u\beta_0/\sqrt{2})\exp(-\sqrt{2}Z_i/\sigma_u)\cdot \\
&\frac{1}{n}\sum_{j=1}^{n}\left[\exp[(\beta_0+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j<Z_i]}-\int_{-\infty}^{Z_i}g(x)\exp[(\beta_0+\sqrt{2}/\sigma_u)x]dx\right]\Big\} \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mu}_{5i}W_i\Big\{(\beta_0+\sigma_u\beta_0^2/(2\sqrt{2}))\exp(-\sqrt{2}Z_i/\sigma_u)\cdot \\
&\frac{1}{n}\sum_{j=1}^{n}\left[Z_j\exp[(\beta_0+\sqrt{2}/\sigma_u)Z_j]I_{[Z_j<Z_i]}-\int_{-\infty}^{Z_i}g(x)x\exp[(\beta_0+\sqrt{2}/\sigma_u)x]dx\right]\Big\},
\end{aligned}
$$

where $\tilde{\mu}_{5i}=(\eta_i g_i-T_{2i}-T_{1i})$. Since $E[\eta g-T_2-T_1|Z]=0$, we could show that (2.17) is also $o_p(1)$. Thus, $\sqrt{n}S_{n5}$ is $o_p(1)$. Similarly, one can show that $\sqrt{n}S_{n6}=o_p(1)$.

For S_{n7} and S_{n8}, by Lemma 3, we have $\sup_z|\hat{g}(z)-g(z)|=O(n^{-2/5}(\log n)^{1/3})$, and by Lemma 4,

$$
\sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i I_{[X_i\leq x]}-E\xi_1 I_{[X_1\leq x]}\right|=O\left(\sqrt{\frac{\log\log n}{n}}\right),\quad \text{a.s..}
$$

Thus $\sqrt{n}S_{n7}$ and $\sqrt{n}S_{n8}$ are $\sqrt{n}O(n^{-4/5}(\log n)^{2/3})$, which is $o_p(1)$.

In summary, we eventually obtain $\sqrt{n}\dot{\hat{L}}_n(\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[e_{i1}+e_{i2}+e_{i3}+e_{i4}]+o_p(1)$,

Also note that $\ddot{\hat{L}}_n(\tilde{\beta})\to\ddot{L}(\beta_0)$ in probability, we conclude the proof of Theorem 2 by applying the CLT and Slutsky theorem. □

Chapter 3

Calibrating Poisson Regression Likelihood with Laplace Measurement Error

3.1 Introduction

Poisson linear regression model is commonly used for fitting the count data. In its classical form, given a covariate $X = x$, possibly multidimensional, the response variable Y has a Poisson distribution with intensity function $\lambda(x) = \exp(x^T\beta)$. Throughout this chapter, we shall assume that the predictor X is univariate. In real applications, due to measurement instrument imperfection or human errors, often times the covariate X cannot be observed directly. Instead, one can observe another variable Z which is related to X via the additive structure $Z = X + U$, where X and U are assumed to be independent, and U is called the measurement error. In the measurement error literature, U is usually assumed to be nondifferential, that is, the conditional distribution of Y given X, Z is the same as the conditional distribution of Y given X. In other words, given the true variable X, knowing the variable Z does not provide any extra information on the distribution of Y. This nondifferential condition will be adopted throughout the chapter. In the following, we denote the density function of X, Z as f and g, respectively. An extensive introduction of Poisson regression and other related regression models on count data can be found in Cameron and Trivedi (2013), and a comprehensive discussion on the measurement error modeling can be found in Fuller (2009) and Carroll et al. (2006).

The estimation of β is the main research interest in Poisson regression models. Like other errors-in-variables models, simply replacing X with the data from Z in standard estimation procedures often result in biased estimate. In fact, although the mean and variance function of the response variable Y given the true regressor X in the Poisson regression are the same, the conditional expectation $E(Y|Z)$ and $\text{Var}(Y|Z)$ are not equal. Guo and Li (2002) showed that the measurement error increases the dispersion, that is, $E(Y|Z) \leq \text{Var}(Y|Z)$ holds. The classical estimation procedures, such as the regression calibration, simulation extrapolation and instrumental variable methods can be applied to Poisson regression directly. For example, conditional on a pseudo-sufficient statistics, Stefanski and Carroll (1987) proposed a conditional score estimator in the generalized linear model. However, when applied to Poisson regression, the analysis is complicated due to a summation of an infinite series in the probability mass function of Y given the pseudo-sufficient statistics. Later, Stefanski (1989) and Nakamura (1990) constructed a corrected score method, in which a function of (Y, Z) and β is found so that the its expectation equals to the expectation of the likelihood function based on (Y, X). When applied to Poisson regression model, the function has the form of $\sum_{i=1}^{n} \left[Y_i Z_i'\beta - \ln Y_i! - \exp(Z_i'\beta - 0.5\beta'\Sigma_u\beta)\right]$. Unfortunately, the above corrected likelihood function is not bounded in β. Guo and Li (2002) relaxed the normality assumption and constructed the exact corrected log-likelihood, so that the corrected score estimator is the solution of $\sum_{i=1}^{n} (Y_i Z_i - E[X_i \exp(X_i\beta)]) = 0$. But the computation of $E[X_i \exp(X_i\beta)]$ is not straightforward even if the distribution of X is known.

In this chapter, we shall propose a more efficient estimation procedure for the Poisson regression with Laplace measurement error. Different from the corrected scores method proposed in Stefanski (1989) and Nakamura (1990), we will derive the probability mass function of Y given the observed data Z. The computation is made possible by the construction of the Tweedie-type formula established in Shi and Song (2015). The proposed estimator is defined as the maximizer of the likelihood function based on the probability mass function of (Y, Z), hence is more efficient than the existing methods proposed in literature.

This chapter is organized as follows. The proposed calibration likelihood estimator, when the density function of X is known and unknown, will be proposed in Section 3.2. Large sample results, including the consistency and the asymptotic normality, of the proposed estimator will be thoroughly investigated in Section 3.3. Simulation studies will be conducted in Section 3.4. Some further discussions, including some future remarks, more simulation studies to compare our two proposed methods, as well as sensitivity analysis will be mentioned in Section 3.5. All the technical proofs will be deferred to Section 3.6.

3.2 Calibrating Poisson Regression Likelihood

We begin with the scenarios in which the density function f of X is known. It is noted that the assumption of known f is rather strict, but it is not rare in real applications. Such examples are abundant in econometrics, nutrition studies and biology literatures. Most importantly, the derivation of the calibrated likelihood estimation procedure based on the known f can help us to construct a semi-parametric estimation procedure for cases where the density function f is unknown.

To be specific, by the nondifferential condition, for any nonnegative integer y, we have

$$\begin{aligned} P(Y = y|Z) &= E[I(Y = y)|Z] = E([I(Y = y)|X, Z]|Z) = E([I(Y = y)|X]|Z) \\ &= E\left(\frac{\lambda(X)^y}{y!}\exp(-\lambda(X))\middle| Z\right) = \frac{1}{y!}E[\exp(yX\beta - \exp(X\beta))|Z]. \end{aligned}$$

The probability mass function $P(Y = y|Z = z)$ will be denoted as $p(y|z, \beta)$. For any twice differentiable function $m(x)$, Shi and Song (2015) showed that if the density function g of Z is also twice continuously differentiable, and U has a Laplace distribution, we have the following Tweedie-type formula

$$E[m(X)|Z] = m(Z) + \frac{1}{g(Z)}\int_z^\infty \left[m'(x) - \frac{\sigma_u m''(x)}{2\sqrt{2}}\right] g(x)\exp\left(\frac{z-x}{\sigma_u/\sqrt{2}}\right)dx$$
$$- \frac{1}{g(Z)}\int_{-\infty}^z \left[m'(x) + \frac{\sigma_u m''(x)}{2\sqrt{2}}\right] g(x)\exp\left(\frac{x-z}{\sigma_u/\sqrt{2}}\right)dx \qquad (3.1)$$

Tweedie formula is originally developed for the conditional expectation $E(X|Z)$ when U follows a normal distribution. As disclosed in Efron (2011), the Tweedie's formula is named after Maurice Kenneth Tweedie and it was first discussed in Robbins (1956). Due to its strong Bayesian flavor, Efron (2011) exclaimed the Tweedie's formula as an "extraordinary Bayesian estimation formula", and a selection bias application of this formula to genomics data was also discussed. However, (3.1) indicates that if U has a Laplace distribution, then a similar Tweedie-type formula can be established for $E(m(X)|X)$, not limited to $E(X|Z)$.

Let $m(x; y, \beta) = \exp(yx\beta - \exp(x\beta))$. Clearly, as a function of x, $m(x; y, \beta)$ is twice differentiable. Define

$$H_-(y, x, \beta) = m'(x; y, \beta) - \frac{\sigma_u m''(x; y, \beta)}{2\sqrt{2}}, \quad H_+(y, x, \beta) = m'(x; y, \beta) + \frac{\sigma_u m''(x; y, \beta)}{2\sqrt{2}}$$

and $G(x, z) = g(x)\exp(-|z - x|/(\sigma_u/\sqrt{2}))$, where m' and m'' denote the first and second derivatives of m with respect to x. Then the log-likelihood function of β, based on a sample $(Z_i, Y_i)_{i=1}^n$ of size n, can be written as

$$L(\beta) = \sum_{i=1}^n \log \frac{1}{Y_i!} E[\exp(Y_i X_i \beta - \exp(X_i\beta))|Z_i]$$
$$\propto \sum_{i=1}^n \log\left[g(Z_i)m(Z_i; Y_i, \beta) + \int_{Z_i}^\infty H_-(Y_i, x, \beta)G(x, Z_i)dx - \int_{-\infty}^{Z_i} H_+(Y_i, x, \beta)G(x, Z_i)dx\right].$$

In fact, for the Poisson regression model discussed in this paper, we have

$$m'(x; y, \beta) = \beta\exp(yx\beta - \exp(x\beta))(y - \exp(x\beta))$$
$$m''(x; y, \beta) = \beta^2\exp(yx\beta - \exp(x\beta))[(y - \exp(x\beta))^2 - \exp(x\beta)].$$

Further denote $\dot{m}'(x; y, \beta), \dot{m}''(x; y, \beta)$ as the derivatives of $m'(x; y, \beta)$, $m''(x; y, \beta)$ with respect to β, respectively. Then we have

$$\begin{aligned}
\dot{m}'(x; y, \beta) = &\ \exp(yx\beta - \exp(x\beta)) \left[x\beta(y - \exp(x\beta))^2 - x\beta\exp(x\beta) + y - \exp(x\beta)\right] \\
\dot{m}''(x; y, \beta) = &\ \exp(yx\beta - \exp(x\beta))[x\beta^2(y - \exp(x\beta))^3 + 2\beta(y - \exp(x\beta))^2 \\
&\quad - 3(y - \exp(x\beta))\exp(x\beta)x\beta^2 - (x\beta + 2)\beta\exp(x\beta)].
\end{aligned}$$

Hence the maximum likelihood estimator of β is the maximizer of $L(\beta)$. Like other likelihood-based estimation procedures, under some regularity conditions, the maximizer of $L(\beta)$ is the solution of the likelihood equation $\partial L(\beta)/\partial\beta = 0$, or

$$\sum_{i=1}^{n} \frac{g(Z_i)\dot{m}(Y_i, Z_i, \beta) + \int_{Z_i}^{\infty} \dot{H}_-(Y_i, x, \beta)G(x, Z_i)dx - \int_{-\infty}^{Z_i} \dot{H}_+(Y_i, x, \beta)G(x, Z_i)dx}{g(Z_i)m(Y_i, Z_i, \beta) + \int_{Z_i}^{\infty} H_-(Y_i, x, \beta)G(x, Z_i)dx - \int_{-\infty}^{Z_i} H_+(Y_i, x, \beta)G(x, Z_i)dx} = 0. \tag{3.2}$$

The solution of (3.2) is denoted as $\tilde{\beta}_n$.

According to Shi and Song (2015), the validity of (3.1) also requires the density function g to be twice continuously differentiable. Note that Z is the convolution of X and U and the Laplace density function is nondifferentiable at the origin, so differentiability of g clearly relies on the smoothness of the density function f of X.

Example 1: Assume that $X \sim N(0, \sigma_x^2)$, and $U \sim$Laplace$(0, \sigma_u^2)$, then

$$g(z) = \frac{1}{2}\exp\left(\frac{\sigma_x^2}{\sigma_u^2}\right)\left[\exp\left(-\frac{z\sqrt{2}}{\sigma_u}\right)\bar{\Phi}_{\sigma_x,\sigma_u}(z) + \exp\left(\frac{z\sqrt{2}}{\sigma_u}\right)\Phi_{\sigma_x,\sigma_u}(z)\right]$$

where

$$\bar{\Phi}_{\sigma_x,\sigma_u}(z) = 1 - \Phi\left(\frac{\sigma_x\sqrt{2}}{\sigma_u} - \frac{z}{\sigma_x}\right), \quad \Phi_{\sigma_x,\sigma_u}(z) = \Phi\left(-\frac{z}{\sigma_x} - \frac{\sigma_x\sqrt{2}}{\sigma_u}\right),$$

and Φ is the CDF of the standard normal distribution. Clearly, $g(z)$ is twice differentiable with respect to z.

Example 2: Assume that $X \sim$Uniform$(-a, a)$ for some $a > 0$, and $X \sim$Laplace$(0, \sigma_u^2)$, then

$$g(z) = \frac{e^{\sqrt{2}(z+a)/\sigma_u} - e^{\sqrt{2}(z-a)/\sigma_u}}{4a} I_{(-\infty,-a)}(z) + \frac{2 - e^{-\sqrt{2}(z+a)/\sigma_u} - e^{\sqrt{2}(z-a)/\sigma_u}}{4a} I_{[-a,a]}(z)$$
$$+ \frac{e^{-\sqrt{2}(z-a)/\sigma_u} - e^{-\sqrt{2}(z+a)/\sigma_u}}{4a} I_{(a,\infty)}(z).$$

One can check that $g(z)$ is not twice differentiable with respect to z. In this case, the Tweedie-type formula (3.1) is not applicable. In fact, by direct calculation, one can obtain that

$$E[m(X; y, \beta)|Z] = \frac{1}{2ag(Z)} \int_{-a}^{a} m(x; y, \beta) \frac{1}{2b} \exp\left(-\frac{|x - Z|}{b}\right) dx$$
$$= \frac{e^{-Z/b}}{4abg(Z)} \int_{[-a,a]\cap(-\infty,Z]} m(x; y, \beta) e^{x/b} dx + \frac{e^{Z/b}}{4abg(Z)} \int_{[-a,a]\cap(Z,\infty)} m(x; y, \beta) e^{-x/b} dx.$$

Thus the MLE of β is defined as the solution of the likelihood equation

$$\sum_{i=1}^{n} \frac{e^{-Z_i/b} \int_{[-a,a]\cap(-\infty,Z_i]} \dot{m}(Y_i, x, \beta) e^{x/b} dx + e^{Z_i/b} \int_{[-a,a]\cap(Z_i,\infty)} \dot{m}(Y_i, x, \beta) e^{-x/b} dx}{e^{-Z_i/b} \int_{[-a,a]\cap(-\infty,Z_i)} m(Y_i, x, \beta) e^{x/b} dx + e^{Z_i/b} \int_{[-a,a]\cap(Z_i,\infty)} m(Y_i, x, \beta) e^{-x/b} dx} = 0.$$

In this chapter, we will focus on the large sample properties the estimator of β defined by (3.2). That is, we will assume that the density g is twice continuously differentiable. The discussion on the cases as in Example 2 deserves an independent study.

Rarely is the density function f of X known in practice, neither is the density function g of Z. The observations from Z allow us to construct a nonparametric estimate of $g(z)$, such as the kernel density estimator. From the likelihood equation (3.2), we can see that the kernel density estimator is only necessary for the first term in both the numerator and denominator of the summand, and all the integration terms can be estimated via a more efficient approach. To be specific, note that

$$\int_{Z_i}^{\infty} H_-(Y_i,x,\beta)G(x,Z_i)dx = \int_{Z_i}^{\infty} H_-(Y_i,x,\beta)\exp(-|Z_i-x|/(\sigma_u/\sqrt{2}))g(x)dx$$
$$= E[H_-(Y_i,Z,\beta)\exp(-|Z_i-Z|/(\sigma_u/\sqrt{2}))I_{[Z_i,\infty)}(Z)|Y_i,Z_i],$$

where Z is independent of $\{Y_i,Z_i\}_{i=1}^n$. So the expectation can be estimated by its sample analogues

$$T_{n1}(Y_i,Z_i,\beta) = \frac{1}{n}\sum_{j=1}^{n} H_-(Y_i,Z_j,\beta)\exp\left(-\frac{\sqrt{2}|Z_i-Z_j|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j).$$

Similarly, $\int_{-\infty}^{Z_i} H_+(Y_i,x,\beta)G(x,Z_i)dx$ can be estimated by

$$T_{n2}(Y_i,Z_i,\beta) = \frac{1}{n}\sum_{j=1}^{n} H_+(Y_i,Z_j,\beta)\exp\left(-\frac{\sqrt{2}|Z_i-Z_j|}{\sigma_u}\right) I_{(-\infty,Z_i)}(Z_j).$$

Let $\dot{T}_{n1}$, $\dot{T}_{n2}$ bet the same as T_{n1} and T_{n2} with H_- and H_+ being replace by $\dot{H}_-$ and $\dot{H}_+$, respectively. Thus, the likelihood equation (3.2) can be replaced by

$$\dot{L}_n(\beta) = \sum_{i=1}^{n} \frac{\hat{g}_n(Z_i)\dot{m}(Y_i,Z_i,\beta) + \dot{T}_{n1}(Y_i,Z_i,\beta) - \dot{T}_{n2}(Y_i,Z_i,\beta)}{\hat{g}_n(Z_i)m(Y_i,Z_i,\beta) + T_{n1}(Y_i,Z_i,\beta) - T_{n2}(Y_i,Z_i,\beta)} = 0, \tag{3.3}$$

where $\hat{g}_n(z)$ is the classic Rosenblatt-Parzen kernel density estimator of $g(z)$. The solution of (3.3) is denoted as $\hat{\beta}_n$.

Instead of estimating the integral terms using empirical averages, we may consider estimating g in the integral with the kernel density estimator also. To be specific, choosing K and w as the kernel function and bandwidth, an alternative estimate T_{n1} can be defined as

$$T_{n1}(Y_i,Z_i,\beta) = \frac{1}{nw}\sum_{j=1}^{n}\int_{Z_i}^{\infty} H_-(Y_i,z,\beta)\exp\left(-\frac{\sqrt{2}|Z_i-z|}{\sigma_u}\right) K\left(\frac{z-Z_j}{w}\right)dz.$$

Other quantities can be similarly redefined. The large sample properties of the resulting estimator of β may not be different from the one defined in (3.3), but clearly, the proof will be more complicated.

3.3 Large Sample Results

In this section, we discuss the large sample properties of the estimators $\tilde{\beta}_n$ and $\hat{\beta}_n$ proposed in Section 3.2, for both cases when the density function f of X is known and unknown. Denote the true value of β as β_0. The following list contains the technical assumptions necessary for presenting and proving the relevant large sample results.

(C1). The density function $g(z)$ of Z is twice continuously differentiable, and the second derivative is bounded.

(C2). The parameter space $\Theta = [a, b]$ for some $a < b$, and for all $\beta \in \Theta$, $E \exp(2\beta X) < \infty$; β_0 is an interior point of Θ.

(C3). $L(\beta)$ is differentiable with respect to β, and β_0 is the unique maximizer of $L(\beta)$, where $L(\beta) = E \log Q(Y, Z; \beta), Q(y, z; \beta) = E[\exp(yX\beta - \exp(X\beta))|Z = z]$.

Condition (C1) is used to guarantee that the kernel estimator of g has the desired order in the asymptotic expansion, and the bounded parameter space ensure the application of dominated convergence theorems is legitimate. Condition (C3) implies that, in the neighborhood of β_0, $\dot{L}(\beta) > 0$ when $\beta < \beta_0$ and $\dot{L}(\beta) < 0$ when $\beta > \beta_0$ for all $\beta \in \Theta = [a, b]$.

The following theorem shows that $\tilde{\beta}_n$, when the distribution of X is assumed to be known, is consistent.

Theorem 6. *Suppose the condition (C1), (C2) hold. Then $\tilde{\beta}_n$ converges to β_0 almost surely.*

To state the asymptotic normality of $\tilde{\theta}_n$, we denote

$$K_1(x) = \min(1, |x|, x^2) \exp(-\max(|a|, |b|)|x|), \quad K_2(x) = \max(1, |x|, x^2) \exp(\max(|a|, |b|)|x|).$$

The asymptotic normality of $\tilde{\theta}_n$ is summarized in the following theorem.

Theorem 7. *In addition to (C1) and (C2), we further assume that* $0 < EK(Y,Z) < \infty$ *for* $K(y,z) = y^2 E[K_2^{y+2} e^{-K_1(X)}|Z = z]$, $E[K_1^y(X)e^{-K_2(X)}|Z = z]$. *Then*

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) \Longrightarrow N(0, 1/I(\beta_0)),$$

where $I(\beta_0)$ *is the Fisher information of* β *based on* $p(y|z,\beta)$.

If the density function of Z is unknown, then we estimate β with $\hat{\beta}_n$, which is defined as the solution of $\dot{L}_n(\beta) = 0$, as shown in (3.3). The following theorem shows that $\hat{\beta}_n$ is consistent.

Theorem 8. *In addition to the conditions of Theorem 7, suppose (C3) holds. Then* $\hat{\beta}_n$ *converges to* β_0 *in probability.*

To state the asymptotic normality of $\hat{\beta}_n$, for $i = 1, 2, \ldots, n$, denote $\eta_i = g(Z_i)m(Y_i, Z_i; \beta_0) + T_1(Y_i, Z_i; \beta_0) - T_2(Y_i, Z_i; \beta_0)$, and $\dot{\eta}_i = g(Z_i)\dot{m}(Y_i, Z_i; \beta_0) + \dot{T}_1(Y_i, Z_i; \beta_0) - \dot{T}_2(Y_i, Z_i; \beta_0)$, $m_i = m(Y_i, Z_i; \beta_0)$, $\dot{\eta}_i$ and $\dot{m}_i$ are similarly defined,

$$\begin{aligned}
M_{1i}(\beta_0) &= E\left[\frac{\dot{\eta}_j}{\eta_j^2} H_-(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_j,\infty)}(Z_i) \middle| Y_i, Z_i\right], \\
M_{2i}(\beta_0) &= E\left[\frac{\dot{\eta}_j}{\eta_j^2} H_+(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j) \middle| Y_i, Z_i\right], \\
N_{1i}(\beta_0) &= E\left[\frac{1}{\eta_j} \dot{H}_-(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_j,\infty)}(Z_i) \middle| Y_i, Z_i\right], \\
N_{2i}(\beta_0) &= E\left[\frac{1}{\eta_j} \dot{H}_+(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j) \middle| Y_i, Z_i\right],
\end{aligned}$$

and

$$\begin{aligned}
S_{1i} &= \frac{(\eta_i \dot{m}_i - \dot{\eta}_i m_i) g(Z_i)}{\eta_i^2} - E\frac{(\eta \dot{m} - \dot{\eta} m) g(Z)}{\eta^2}, \\
S_{2i} &= \frac{M_{1i} - M_{2i}}{2} - \frac{\dot{\eta}_i (T_{1i} - T_{2i})}{2\eta_i^2}, \quad S_{3i} = \frac{N_{1i} - N_{2i}}{2} - \frac{\dot{T}_{1i} - \dot{T}_{2i}}{2\eta_i}.
\end{aligned}$$

Then we have

Theorem 9. *Assume that* $\sigma^2 = E[S_{11} - S_{21} + S_{31} + \dot{\eta}/\eta]^2 < \infty$*, then the estimator* $\hat{\beta}_n$ *is asymptotically normal:* $\sqrt{n}(\hat{\beta}_n - \beta_0) \Longrightarrow (0, \sigma^2/I^2(\beta_0))$.

From Theorem 9, one can see that the effect of measurement error is reflected by the extra term σ^2 comparing to the result in Theorem 7, which reduces to $I(\theta_0)$ if $\sigma_u^2 = 0$.

3.4 Simulation Studies

In this section, some simulation studies are conducted to evaluate the finite sample performance of the proposed estimate $\hat{\beta}_n$. We considers both situations in which the distribution of X is assumed to be known and unknown.

3.4.1 Parametric Case

In this case, X is generated from the standard normal distribution. The true value of β_0 is chosen to be 1. To see the effect of measurement error on the estimator, the scale parameter b is chosen to be $0.2, 0.5, 0.8$ and 1. Note that the variance $\sigma_u^2 = 2b^2$, and four sample sizes, $n = 100, 200, 300, 500$ are considered. For each setup, the simulation is repeated 500 times, and the mean, bias, and mean squared errors (MSEs) are used as the criteria to evaluate the finite sample performance of the estimation procedures. For comparison purpose, we also calculate the estimator proposed in Guo and Li (2002), denoted as GL in Table 3.1.

From Table 3.1, it can be seen that for both estimators, when the scale parameter b is fixed, the MSEs and biases get smaller when the sample size gets larger. While for fixed sample size n, when b or σ_u^2 is small, for example, $b = 0.2$, the proposed method has slightly higher MSE and bias than Guo and Li (2002)'s method. However, as b or σ_u^2 increases, our method outperforms Guo and Li (2002)'s method. The reason that why the proposed estimator is inferior to GL estimator for smaller b values needs further investigation.

		GL			$\hat{\beta}_n$		
b	n	Mean	MSE	Bias	Mean	MSE	Bias
0.2	100	0.9637	0.0358	-0.0363	0.9556	0.0387	-0.0444
	200	0.9856	0.0159	-0.0144	0.9713	0.0329	-0.0287
	300	0.9887	0.0103	-0.0113	0.9826	0.0305	-0.0174
	500	0.9919	0.0066	-0.0081	0.9847	0.0248	-0.0153
0.5	100	0.9606	0.0408	-0.0394	0.9711	0.0189	-0.0289
	200	0.9729	0.0176	-0.0271	0.9892	0.0074	-0.0108
	300	0.9869	0.0141	-0.0131	0.9932	0.0077	-0.0068
	500	0.9938	0.0076	-0.0062	1.0006	0.0040	0.0006
0.8	100	0.9541	0.0510	-0.0459	0.9781	0.0180	-0.0219
	200	0.9821	0.0204	-0.0179	0.9908	0.0106	-0.0092
	300	0.9849	0.0128	-0.0151	0.9935	0.0083	-0.0065
	500	0.9874	0.0101	-0.0126	0.9951	0.0057	-0.0049
1	100	0.9480	0.0609	-0.0520	0.9786	0.0157	-0.0214
	200	0.9795	0.0236	-0.0205	0.9886	0.0130	-0.0114
	300	0.9840	0.0148	-0.0160	0.9904	0.0100	-0.0096
	500	0.9970	0.0111	-0.0030	0.9987	0.0074	-0.0013

Table 3.1: *Estimates when the distribution of X is known*

3.4.2 Semi-parametric Case

When the distribution of X is unknown, the estimator in Guo and Li (2002) is no longer applicable since $E\exp(X\beta)$ would not be available. For comparison purpose, the proposed estimator will be compared with the estimation method that combines Guo and Li (2002) and Hong and Tamer (2003)'s approach as mentioned in 3.1, denoted as GLHT in Table 3.2.

To estimate the density function of Z, we choose the standard normal density function as the kernel function. For the bandwidth of the kernel, we choose $n^{-1/5}$ which is the optimal order of the bandwidth in kernel density estimation at which the kernel density estimate achieves optimal convergence rate in the MSE sense. Other settings are the same as in Section 3.4.1. The simulation results are summarized in Table 3.2.

From Table 3.2, one can see for fixed sample size, when b or σ_u^2 is small, the performance of the proposed method is similar to the method combining Guo and Li (2002) and Hong and Tamer (2003)'s approach. However, as b increase, the MSEs and biases of the combined method are much higher than the proposed method. This result is consistent with our discussion in Section 1, that the modified likelihood function of the combined method is indeed not bounded.

		GLHT			$\hat{\beta}_n$		
b	n	Mean	MSE	Bias	Mean	MSE	Bias
0.2	100	0.9940	0.0091	-0.0060	1.0430	0.0345	0.0430
	200	0.9887	0.0175	-0.0113	1.0329	0.0287	0.0329
	300	0.9023	0.1039	-0.0977	1.0242	0.0119	0.0242
	500	0.9788	0.0214	-0.0212	1.0161	0.0107	0.0161
0.5	100	1.5655	0.6195	0.5655	1.1111	0.1610	0.1111
	200	1.5370	0.5751	0.5369	1.1053	0.1383	0.1053
	300	1.4701	0.5183	0.4701	1.1531	0.1436	0.1531
	500	1.3280	0.4426	0.3280	1.1179	0.1227	0.1179
0.8	100	1.9861	0.9964	0.9861	1.0533	0.2250	0.0533
	200	1.9861	0.9964	0.9861	1.1068	0.2125	0.1068
	300	1.9967	0.9987	0.9967	1.1275	0.2015	0.1275
	500	1.9907	0.9957	0.9907	1.1230	0.1800	0.1230
1	100	2.0000	0.9999	1.0000	1.0441	0.2640	0.0441
	200	2.0000	0.9999	1.0000	1.1070	0.2713	0.1070
	300	2.0000	0.9999	1.0000	1.0941	0.2404	0.0941
	500	2.0000	0.9999	1.0000	1.0646	0.2141	0.0646

Table 3.2: *Estimates when the distribution of X is unknown*

3.5 Discussion

3.5.1 Future Remarks

As we mentioned before, the validity of (3.1) requires the density function g to be twice continuously differentiable, which may not be held, as the example in Example 2 on uniform X suggested. In this case, a direct estimate of the density function of X can be used. By using the standard normal kernel, the deconvolution kernel density estimator of f is defined as

$$\hat{f}_n(x) = \frac{1}{nw}\sum_{j=1}^{n} L_n\left(\frac{x - Z_j}{w}\right), \qquad \text{where} \qquad L_n(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\left[1 - \frac{\sigma_u^2}{2w^2}(x^2 - 1)\right].$$

Note that $E[m(y, X, \beta)|Z] = g^{-1}(Z)\int m(y, x, \beta)f(x)h(Z - x)dx$, so we can estimate the above expectation, denoted as $M(y, Z, \beta)$ by

$$\hat{M}(y, Z, \beta) = \frac{1}{nw\hat{g}_n(Z)}\sum_{j=1}^{n}\int m(y, x, \beta)L_n\left(\frac{x - Z_j}{w}\right)\frac{1}{2b}\exp\left(-\frac{|x - Z|}{b}\right)dx.$$

Therefore, an estimate of β can be obtained by solving the following estimated likelihood equation

$$\sum_{i=1}^{n}\frac{\dot{\hat{M}}(Y_i, Z_i, \beta)}{\hat{M}(Y_i, Z_i, \beta)} = \sum_{i=1}^{n}\frac{\sum_{j=1}^{n}\int \dot{m}(Y_i, x, \beta)L_n\left(\frac{x - Z_j}{w}\right)\exp\left(-\frac{|x - Z_i|}{b}\right)dx}{\sum_{j=1}^{n}\int m(Y_i, x, \beta)L_n\left(\frac{x - Z_j}{w}\right)\exp\left(-\frac{|x - Z_i|}{b}\right)dx} = 0. \qquad (3.4)$$

The large sample properties of such an estimator deserves an independent study.

Also, note that in this chapter, we only considered a simple case where the explanatory variable X is univariate. It is well known that the definition of the multivariate Laplace distribution is not unique. Although we have developed some Tweedie-type formulae for two multivariate Laplace distributions in Shi and Song (2015) and Song et al. (2021), but these Tweedie-type formulae have complicated forms and seem too hard to apply to the Poisson regression with multivariate X variables. We will continue the exploration for simpler versions so that it can be used for constructing more efficient estimation procedures.

3.5.2 Simulation Studies to Compare WLS and MLE

Besides, some simulation studies are also conducted to evaluate the finite sample performance of our proposed weighted least squares estimator (WLS) in chapter 2 and maximum likelihood estimator (MLE) in this chapter, when the distribution of X is unknown.

Similar to the settings in Section 2.4, X is generated from the standard normal distribution. The true value of β_0 is chosen to be 1. And to see the effect of measurement error on the estimator, the variance σ_u^2 is chosen to be $1.25, 1, 0.75, 0.5, 0.25, 0.1$. And four sample sizes, $n = 200, 300, 500, 1000$ are considered. The bandwidth of the kernel density estimator is chosen to be $n^{-1/5}$ and $10n^{-1/5}$. For each setup, the simulation is repeated 500 times, and the means, bias, variances and mean squared errors (MSEs) are used as the criteria to evaluate the finite sample performance of the proposed estimation procedures. The simulation results are summarized in Table 3.3 and Table 3.4. Our proposed weighted least square estimator and maximum likelihood estimator are denoted by $\hat{\beta}_{WLS}$ and $\hat{\beta}_{MLE}$ respectively.

From Table 3.3, it can be seen that when the variance is small, $\hat{\beta}_{MLE}$ has slightly bigger biases and variances compared with $\hat{\beta}_{WLS}$. While as σ_u^2 increases, $\hat{\beta}_{MLE}$ has relatively smaller biases but still bigger variances.

From Table 3.4, one can see that for fixed small variance σ_u^2, $\hat{\beta}_{MLE}$ tends to have smaller biases than $\hat{\beta}_{WLS}$ when the sample size increases. While for fixed big variance, $\hat{\beta}_{MLE}$ has smaller biases than $\hat{\beta}_{WLS}$ under both small and big sample size. However, we do realize that under both small and big variances, the variances of $\hat{\beta}_{MLE}$ is bigger than $\hat{\beta}_{WLS}$, which needs further investigation.

3.5.3 Sensitivity Analysis

To evaluate the performance of the two proposed estimators in practices, we considered the dataset from the website https://www.theanalysisfactor.com/generalized-linear-models-in-r-part-6-poisson-regression-count-variables/. The dataset consists of 109 observations on the counts of high school students diagnosed with an infectious disease within a period of days from an initial outbreak. Figure 3.1 is the scatterplot of counts of students (y) versus days from the initial outbreak (x).

A sensitivity analysis is conducted in this section. Some Laplace errors are added to the covariate x. Since the variance of x is 1202.911, the variance of measurement error σ_u^2 is chosen to be 1 to 351 with equal step 50, from very small error contamination to moderately

		$\hat{\beta}_{WLS}$				$\hat{\beta}_{MLE}$			
σ_u^2	n	Mean	Bias	Variance	MSE	Mean	Bias	Variance	MSE
1.25	200	0.625	-0.375	0.004	0.145	0.680	-0.320	0.425	0.527
	300	0.652	-0.348	0.003	0.125	0.720	-0.280	0.413	0.491
	500	0.676	-0.324	0.003	0.108	0.757	-0.243	0.376	0.435
	1000	0.706	-0.294	0.002	0.089	0.778	-0.222	0.358	0.407
1	200	0.669	-0.331	0.005	0.115	0.702	-0.298	0.405	0.494
	300	0.700	-0.300	0.004	0.094	0.743	-0.257	0.379	0.445
	500	0.728	-0.272	0.004	0.078	0.716	-0.284	0.344	0.424
	1000	0.768	-0.232	0.002	0.056	0.853	-0.147	0.333	0.355
0.75	200	0.741	-0.259	0.006	0.074	0.776	-0.224	0.357	0.407
	300	0.771	-0.229	0.004	0.057	0.786	-0.214	0.370	0.416
	500	0.801	-0.199	0.004	0.044	0.778	-0.222	0.373	0.423
	1000	0.837	-0.163	0.003	0.030	0.879	-0.121	0.293	0.308
0.5	200	0.825	-0.175	0.008	0.038	0.735	-0.265	0.359	0.429
	300	0.860	-0.140	0.006	0.026	0.836	-0.164	0.361	0.388
	500	0.885	-0.115	0.004	0.017	0.838	-0.162	0.301	0.327
	1000	0.919	-0.081	0.003	0.009	0.900	-0.100	0.246	0.255
0.25	200	0.915	-0.085	0.009	0.016	0.907	-0.093	0.231	0.240
	300	0.944	-0.056	0.005	0.008	0.885	-0.115	0.222	0.235
	500	0.963	-0.037	0.003	0.004	0.938	-0.062	0.199	0.203
	1000	0.979	-0.021	0.001	0.002	0.955	-0.045	0.121	0.123
0.1	200	0.977	-0.023	0.006	0.007	0.971	-0.029	0.077	0.078
	300	0.987	-0.013	0.004	0.004	0.979	-0.021	0.070	0.071
	500	0.996	-0.004	0.002	0.002	1.004	0.004	0.031	0.032
	1000	1.000	0.000	0.001	0.001	0.994	-0.006	0.019	0.019

Table 3.3: $\hat{\beta}_{WLS}$ *and* $\hat{\beta}_{MLE}$ *when* $h = n^{-1/5}$

σ_u^2	n	$\hat{\beta}_{WLS}$				$\hat{\beta}_{MLE}$			
		Mean	Bias	Variance	MSE	Mean	Bias	Variance	MSE
1.25	200	0.532	-0.468	0.018	0.237	0.549	-0.451	0.428	0.632
	300	0.519	-0.481	0.010	0.242	0.613	-0.387	0.434	0.583
	500	0.526	-0.474	0.006	0.231	0.667	-0.333	0.429	0.540
	1000	0.538	-0.462	0.003	0.217	0.676	-0.324	0.427	0.532
1	200	0.566	-0.434	0.019	0.208	0.588	-0.412	0.412	0.582
	300	0.562	-0.438	0.014	0.206	0.662	-0.338	0.438	0.552
	500	0.568	-0.432	0.009	0.196	0.665	-0.335	0.427	0.539
	1000	0.580	-0.420	0.003	0.180	0.681	-0.319	0.398	0.500
0.75	200	0.644	-0.356	0.024	0.150	0.648	-0.352	0.457	0.580
	300	0.631	-0.369	0.015	0.152	0.664	-0.336	0.420	0.533
	500	0.636	-0.364	0.011	0.143	0.620	-0.380	0.411	0.556
	1000	0.628	-0.372	0.004	0.143	0.702	-0.298	0.365	0.454
0.5	200	0.751	-0.249	0.024	0.086	0.646	-0.354	0.471	0.596
	300	0.726	-0.274	0.017	0.092	0.648	-0.352	0.400	0.524
	500	0.718	-0.282	0.011	0.091	0.740	-0.260	0.359	0.426
	1000	0.715	-0.285	0.006	0.088	0.737	-0.263	0.295	0.364
0.25	200	0.860	-0.140	0.018	0.038	0.736	-0.264	0.370	0.440
	300	0.850	-0.150	0.015	0.037	0.748	-0.252	0.369	0.432
	500	0.838	-0.162	0.010	0.036	0.782	-0.218	0.277	0.325
	1000	0.843	-0.157	0.006	0.030	0.871	-0.129	0.177	0.193
0.1	200	0.961	-0.039	0.010	0.012	0.881	-0.119	0.200	0.214
	300	0.943	-0.057	0.007	0.010	0.927	-0.073	0.125	0.130
	500	0.944	-0.056	0.005	0.008	0.968	-0.032	0.046	0.047
	1000	0.935	-0.065	0.003	0.008	0.967	-0.033	0.024	0.025

Table 3.4: $\hat{\beta}_{WLS}$ *and* $\hat{\beta}_{MLE}$ *when* $h = 10n^{-1/5}$

large noises. The bandwidth of the kernel density estimator is chosen to be $n^{-1/5}$. For convenience, the weight function of our weighted least square estimator is chosen to be $e^{-z^2/2}$. To conduct the sensitivity analysis, the intercept is included in the model, i.e., the intensity function $\lambda(x)$ of the Poisson regression model is chosen to be $\exp(\beta_0+\beta_1 x)$, instead of $\exp(\beta x)$. For each different σ_u^2, the simulation is repeated 500 times, and the means of β_0 and β_1 are calculated. The results are summarized in Table 3.5.

Note that the estimates of β_0 and β_1 are 1.9902 and -0.0175 based on the data from (Y, X), respectively. From Table 3.5, it could be seen that when the variance of measurement error σ_u^2 is small, or when there is almost no measurement error, the performance of our proposed maximum likelihood estimator (MLE) is similar to the naive estimator and is slightly better than our proposed weighted least square estimator (WLS). While as the variance increases, our MLE estimator outperforms the naive estimator with respect to both β_0 and β_1, and our WLS estimator outperforms the naive estimator with respect to β_0 and performs similar with the naive estimator with respect to β_1.

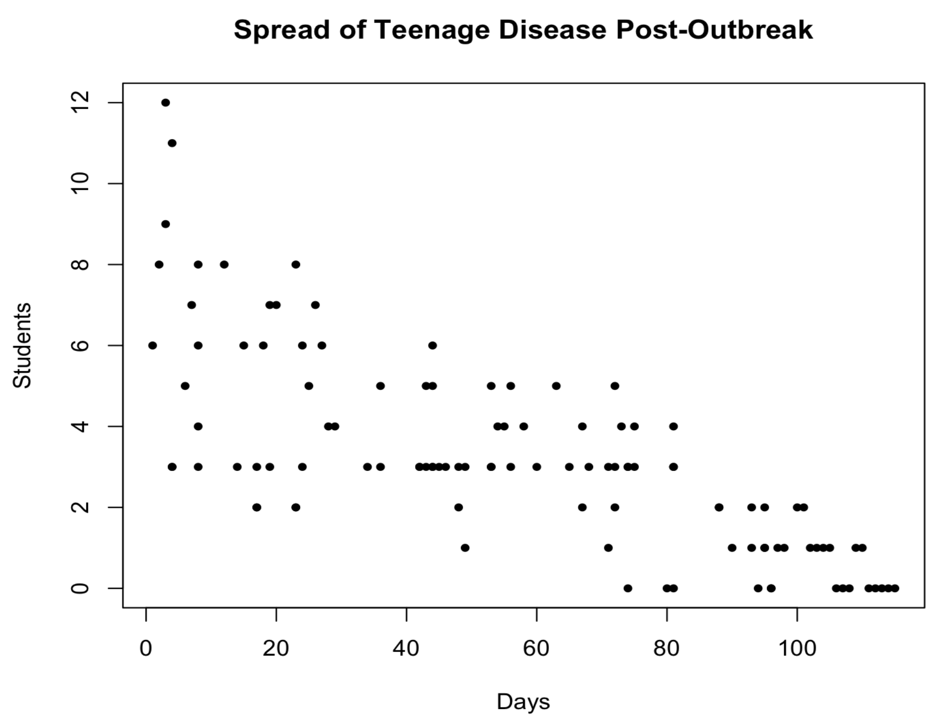

Figure 3.1: *Scatterplot*

	$\hat{\beta}_{Naive}$		$\hat{\beta}_{MLE}$		$\hat{\beta}_{WLS}$	
σ_u^2	β_0	β_1	β_0	β_1	β_0	β_1
1	1.9894	-0.0174	1.9809	-0.0175	1.7404	0.0997
51	1.9527	-0.0166	1.9502	-0.0168	1.9000	-0.0283
101	1.9157	-0.0159	1.9186	-0.0162	1.9222	-0.0323
151	1.8852	-0.0152	1.8926	-0.0156	1.8675	-0.0267
201	1.8581	-0.0146	1.8705	-0.0151	1.8718	-0.0268
251	1.8318	-0.0141	1.8473	-0.0147	1.8673	-0.0284
301	1.8084	-0.0136	1.8278	-0.0142	1.8304	-0.0247
351	1.7848	-0.0130	1.8076	-0.0138	1.8472	-0.0242

Table 3.5: *$\hat{\beta}_{Naive}$, $\hat{\beta}_{MLE}$ and $\hat{\beta}_{WLS}$ when $h = n^{-1/5}$*

3.6 Appendix: Proof of Main Results

The proof the consistency of $\tilde{\beta}_n$ is facilitated by a lemma whose proof can be found in Ferguson (2017). For the sake of completeness, the lemma is reproduced below.

Lemma 10. *Let $X_1, X_2, \dots$ be i.i.d. with density $p(x;\theta)$, $\theta \in \Theta$, and let θ_0 denote the true value of θ. If*

(F1). Θ is compact,

(F2). $p(x;\theta)$ is upper semicontinuous in θ for all x,

(F3). there exists a function $K(x)$ such that $E_{\theta_0}|K(X)| < \infty$ and

$$U(x;\theta) = \log p(x;\theta) - \log p(x;\theta_0) \le K(x), \quad \text{for all } x \text{ and } \theta,$$

(F4). for all $\theta \in \Theta$ and sufficiently small $\rho > 0$, $\sup_{|\theta'-\theta|<\rho} p(x;\theta')$ is measurable in x,

(F5). $p(x;\theta) = p(x;\theta_0)$ a.e. implies $\theta = \theta_0$,

then, for any sequence of MLEs $\hat{\theta}_n$ of θ_0, $\hat{\theta}_n \to \theta_0$ almost surely.

The proof of Theorem 6. Applying Lemma 10 to

$$p(y,z;\beta) = (y!)^{-1} E[\lambda^y(X,\beta)\exp(-\lambda(X,\beta))|Z = z]$$

with (y, z) being treated as the variable x. It is easy to see that (F1), (F2) and (F4) hold in the current setup, so it suffices to show that (F3) and (F5) also hold.

To show the identifiability condition (F3) in Lemma 10, we consider the difference

$$\left|\log \frac{1}{y!}E[\lambda^y(X,\beta)\exp(-\lambda(X,\beta))|Z] - \log \frac{1}{y!}E[\lambda^y(X,\beta_0)\exp(-\lambda(X,\beta_0))|Z]\right|.$$

By the triangular inequality, the difference is bounded above by

$$|\log E[\lambda^y(X,\beta)\exp(-\lambda(X,\beta))|Z]| + |\log E[\lambda^y(X,\beta_0)\exp(-\lambda(X,\beta_0))|Z]|.$$

Since for any $\lambda > 0$, $\lambda^y \exp(-\lambda) \le y^y \exp(-y)$, so the difference is further bounded above by $2|\log(y^y \exp(-y))| \le 2y|\log y| + 2y$. Thus, (F3) is satisfied because of $E[Y|\log Y| + Y] < \infty$ under the condition of $EY^2 < \infty$. In fact, $EY^2 = E[E[Y^2|X]] = E[\exp(\beta_0 X) + \exp(2\beta_0 X)] < \infty$ implies $E[\exp(2\beta_0 X)] < \infty$, which is finite by the condition (C2).

Let β_1, $\beta_2 \in \Theta$, and we assume that

$$E[\lambda^y(X,\beta_1)\exp(-\lambda(X,\beta_1))|Z = z] = E[\lambda^y(X,\beta_2)\exp(-\lambda(X,\beta_2))|Z = z] \tag{3.5}$$

for almost every z. Note that for any $\beta \in \Theta$,

$$\begin{aligned}&\int \lambda^y(x,\beta)\exp(-\lambda(x,\beta))f_{X|Z}(x,z)dx \\ &= g^{-1}(z)\int \lambda^y(x,\beta)\exp(-\lambda(x,\beta))f_X(x)\frac{1}{\sqrt{2}\sigma_u}\exp\left(-\frac{\sqrt{2}|z-x|}{\sigma_u}\right)dx,\end{aligned}$$

and, as a location family, the Laplace distribution is complete, so (3.5) implies

$$\lambda^y(x,\beta_1)\exp(-\lambda(x,\beta_1)) = \lambda^y(x,\beta_2)\exp(-\lambda(x,\beta_2))$$

for all x, z and y. Recall that $\lambda(x,\beta) = \exp(x\beta)$, the above equality implies $\beta_1 = \beta_2$. Therefore, the identifiability condition (F4) holds in Lemma 10. This concludes the proof of

Theorem 6. □

To show the asymptotic normality of the proposed MLE $\hat{\beta}_n$, we need the classical Cramér's theorem. For the sake of completeness, the Cramér's theorem is reproduced here.

Lemma 11. *Let $X_1, X_2, \dots$ be i.i.d. with density $p(x;\beta)$, and let β_0 denote the true value of the parameter. If*

(N1). Θ is an open subset of $\mathbb{R}$;

(N2). Second partial derivatives of $p(x;\beta)$ with respect to β exist and are continuous for all x, and may be passed under the integral sign in $\int p(x;\beta)dx$;

(N3). There exists a function $K(x)$ such that $E_{\beta_0}K(X) < \infty$ and each component of $\dot{\Psi}(x;\beta) = \partial^2 \log p(x;\beta)/\partial\beta^2$ is bounded in absolute value by $K(x)$ uniformly in some neighborhood of β_0;

(N4). $I(\beta_0) = -E_{\beta_0}(\dot{\Psi}(x;\beta_0))$ is positive definite;

(N5). $p(x;\beta) = p(x;\beta_0)$ a.e. implies $\beta = \beta_0$.

Then there exists a strongly consistent sequence $\hat{\beta}_n$ of roots of the likelihood equation such that $\sqrt{n}(\hat{\beta}_n - \beta_0) \Longrightarrow N(0, I^{-1}(\beta_0))$.

The proof of Theorem 7. We begin with by checking the condition (N2) in Lemma 11. That is, the second partial derivative of $p(y,z;\beta)$ with respect to β exists and may be passed under the sum sign in $\sum_y \int p(y,z;\beta)dz$. Note that

$$\sum_{y=0}^{\infty} \int \frac{\partial}{\partial\beta} p(y,z;\beta)dz = \sum_{y=0}^{\infty} \frac{1}{y!} \int \frac{\partial}{\partial\beta} E[\lambda^y(X,\beta)\exp(-\lambda(X,\beta))|Z=z]g(z)dz,$$

and

$$\begin{aligned}
&\frac{\partial}{\partial\beta}(\lambda^y(x,\beta)\exp(-\lambda(x,\beta))) \\
&= y\lambda^{y-1}(x,\beta)\lambda'(x,\beta)\exp(-\lambda(x,\beta)) - \lambda^y(x,\beta)\exp(-\lambda(x,\beta))\lambda'(x,\beta).
\end{aligned}$$

Recall that the parameter space of β is a closed interval $\Theta = [a, b], a, b \in \mathbb{R}$, this implies for the functions $K_1(x)$ and $K_2(x)$ defined in Section 3.3, we have $0 \leqslant K_1(x) \leqslant \lambda(x, \beta) \leqslant K_2(x)$, $0 \leqslant K_1(x) \leqslant \lambda'(x, \beta) \leqslant K_2(x)$, then

$$\begin{aligned}\left|\frac{\partial}{\partial \beta}(\lambda^y(x, \beta)) \exp(-\lambda(x, \beta))\right| &\leq yK_2^{y-1}(x)K_2(x) \exp(-K_1(x)) + K_2^{y+1}(x) \exp(-K_1(x)) \\ &= yK_2^y(x) \exp(-K_1(x)) + K_2^{y+1}(x) \exp(-K_1(x)) \\ &\leq (y+1)K_2^{y+1}(x) \exp(-K_1(x)).\end{aligned}$$

This, together with the condition $E[K_2^{y+1}(X) \exp(-K_1(X))] < \infty$ and Fubini's Theorem, we have

$$\begin{aligned}&\sum_{y=0}^{\infty} \int \frac{\partial}{\partial \beta} p(y, z; \beta) dz = \sum_{y=0}^{\infty} \frac{1}{y!} \int E\left[\frac{\partial}{\partial \beta} \lambda^y(X, \beta) \exp(-\lambda(X, \beta)) | z\right] g(z) dz \\ &= \sum_{y=0}^{\infty} \frac{1}{y!} \int E[(y\lambda^{y-1}(X, \beta) \exp(-\lambda(X, \beta)) - \lambda^y(X, \beta) \exp(-\lambda(X, \beta)))\lambda'(X, \beta)|z] g(z) dz \\ &= \int E\left[\sum_{y=0}^{\infty} \frac{1}{y!}[y\lambda^{y-1}(X, \beta) \exp(-\lambda(X, \beta)) - \lambda^y(X, \beta) \exp(-\lambda(X, \beta))]\lambda'(X, \beta)|z\right] g(z) dz \\ &= \int E[(1-1)\lambda'(X, \beta)|z] g(z) dz = 0.\end{aligned}$$

Similarly, from the fact that $0 \leqslant K_1(X) \leqslant |\lambda''(X, \beta)| \leqslant K_2(X)$, we can show that the second partial derivatives can also be passed under the integral sign. Thus, the condition (N2) in Cramer's theorem holds.

Finally, let us show that each component of $\dot{\Psi}(y, z, \beta)$ is bounded in absolute value by $K(y, z)$ uniformly in some neighborhood of β_0. Note that

$$\begin{aligned}\dot{\Psi}(y, z, \beta) &= \frac{\partial^2}{\partial \beta^2} \log p(y, z; \beta) = \frac{\partial^2}{\partial \beta^2} \log E[\lambda^y(X, \beta) \exp(-\lambda(X, \beta))|Z = z] \\ &= \frac{\partial}{\partial \beta} \frac{E[y\lambda^{y-1}(X, \beta) \exp(-\lambda(X, \beta))\lambda'(X, \beta) - \lambda^y(X, \beta) \exp(-\lambda(X, \beta))\lambda'(X, \beta)|Z]}{E[\lambda^y(X, \beta) \exp(-\lambda(X, \beta))|Z = z]} \\ &= \frac{M(y, z)}{(E[\lambda^y(X, \beta) \exp(-\lambda(X, \beta))|Z = z])^2},\end{aligned}$$

where $M(y,z)$ can be written as

$$
\begin{aligned}
&E[y(y-1)\lambda^{y-2}(X,\beta)\exp(-\lambda(X,\beta))(\lambda'(X,\beta))^2\\
&-y\lambda^{y-1}(X,\beta)\exp(-\lambda(X,\beta))(\lambda'(X,\beta))^2\\
&+y\lambda^{y-1}(X,\beta)\exp(-\lambda(X,\beta))\lambda''(X,\beta)-y\lambda^{y-1}(X,\beta)\exp(-\lambda(X,\beta))(\lambda'(X,\beta))^2\\
&+\lambda^{y}(X,\beta)\exp(-\lambda(X,\beta))(\lambda'(X,\beta))^2\\
&-\lambda^{y}(X,\beta)\exp(-\lambda(X,\beta))\lambda''(X,\beta)|Z=z]E[\lambda^{y}(X,\beta)\exp(-\lambda(X,\beta))|Z=z]\\
&-(E[y\lambda^{y-1}(X,\beta)\exp(-\lambda(X,\beta))\lambda'(X,\beta)-\lambda^{y}(X,\beta)\exp(-\lambda(X,\beta))\lambda'(X,\beta)|Z=z])^2
\end{aligned}
$$

Note that $E[\lambda^y(X,\beta)\exp(-\lambda(X,\beta))|Z] \geqslant E[K_1^y(X)\exp(-K_2(X))|Z]$, then from the condition set in the theorem, we can show that $|\dot{\Psi}(y,z,\beta)|$ is bounded above by a function of (y,z) whose expectation under the joint distribution of (Y,Z) is finite. Hence, by Crámer's Theorem, we conclude that $\sqrt{n}(\hat{\beta}_n-\beta_0) \overset{\mathcal{L}}{\Longrightarrow} N(0,I^{-1}(\beta_0))$, where $I(\beta)$ is the Fisher information number $I(\beta)=E\left[\partial \log p(Y,Z;\beta)/\partial\beta\right]^2$ which can be written as

$$
E\left[\frac{g(Z)\dot{m}(Y,Z,\beta)+\int_Z^\infty \dot{H}_-(Y,x,\beta)G(x,Z)dx-\int_{-\infty}^Z \dot{H}_+(Y,x,\beta)G(x,Z)dx}{g(Z)m(Y,Z,\beta)+\int_Z^\infty H_-(Y,x,\beta)G(x,Z)dx-\int_{-\infty}^Z H_+(Y,x,\beta)G(x,Z)dx}\right]^2.
$$

□

Now, let us proceed to the proof of the consistency of $\hat{\beta}_n$.

The proof of Theorem 8. . To show the consistency of $\hat{\beta}_n$, it suffices to show that

$$
\sup_{\beta\in\Theta}|\dot{L}_n(\beta)-\dot{L}(\beta)|=o_p(1). \tag{3.6}
$$

Indeed, (3.6) implies that $\dot{L}_n(\hat{\beta}_n)-\dot{L}(\hat{\beta}_n)=o_p(1)$, which further implies that $\dot{L}(\hat{\beta}_n)=o_p(1)$. Since β_0 is the unique solution of $\dot{L}(\beta)=0$, so $\hat{\beta}_n\to\beta_0$ in probability. Note that

$\dot{L}_n(\beta) - \dot{L}(\beta) = \dot{L}_n(\beta) - \dot{\tilde{L}}_n(\beta) + \dot{\tilde{L}}_n(\beta) - \dot{L}(\beta)$, so it suffices to show that

$$\sup_{\beta\in\Theta} |\dot{L}_n(\beta) - \dot{\tilde{L}}_n(\beta)| = o_p(1), \quad \sup_{\beta\in\Theta} |\dot{\tilde{L}}_n(\beta) - \dot{L}(\beta)| = o_p(1), \tag{3.7}$$

where

$$\dot{\tilde{L}}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{g_i \dot{m}_i(\beta) + \dot{T}_{1i}(\beta) - \dot{T}_{2i}(\beta)}{g_i m_i(\beta) + T_{1i}(\beta) - T_{2i}(\beta)}.$$

For brevity, denote

$$\begin{aligned}
T_{1i}(\beta) &= E[T_{n1}(Y_i, Z_i, \beta_0)|Y_i, Z_i] = \int_{Z_i}^{\infty} H_-(Y_i, x, \beta)G(x, Z_i)dx \\
T_{2i}(\beta) &= E[T_{n2}(Y_i, Z_i, \beta_0)|Y_i, Z_i] = \int_{-\infty}^{Z_i} H_+(Y_i, x, \beta)G(x, Z_i)dx,
\end{aligned}$$

$\hat{g}_{ni} = \hat{g}_n(Z_i)$, $T_{n1i}(\beta) = T_{n1}(Y_i, Z_i, \beta)$, $\eta_i(\beta) = g_i m(Z_i, \beta) + T_{1i}(\beta) - T_{2i}(\beta)$, $\Delta T_{n1i}(\beta) = T_{n1}(Y_i, Z_i, \beta) - T_{1i}(\beta), \Delta T_{n2i}(\beta) = T_{n2}(Y_i, Z_i, \beta) - T_{2i}(\beta)$, and other quantities are similarly defined. Note that

$$\begin{aligned}
&\dot{L}_n(\beta) - \dot{\tilde{L}}_n(\beta) \\
= &\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}_{ni}\dot{m}_i(\beta) + \dot{T}_{n1i}(\beta) - \dot{T}_{n2i}(\beta)}{\hat{g}_{ni} m_i(\beta) + T_{n1i}(\beta) - T_{n2i}(\beta)} - \frac{1}{n}\sum_{i=1}^{n} \frac{g_i\dot{m}_i(\beta) + \dot{T}_{1i}(\beta) - \dot{T}_{2i}(\beta)}{g_i m_i(\beta) + T_{1i}(\beta) - T_{2i}(\beta)} \\
= &\frac{1}{n}\sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)\dot{m}_i(\beta) + \Delta\dot{T}_{n1i}(\beta) - \Delta\dot{T}_{n2i}(\beta)}{(\hat{g}_{ni} - g_i)m_i(\beta) + \Delta T_{n1i}(\beta) - \Delta T_{n2i}(\beta) + \eta_i(\beta)} \\
&- \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{\eta}_i(\beta)}{\eta_i(\beta)} \left(\frac{(\hat{g}_{ni} - g_i)m_i(\beta) + \Delta T_{n1i}(\beta) - \Delta T_{n2i}(\beta)}{(\hat{g}_{ni} - g_i)m_i(\beta) + \Delta T_{n1i}(\beta) - \Delta T_{n2i}(\beta) + \eta_i(\beta)} \right),
\end{aligned}$$

Note that for $\beta \in [a, b]$, we have $m(x, y; \beta)$ and $\partial m(x, y; \beta)/\partial\beta$ are both bounded below and above by $K_1^y(x)e^{-K_2(x)}$ and $y^2 E[K_2^{y+2}(X)e^{-K_1(x)}]$.

Let's first consider the asymptotic behavior of $\sum_{i=1}^n [(\hat{g}_n(Z_i) - g(Z_i))m(Y_i, Z_i, \beta)]^4$ and $\sum_{i=1}^n [T_{nj}(Y_i, Z_i, \beta_0) - T_j(Y_i, Z_i, \beta)]^4, \quad j = 1, 2.$

First, we have

$$E\left(\sum_{i=1}^{n}[(\hat{g}_n(Z_i)-g(Z_i))m(Y_i,Z_i,\beta)]^4\right)=nE[(\hat{g}_n(Z_1)-g(Z_1))m(Y_1,Z_1,\beta)]^4$$
$$\leq nE[(\hat{g}_n(Z_1)-g(Z_1))[K_2(Z_1)]^{Y_1}\exp(-K_1(Z_1))]^4.$$

Denote $U(Z,Y)=[K_2(Z)]^Y\exp(-K_1(Z))$ and rewrite $\hat{g}_n(Z_1)$ as

$$\frac{1}{nh}K(0)+\frac{1}{(n-1)h}\sum_{i=2}^{n}K\left(\frac{Z_i-Z_1}{h}\right)-\frac{1}{n(n-1)h}\sum_{i=2}^{n}K\left(\frac{Z_i-Z_1}{h}\right).$$

So,

$$E[(\hat{g}_n(Z_1)-g(Z_1))U(Y_1,Z_1)]^4$$
$$\leq\frac{27}{n^4h^4}K^4(0)EU^4(Y_1,Z_1)+27E\left[\frac{1}{(n-1)h}\sum_{i=2}^{n}K\left(\frac{Z_i-Z_1}{h}\right)-g(Z_1)\right]^4U^4(Y_1,Z_1)$$
$$+27E\left[\frac{1}{n(n-1)h}\sum_{i=2}^{n}K\left(\frac{Z_i-Z_1}{h}\right)\right]^4U^4(Y_1,Z_1).$$

Denote

$$\xi_{i1}=\frac{1}{h}K\left(\frac{Z_i-Z_1}{h}\right)-E\left(\frac{1}{h}K\left(\frac{Z_i-Z_1}{h}\right)\Bigg|Z_1\right),\quad \xi_1=E\left(\frac{1}{h}K\left(\frac{Z_i-Z_1}{h}\right)\Bigg|Z_1\right),$$

Then

$$E\left[\frac{1}{(n-1)h}\sum_{i=2}^{n}K\left(\frac{Z_i-Z_1}{h}\right)-g(Z_1)\right]^4U^4(Y_1,Z_1)$$
$$\leq 8E\left[\frac{1}{(n-1)}\sum_{i=2}^{n}\xi_{i1}\right]^4U^4(Y_1,Z_1)+8E\left[\xi_1-g(Z_1)\right]^4U^4(Y_1,Z_1).$$

Note that

$$E\left[\frac{U(Y_1,Z_1)}{n-1}\sum_{i=2}^{n}\xi_{i1}\right]^4=\frac{E\xi_{21}^4U^4(Y_1,Z_1)}{(n-1)^3}+\frac{3(n-2)E\xi_{21}^2\xi_{31}^2U^4(Y_1,Z_1)}{(n-1)^3}=O\left(\frac{1}{n^2h^2}\right)$$

and $E\left[\xi_1 - g(Z_1)\right]^4 U^4(Y_1, Z_1) = O(h^8)$. It is also easy to see that, by the boundedness of K,

$$E\left[\frac{1}{n(n-1)h}\sum_{i=2}^{n} K\left(\frac{Z_i - Z_1}{h}\right)\right]^4 U^4(Y_1, Z_1) = O\left(\frac{1}{n^4}\right).$$

Therefore, we have

$$E\left(\sum_{i=1}^{n}[(\hat{g}_n(Z_i) - g(Z_i))m(Y_i, Z_i, \beta)]^4\right) = O\left(\frac{1}{nh^2}\right) + O(nh^8)$$

uniformly in $\beta \in \Theta$, and

$$\max_{1\le i\le n,\,\beta\in\Theta} |(\hat{g}_n(Z_i) - g(Z_i))m(Y_i, Z_i, \beta)| = O_p\left(\sqrt[4]{\frac{1}{nh^2} + nh^8}\right) \tag{3.8}$$

On the other hand, we have $E\left[\sum_{i=1}^{n}(T_{n1i}(\beta) - T_{1i}(\beta))^4\right] = nE(T_{n11} - T_{11})^4$. Define

$$\tilde{H}_-(Y_i, Z_i, Z_j) = H_-(Y_i, Z_j, \beta)\exp\left(-\frac{\sqrt{2}|Z_i - Z_j|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j).$$

Then $T_{n11}(\beta) = n^{-1}\tilde{H}_-(Y_1, Z_1, Z_1) + n^{-1}\sum_{j=2}^{n}\tilde{H}_-(Y_1, Z_1, Z_j)$

$$T_{11}(\beta) = E[T_{n11}|Y_1, Z_1] = \frac{1}{n}\tilde{H}_-(Y_1, Z_1, Z_1) + \frac{1}{n}\sum_{j=2}^{n} E[\tilde{H}_-(Y_1, Z_1, Z_j)|Y_1, Z_1].$$

So

$$E\left[\sum_{i=1}^{n}(T_{n1i}(\beta) - T_{1i}(\beta))^4\right] = \frac{1}{n^3}E\left[\sum_{j=2}^{n}\left(\tilde{H}_-(Y_1, Z_1, Z_j) - E[\tilde{H}_-(Y_1, Z_1, Z_2)|Y_1, Z_1]\right)\right]^4$$

is the order of $O(n^{-1})$. By the compactness of Θ, one can show that the above holds uniformly in β. Therefore

$$\max_{1\le i\le n,\beta\in\Theta} |T_{n1i}(\beta) - T_{1i}(\beta)| \le \left[\sum_{i=1}^{n}(T_{n1i}(\beta) - T_{1i}(\beta))^4\right]^{1/4} = O_p\left(\frac{1}{\sqrt[4]{n}}\right). \tag{3.9}$$

Similarly, we also have

$$\max_{1\le i\le n,\beta\in\Theta} |T_{n2i}(\beta) - T_{2i}(\beta)| = O_p\left(n^{-1/4}\right). \tag{3.10}$$

From (3.8), (3.9) and (3.10), one can see that

$$\left|\dot{L}_n(\beta) - \dot{\hat{L}}_n(\beta)\right| \le o_p(1)\cdot\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{|\eta_i(\beta)|} + o_p(1)\right] - o_p(1)\cdot\frac{1}{n}\sum_{i=1}^{n}\left[\left|\frac{\dot{\eta}_i(\beta)}{\eta_i^2(\beta)}\right| + o_p(1)\right] = o_p(1)$$

by the conditions of the theorem. Finally, the second claim in (3.7) can be justified by using law of large numbers for each $\beta\in\Theta$, the continuity of $\dot{\hat{L}}_n(\beta) - \dot{L}(\beta)$ as a function of $\beta\in\Theta$ and the compactness of Θ. □

Now, let's prove the asymptotic normality of $\hat{\beta}_n$.

Proof. We begin with showing that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\hat{g}_n(Z_i)\dot{m}(Y_i,Z_i,\beta_0) + \dot{T}_{n1}(Y_i,Z_i,\beta_0) - \dot{T}_{n2}(Y_i,Z_i,\beta_0)}{\hat{g}_n(Z_i)m(Y_i,Z_i,\beta_0) + T_{n1}(Y_i,Z_i,\beta_0) - T_{n2}(Y_i,Z_i,\beta_0)}$$

is asymptotically normal. For the sake of convenience, denote $m_i = \dot{m}(Y_i, Z_i; \beta_0)$, $T_{n1i} = T_{n1}(Y_i, Z_i; \beta_0)$, $T_{n2i} = T_{n2}(Y_i, Z_i; \beta_0)$, and $\dot{m}_i, \dot{T}_{n1i}, \dot{T}_{n2i}$ are similarly defined. Note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\hat{g}_{ni}\dot{m}_i + \dot{T}_{n1i} - \dot{T}_{n2i}}{\hat{g}_{ni}m_i + T_{n1i} - T_{n2i}} \qquad (3.11)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{g}_{ni}-g_i)\dot{m}_i + (\dot{T}_{n1i}-\dot{T}_{1i}) - (\dot{T}_{n2i}-\dot{T}_{2i}) + g_i\dot{m}_i + \dot{T}_{1i} - \dot{T}_{2i}}{(\hat{g}_{ni}-g_i)m_i + (T_{n1i}-T_{1i}) - (T_{n2i}-T_{2i}) + g_i m_i + T_{1i} - T_{2i}}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{g}_{ni}-g_i)\dot{m}_i + (\dot{T}_{n1i}-\dot{T}_{1i}) - (\dot{T}_{n2i}-\dot{T}_{2i}) + g_i\dot{m}_i + \dot{T}_{1i} - \dot{T}_{2i}}{g_i m_i + T_{1i} - T_{2i}}\\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{g}_{ni}-g_i)\dot{m}_i + (\dot{T}_{n1i}-\dot{T}_{1i}) - (\dot{T}_{n2i}-\dot{T}_{2i}) + g_i\dot{m}_i + \dot{T}_{1i} - \dot{T}_{2i}}{g_i m_i + T_{1i} - T_{2i}}\cdot\\
&\qquad\left(\frac{(\hat{g}_{ni}-g_i)m_i + (T_{n1i}-T_{1i}) - (T_{n2i}-T_{2i})}{(\hat{g}_{ni}-g_i)m_i + (T_{n1i}-T_{1i}) - (T_{n2i}-T_{2i}) + g_i m_i + T_{1i} - T_{2i}}\right).
\end{aligned}$$

For the sake of brevity, denote $\mathrm{d}_{ni} = (\hat{g}_{ni} - g_i)m_i + (T_{n1i} - T_{1i}) - (T_{n2i} - T_{2i}), \eta_i = g_i m_i + T_{1i} - T_{2i}, and_i = g_i \dot{m}_i + \dot{T}_{1i} - \dot{T}_{2i}$. Then the second term in (3.11) can be written as the sum of the following twelve terms

$$S_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)^2 m_i \dot{m}_i}{(d_{ni} + \eta_i)\eta_i}, \qquad S_{n2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)(T_{n1i} - T_{1i})\dot{m}_i}{(d_{ni} + \eta_i)\eta_i},$$

$$S_{n3} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)(T_{n2i} - T_{2i})\dot{m}_i}{(d_{ni} + \eta_i)\eta_i}, \quad S_{n4} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)(\dot{T}_{n1i} - \dot{T}_{1i})m_i}{(d_{ni} + \eta_i)\eta_i},$$

$$S_{n5} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\dot{T}_{n1i} - \dot{T}_{1i})(T_{n1i} - T_{1i})}{(d_{ni} + \eta_i)\eta_i}, \qquad S_{n6} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\dot{T}_{n1i} - \dot{T}_{1i})(T_{n2i} - T_{2i})}{(d_{ni} + \eta_i)\eta_i},$$

$$S_{n7} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)(\dot{T}_{n2i} - \dot{T}_{2i})m_i}{(d_{ni} + \eta_i)\eta_i}, \quad S_{n8} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\dot{T}_{n2i} - \dot{T}_{2i})(T_{n1i} - T_{1i})}{(d_{ni} + \eta_i)\eta_i},$$

$$S_{n9} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\dot{T}_{n2i} - \dot{T}_{2i})(T_{n2i} - T_{2i})}{(d_{ni} + \eta_i)\eta_i}, \qquad S_{n10} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i(\hat{g}_{ni} - g_i)m_i}{(d_{ni} + \eta_i)\eta_i},$$

$$S_{n11} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i(T_{n1i} - T_{1i})}{(d_{ni} + \eta_i)\eta_i}, \qquad S_{n12} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i(T_{n2i} - T_{2i})}{(d_{ni} + \eta_i)\eta_i}.$$

To show that the first 9 terms of the second term in (3.11) is the order of $o_p(1)$, it suffices to show $E(S_{nj})^2 = o(1)$ for $j = 1, 2, \ldots, 9$. We only present the proofs for the cases $j = 1, 2$. Note that

$$\begin{aligned}
ES_{n1}^2 &= E\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)^2 m_i \dot{m}_i}{(d_{ni} + \eta_i)\eta_i}\right]^2 \\
&= \frac{1}{n} E\left[\sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)^4 m_i^2 \dot{m}_i^2}{(d_{ni} + \eta_i)^2 \eta_i^2}\right] + \frac{1}{n} E\left[\sum_{i \neq j} \frac{(\hat{g}_{ni} - g_i)^2 (\hat{g}_{nj} - g_j)^2 m_i \dot{m}_i m_j \dot{m}_j}{(d_{ni} + \eta_i)(d_{nj} + \eta_j)\eta_i \eta_j}\right] \\
&= E\left[\frac{(\hat{g}_{n1} - g_1)^4 m_1^2 \dot{m}_1^2}{(d_{n1} + \eta_1)^2 \eta_1^2}\right] + (n-1)E\left[\frac{(\hat{g}_{n1} - g_1)^2 (\hat{g}_{n2} - g_2)^2 m_1 \dot{m}_1 m_2 \dot{m}_2}{(d_{n1} + \eta_1)(d_{n2} + \eta_2)\eta_1 \eta_2}\right] \\
&\le nE\left[\frac{(\hat{g}_{n1} - g_1)^4 m_1^2 \dot{m}_1^2}{(d_{n1} + \eta_1)^2 \eta_1^2}\right].
\end{aligned}$$

Before proceeding, we want to show that $P(\eta_1 + d_{n1} < \eta_1/2, i.o.) = 0$. In fact,

$$\sum_{n=1}^{\infty} P(\eta_1 + d_{n1} < \eta_1/2)) \le \sum_{n=1}^{\infty} P(|d_{n1}| \ge \eta_1/2) \le c \sum_{n=1}^{\infty} E\left[\frac{d_{n1}^{2r}}{\eta_1^{2r}}\right]$$
$$\le c \sum_{n=1}^{\infty} E\left(\frac{(\hat{g}_{n1} - g_1)m_1}{\eta_1}\right)^{2r} + c \sum_{n=1}^{\infty} E\left(\frac{T_{n11} - T_{11}}{\eta_1}\right)^{2r} + c \sum_{n=1}^{\infty} E\left(\frac{T_{n21} - T_{21}}{\eta_1}\right)^{2r}$$
$$= c \sum_{n=1}^{\infty} \left[O\left(\frac{1}{(nh)^r}\right) + O(h^{4r}) + O\left(\frac{1}{n^r}\right)\right] < \infty.$$

Therefore, when n is large enough,

$$ES_{n1}^2 \le nE\left[\frac{(\hat{g}_{n1} - g_1)^4 m_1^2 \dot{m}_1^2}{(d_{n1} + \eta_1)^2 \eta_1^2}\right] \le 4nE\left[\frac{(\hat{g}_{n1} - g_1)^4 m_1^2 \dot{m}_1^2}{\eta_1^4}\right] = o_p(1).$$

Similarly, for S_{n2}, we have

$$\begin{aligned}
ES_{n2}^2 &= E\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)(T_{n1i} - T_{1i})\dot{m}_i}{(d_{ni} + \eta_i + \delta_n)\eta_i}\right]^2 \\
&= \frac{1}{n} E\left[\sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)^2 (T_{n1i} - T_{1i})^2 \dot{m}_i^2}{(d_{ni} + \eta_i)^2 \eta_i^2}\right] \\
&\quad + \frac{1}{n} E\left[\sum_{i \ne j} \frac{(\hat{g}_{ni} - g_i)(\hat{g}_{nj} - g_j)(T_{n1i} - T_{1i})(T_{n1j} - T_{1j})\dot{m}_i \dot{m}_j}{(d_{ni} + \eta_i)(d_{nj} + \eta_j)\eta_i \eta_j}\right] \\
&= E\left[\frac{(\hat{g}_{n1} - g_1)^2 (T_{n11} - T_{11})^2 \dot{m}_1^2}{(d_{n1} + \eta_1)^2 \eta_1^2}\right] \\
&\quad + (n-1) E\left[\frac{(\hat{g}_{n1} - g_1)(\hat{g}_{n2} - g_2)(T_{n11} - T_{11})(T_{n12} - T_{12})\dot{m}_1 \dot{m}_2}{(d_{n1} + \eta_1)(d_{n2} + \eta_2)\eta_1 \eta_2}\right] \\
&\le n \sqrt{E\left[\frac{(\hat{g}_{n1} - g_1)^4 \dot{m}_1^4}{(d_{n1} + \eta_1)^2 \eta_1^2}\right] \cdot E\left[\frac{(T_{n11} - T_{11})^4}{(d_{n1} + \eta_1)^2 \eta_1^2}\right]} = o(1).
\end{aligned}$$

Furthermore, we can show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (\hat{g}_{ni} - g_i) m_i}{\eta_i^2} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (\hat{g}_{ni} - g_i) m_i}{(d_{ni} + \eta_i)\eta_i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (\hat{g}_{ni} - g_i) m_i d_{ni} \eta_i}{(d_{ni} + \eta_i)\eta_i^3} = o_p(1),$$

and

$$S_{n11} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (T_{n1i} - T_{1i})}{\eta_i^2} + o_p(1), \quad S_{n12} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (T_{n2i} - T_{2i})}{\eta_i^2} + o_p(1).$$

Then we eventually have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{g}_{ni}\dot{m}_i + \dot{T}_{n1i} - \dot{T}_{n2i}}{\hat{g}_{ni} m_i + T_{n1i} - T_{n2i}} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{g}_{ni} - g_i)\eta_i \dot{m}_i + \eta_i(\dot{T}_{n1i} - \dot{T}_{1i}) - \eta_i(\dot{T}_{n2i} - \dot{T}_{2i}) + \eta_i \dot{\eta}_i}{\eta_i^2} \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (\hat{g}_{ni} - g_i) m_i}{\eta_i^2} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (T_{n1i} - T_{1i})}{\eta_i^2} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i (T_{n2i} - T_{2i})}{\eta_i^2} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{(\hat{g}_{ni} - g_i)(-\dot{\eta}_i m_i + \eta_i \dot{m}_i)}{\eta_i^2} \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\dot{T}_{n1i} - \dot{T}_{1i}}{\eta_i} \right] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\dot{\eta}_i (T_{n1i} - T_{1i})}{\eta_i^2} \right] + o_p(1) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\dot{T}_{n2i} - \dot{T}_{2i}}{\eta_i} \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\dot{\eta}_i (T_{n2i} - T_{2i})}{\eta_i^2} \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i}{\eta_i} + o_p(1).
\end{aligned}$$

At this stage, we will introduce two lemmas.

Lemma 12. *Assume that $\mu(x)$ is a continuous function, and the density function g of Z is twice differentiable with bounded second derivative. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mu(Z_i)[\hat{g}_n(Z_i) - g(Z_i)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mu(Z_i) g(Z_i) - E\mu(Z) g(Z)] + o_p(1).$$

The proof of Lemma 12 can be found in Shi et al. (2019).

Lemma 13. *Under some regularity conditions, for $k = 1, 2$, we have*

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\dot{\eta}_i T_{nki}}{\eta_i^2} - E\left[\frac{\dot{\eta} T_k}{\eta^2} \right] \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\dot{\eta}_i}{2\eta_i^2} T_{ki}(\beta_0) + \frac{1}{2} M_{ki}(\beta_0) - E\left[\frac{\dot{\eta} T_k}{\eta^2} \right] \right) + o_p(1), \\
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\dot{T}_{nki}}{\eta_i} - E\left[\frac{\dot{T}_k}{\eta} \right] \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\dot{T}_{ki}(\beta_0)}{2\eta_i} + \frac{1}{2} N_{ki}(\beta_0) - E\left[\frac{\dot{T}_k}{\eta} \right] \right) + o_p(1),
\end{aligned}$$

where

$$M_{1i}(\beta_0) = E\left[\frac{\dot{\eta}_j}{\eta_j^2} H_-(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_j,\infty)}(Z_i) \middle| Y_i, Z_i\right],$$
$$M_{2i}(\beta_0) = E\left[\frac{\dot{\eta}_j}{\eta_j^2} H_+(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j) \middle| Y_i, Z_i\right],$$
$$N_{1i}(\beta_0) = E\left[\frac{1}{\eta_j} \dot{H}_-(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_j,\infty)}(Z_i) \middle| Y_i, Z_i\right],$$
$$N_{2i}(\beta_0) = E\left[\frac{1}{\eta_j} \dot{H}_+(Y_j, Z_i, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_j - Z_i|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j) \middle| Y_i, Z_i\right].$$

Proof. Denote

$$W_{ij1} = H_-(Y_i, Z_j, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_i - Z_j|}{\sigma_u}\right) I_{[Z_i,\infty)}(Z_j),$$

and $V_i = \dot{\eta}_i/\eta_i^2$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i T_{n1i}}{\eta_i^2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \left[\frac{1}{n} \sum_{j=1}^{n} W_{ij1}\right] = \frac{1}{n\sqrt{n}} \sum_{i=1}^{n} V_i W_{ii1} + \frac{1}{n\sqrt{n}} \sum_{i \neq j} V_i W_{ij1}.$$

The assumption of $EV_1W_{111} < \infty$ implies the first term on the right is $o_p(1)$. Denote $U_{ij1} = (V_iW_{ij1} + V_jW_{ji1})/2$,

$$\frac{1}{n\sqrt{n}} \sum_{i \neq j} V_i W_{ij1} = \frac{n-1}{\sqrt{n}} \cdot \frac{2}{n(n-1)} \sum_{i<j} U_{ij1} = \frac{n-1}{\sqrt{n}} U_n.$$

It is easy to see that U_n is a U-statistic. If $E[VW]^2 < \infty$, and $\mathrm{Var}[E(U_{121}|Z_1, Y_1)] > 0$, then

$$\sqrt{n}[U_n - E(V_1W_{121})] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [E(U_{ij1}|Z_i, Y_i) - E(V_1W_{121})] + o_p(1),$$

where $i \neq j$. Note that

$$E(U_{ij1}|Z_i, Y_i) = \frac{1}{2} E\left[V_iW_{ij1}|Z_i, Y_i\right] + \frac{1}{2} E\left[V_jW_{ji1}|Z_i, Y_i\right]$$

and

$$
\begin{aligned}
&E\left[V_i W_{ij1} | Z_i, Y_i\right] = V_i E[W_{ij1}|Z_i, Y_i] \\
&= \frac{\dot{\eta}_i}{\eta_i^2} E\left[H_-(Y_i, Z_j, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_i - Z_j|}{\sigma_u} \right) I_{[Z_i,\infty)}(Z_j) \middle| Y_i, Z_i \right] \\
&= \frac{\dot{\eta}_i}{\eta_i^2} \int_{Z_i}^{\infty} H_-(Y_i, x, \beta_0) G(x, Z_i) dx = \frac{\dot{\eta}_i}{\eta_i^2} T_{1i}(\beta_0).
\end{aligned}
$$

On the other hand, we can show that $E\left[V_j W_{ji1}|Z_i, Y_i\right] = M_{1i}(\beta_0)$. Now, take

$$
W_{ij2} = H_+(Y_i, Z_j, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_i - Z_j|}{\sigma_u} \right) I_{(-\infty,Z_i]}(Z_j),
$$

then similarly, we can show that

$$
\begin{aligned}
&E\left[V_i W_{ij2} | Z_i, Y_i\right] = V_i E[W_{ij2}|Z_i, Y_i] \\
&= \frac{\dot{\eta}_i}{\eta_i^2} E\left[H_+(Y_i, Z_j, \beta_0) \exp\left(-\frac{\sqrt{2}|Z_i - Z_j|}{\sigma_u} \right) I_{(-\infty,Z_i]}(Z_j) \middle| Y_i, Z_i \right] \\
&= \frac{\dot{\eta}_i}{\eta_i^2} \int_{-\infty}^{Z_i} H_+(Y_i, x, \beta_0) G(x, Z_i) dx = \frac{\dot{\eta}_i}{\eta_i^2} T_{2i}(\beta_0)
\end{aligned}
$$

and $E\left[V_j W_{ji2}|Z_i, Y_i\right] = M_{2i}(\beta_0)$. This concludes the proof. □

From the above lemma, we showed that

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i(T_{n1i} - T_{1i})}{\eta_i^2} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{\eta}_i(T_{n2i} - T_{2i})}{\eta_i^2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{M_{1i} - M_{2i}}{2} - \frac{\dot{\eta}_i(T_{1i} - T_{2i})}{2\eta_i^2} \right].
$$

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{T}_{n1i} - \dot{T}_{1i}}{\eta_i} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\dot{T}_{n2i} - \dot{T}_{2i}}{\eta_i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{N_{1i} - N_{2i}}{2} - \frac{\dot{T}_{1i} - \dot{T}_{2i}}{2\eta_i} \right].
$$

For the sake of brevity, denote

$$S_{1i} = \frac{(\eta_i \dot{m}_i - \dot{\eta}_i m_i) g(Z_i)}{\eta_i^2} - E\frac{(\eta \dot{m} - \dot{\eta} m) g(Z)}{\eta^2},$$
$$S_{2i} = \frac{M_{1i} - M_{2i}}{2} - \frac{\dot{\eta}_i (T_{1i} - T_{2i})}{2\eta_i^2}, \quad S_{3i} = \frac{N_{1i} - N_{2i}}{2} - \frac{\dot{T}_{1i} - \dot{T}_{2i}}{2\eta_i}.$$

Then base on Lemma 12 and 13, we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{g}_{ni} \dot{m}_i + \dot{T}_{n1i} - \dot{T}_{n2i}}{\hat{g}_{ni} m_i + T_{n1i} - T_{n2i}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[S_{1i} - S_{2i} + S_{3i} + \frac{\dot{\eta}_i}{\eta_i} \right] + o_p(1).$$

This implies that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{g}_{ni} \dot{m}_i + \dot{T}_{n1i} - \dot{T}_{n2i}}{\hat{g}_{ni} m_i + T_{n1i} - T_{n2i}} \Longrightarrow N(0, \sigma^2),$$

where $\sigma^2 = E(S_{11} - S_{21} + S_{31} + \dot{\eta}/\eta)^2$. And note that the estimator $\hat{\beta}_n$ satisfies the following equation

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{g}_n(Z_i) \dot{m}(Y_i, Z_i, \hat{\beta}_n) + \dot{T}_{n1}(Y_i, Z_i, \hat{\beta}_n) - \dot{T}_{n2}(Y_i, Z_i, \hat{\beta}_n)}{\hat{g}_n(Z_i) m(Y_i, Z_i, \hat{\beta}_n) + T_{n1}(Y_i, Z_i, \hat{\beta}_n) - T_{n2}(Y_i, Z_i, \hat{\beta}_n)} = 0.$$

By Taylor expansion, we have

$$\begin{aligned} 0 = & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{g}_n(Z_i) \dot{m}(Y_i, Z_i, \beta_0) + \dot{T}_{n1}(Y_i, Z_i, \beta_0) - \dot{T}_{n2}(Y_i, Z_i, \beta_0)}{\hat{g}_n(Z_i) m(Y_i, Z_i, \beta_0) + T_{n1}(Y_i, Z_i, \beta_0) - T_{n2}(Y_i, Z_i, \beta_0)} \\ & + \sqrt{n}(\hat{\beta}_n - \beta_0) \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \left[\frac{\hat{g}_n(Z_i) \dot{m}(Y_i, Z_i, \tilde{\beta}_n) + \dot{T}_{n1}(Y_i, Z_i, \tilde{\beta}_n) - \dot{T}_{n2}(Y_i, Z_i, \tilde{\beta}_n)}{\hat{g}_n(Z_i) m(Y_i, Z_i, \tilde{\beta}_n) + T_{n1}(Y_i, Z_i, \tilde{\beta}_n) - T_{n2}(Y_i, Z_i, \tilde{\beta}_n)} \right]. \end{aligned}$$

The consistency of $\hat{\beta}_n$ to β_0 implies that $\tilde{\beta}_n$ also converges to β_0 in probability. Therefore,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \left[\frac{\hat{g}_n(Z_i) \dot{m}(Y_i, Z_i, \tilde{\beta}_n) + \dot{T}_{n1}(Y_i, Z_i, \tilde{\beta}_n) - \dot{T}_{n2}(Y_i, Z_i, \tilde{\beta}_n)}{\hat{g}_n(Z_i) m(Y_i, Z_i, \tilde{\beta}_n) + T_{n1}(Y_i, Z_i, \tilde{\beta}_n) - T_{n2}(Y_i, Z_i, \tilde{\beta}_n)} \right] \\ & \to E \frac{\partial}{\partial \beta} \left[\frac{g(Z) \dot{m}(Y, Z, \beta_0) + \dot{T}_1(Y, Z, \beta_0) - \dot{T}_2(Y, Z, \beta_0)}{g(Z) m(Y, Z, \beta_0) + T_1(Y, Z, \beta_0) - T_2(Y, Z, \beta_0)} \right] \end{aligned}$$

which is the Fisher information number based on the joint density function of Y and Z. Thus, we conclude the main result on the asymptotic normality of $\hat{\beta}_n$. □

Bibliography

Alan Agresti. *Categorical Data Analysis*. John Wiley & Sons, Inc., 2002.

John P Buonaccorsi. *Measurement error: models, methods, and applications*. Chapman and Hall/CRC, 2010.

A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.

Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC, 2006.

X.R. Chen. Unifrom convergence rates of kernel density estimation. *Journal of Systems Science and Mathematical Sciences*, 3(4):263–272, 1983.

Chi-Lun Cheng and John W. Van Ness. *Statistical Regression with Measurement Error: Kendall's Library of Statistics 6*. Wiley, 2010.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106:1602–1614, 12 2011.

Thomas S Ferguson. *A course in large sample theory*. Routledge, 2017.

Wayne A Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.

Jie Q Guo and Tong Li. Poisson regression models with errors-in-variables: implication and treatment. *Journal of statistical planning and inference*, 104(2):391–401, 2002.

H. Hong and E.A. Tamer. Simple estimator for nonlinear error in variable models. *Journal of Econometrics*, 117(1):1–19, 2003.

McCullagh and Nelder. *Generalized linear models*. Chapman and Hall, 1989.

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

H. Robbins. An empirical bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume I, pages 157–163. University of California Press, Berkeley and Los Angeles, 1956.

Jianhong Shi and Weixing Song. Nonlinear statistics inferences with laplace measurement error. *J. Sys. Sci & Math. Scis.*, 35(12):1510–1528, 2015.

Jianhong Shi, Jing Feng, and Weixing Song. Estimation in linear regression with laplace measurement error using tweedie-type formula. *J. Syst. Sci. Complex.*, 32(4):1211–1230, 2019.

Weixing Song, Xiuqin Bai, and Jianhong Shi. Tweedie-type formulae for a multivariate laplace distribution. *submitted to Statistics and Probability Letters*, 2021.

Leonard A Stefanski. Unbiased estimation of a nonlinear function a normal mean with application to measurement err oorf models. *Communications in Statistics-Theory and Methods*, 18(12):4335–4358, 1989.

Leonard A Stefanski and Raymond J Carroll. Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, 74(4):703–716, 1987.

M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall/CRC, 1995.

Liqun Wang. Estimation of nonlinear models with berkson measurement errors. *The Annals of Statistics*, 32(6):2559–2579, 2004.

R.W.M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447, 1974.