

# Inference on a new sigmoid regression model with unknown support and unbounded likelihood function

*HaiYing Wang*

hwzq7@mail.missouri.edu

*Nancy Flournoy*

flournoy@missouri.edu

*Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.*

## **Abstract**

In this paper we introduce a non-regular sigmoid shape regression model with boundaries of support for the response variable being two unknown parameters; the likelihood function is unbounded, for which the global maximizers are not consistent estimators. Although the two sample extremes, the smallest and the largest observations, are consistent estimators of the two unknown boundaries, they have slow convergence rate and are asymptotically biased. Improved estimators are developed by correcting for the asymptotic biases of the two sample extremes for the one sample case; but these proposed estimators do not obtain the optimal convergence rate either. To obtain efficient estimation, we resort to the local maximizers of the likelihood function, i.e., the solution to the likelihood equations that is obtained by setting to zero the gradient of the log-likelihood function with respect to the parameters. We prove that, with probability approaching one as the sample size goes to infinity, there exists a solution to the likelihood equation that is consistent at the rate of the square root of the sample size and it is asymptotically normally distributed.

*Keywords:* **Asymptotics; Bioassay; Consistency; Logistic model; Maximum likelihood estimation; Parameter dependent support.**

# Inference on a new sigmoid regression model with unknown support and unbounded likelihood function

Haiying Wang and Nancy Flournoy

hwzq7@mail.missouri.edu flournoyn@missouri.edu

*Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.*

January 10, 2012

## Abstract

In this paper we introduce a non-regular sigmoid shape regression model with boundaries of support for the response variable being two unknown parameters; the likelihood function is unbounded, for which the global maximizers are not consistent estimators. Although the two sample extremes, the smallest and the largest observations, are consistent estimators of the two unknown boundaries, they have slow convergence rate and are asymptotically biased. Improved estimators are developed by correcting for the asymptotic biases of the two sample extremes for the one sample case; but these proposed estimators do not obtain the optimal convergence rate either. To obtain efficient estimation, we resort to the local maximizers of the likelihood function, i.e., the solution to the likelihood equations that is obtained by setting to zero the gradient of the log-likelihood function with respect to the parameters. We prove that, with probability approaching one as the sample size goes to infinity, there exists a solution to the likelihood equation that is consistent at the rate of the square root of the sample size and it is asymptotically normally distributed.

*Keywords:* **Asymptotics; Bioassay; Consistency; Logistic model; Maximum likelihood estimation; Parameter dependent support.**

## 1 Introduction

Assume an independent sample  $\{(x_i, Y_i), i = 1, \dots, n\}$  is from the following model,

$$\log\left(\frac{B - Y}{Y - A}\right) = a + bx + \varepsilon, \quad (1)$$

where  $Y$  is the response variable and  $x$  is a nonrandom scalar covariate, such as dose;  $\varepsilon$  is an error;  $a, b, \sigma, A$  and  $B$  are unknown parameters. This model can also be presented as

$$Y = B - \frac{B - A}{1 + e^{-(a+bx+\varepsilon)}},$$

by which one can see that this model is connected to the four parameter logistic model,

$$Y = B - \frac{B - A}{1 + e^{-(a+bx)}} + \varepsilon. \quad (2)$$

The four parameter logistic model for continuous responses is also called the  $E_{\max}$  model, and it is frequently used for curve-fitting analysis in immunoassays such as ELISAs and other bioassays. See, for example, DeLean *et al.* (1978), Vølund (1978), Holford and Sheiner (1981), Ratkowsky and Reedy (1986), Nix and Wild (2001), MacDougall (2006), Dragalin *et al.* (2007) and Sebaugh (2011). The  $E(Y | x)$  of model (2) is often used in phase I clinical trials to model the mean response for Bernoulli random variables. The applications considered here, and in the aforementioned references, focus on continuous random variables. We take  $\varepsilon$  to have a normal distribution  $N(0, \sigma^2)$  for models (1) and (2).

A drawback of the four parameter logistic model is that parameters  $A$  and  $B$  are often interpreted as the minimum and maximum of possible responses, although model (2) allows the responses to be unbounded. Another inadequacy of model (2) is that the responses  $Y$  have the same variance for all possible values of the covariate  $x$ , which is often violated in practice. Leonov and Miller (2009) tackled this problem by letting the variance of the model error depend on the covariate, but the range of possible responses remained to be bounded. Our model (1) has bounded responses and the distribution of the response for a given dose is skewed like a beta distribution. Figure 1 displays simulated data from models (1) and (2). Note that observations from model (2) may fall far outside the two hypothetical bounds, while data from model (1) always stays between the two boundaries. Additionally, data from model (2) still has large variation in the two tails, whereas variation in the tails is skewed and very small for model (1), and this scenario is observed frequently in real data.

A special case of model (1) is, by setting  $b = 0$  and replacing  $a$  by  $\mu$ ,

$$\log \left( \frac{B - Y}{Y - A} \right) = Z, \quad (3)$$

where  $Z \sim N(\mu, \sigma^2)$ , and  $\mu$ ,  $\sigma$ ,  $A$  and  $B$  are unknown parameters. Model (3) has some similarities to the three parameter log-normal distribution, in which

$$\log(Y - A) = Z \sim N(\mu, \sigma^2),$$

and which also has an unbounded likelihood function (see Hill, 1963). Although the three parameter log-normal distribution has been studied by many, including Cohen (1951), Hill (1963), Harter and Moore (1966), Giesbrecht and Kempthorne (1976) and Cohen and Whitten (1980), the theoretical properties of the proposed methods were not addressed rigorously in these papers. Cheng and Amin (1983) proposed an estimation method called maximum product spacings and proved the asymptotic normality of the proposed estimator and the local maximum likelihood estimator for the log-normal distribution. However, a rigorous proof for the consistency of the local maximum likelihood estimator was not provided.

There is a large body of literature rigorously developing methods of statistical inference for models with parameter-dependent support, including Woodroffe (1972), Weiss and Wolfowitz (1973), Woodroffe (1974), Smith (1985), Cheng and Iles (1987) and Smith (1994).

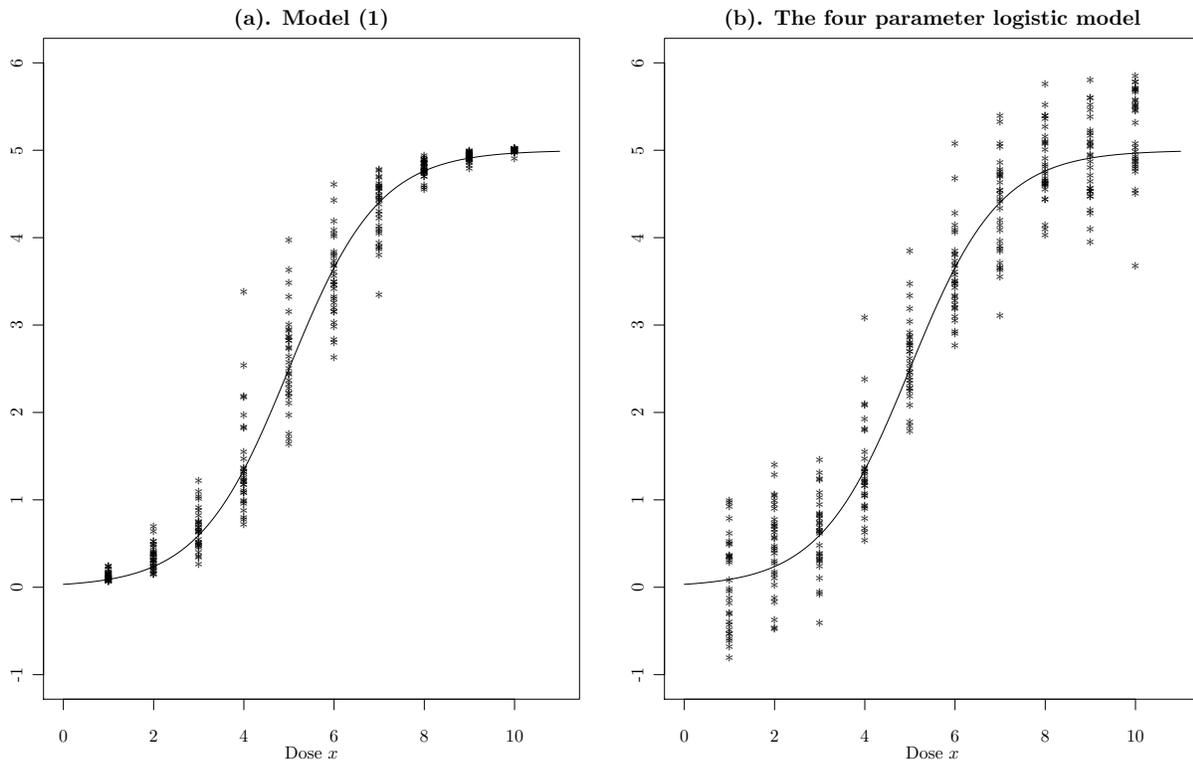


Figure 1: Simulated data from models (1) and (2) with  $a = 5$ ,  $b = -1$ ,  $\sigma = 0.5$ ,  $A = 0$  and  $B = 5$ . The sigmoid curve is the median response for model (1) while it is the mean response for model (2).

More references on non-regular models and estimation approaches for them can be found in Cheng and Traylor (1995) and the references and discussions therein.

This paper is closely related to the work of Smith (1985), in which instead of using the global maximizer of the likelihood function, the solution to the likelihood equation is used to estimate unknown parameters. This idea was originally proposed by Harter and Moore (1966). The theory of local maximum likelihood estimation was established for a broad class of non-regular models without covariates in Smith (1985) by an elegant mathematical derivation. A key requirement in their proof is that the difference between the sample minimum and the lower bound of the support of a distribution has a non-degenerate distribution asymptotically. However, as is shown in Section 2, this quantity for samples from model (3) always converges to a constant. Smith (1994) extended the results of Smith (1985) to a class of non-regular regression models, but model (1) does not meet the assumptions required in their analysis. In this paper, we find another technique to prove the existence of the consistent maximum likelihood estimator. With small modifications, this technique applies to the consistency of local maximum likelihood estimator for the well known three parameter log-normal distribution, for which a rigorous proof has been missing for a long time. Uniqueness of the local maximizer of the likelihood function is also investigated and a theorem similar to Theorem 2 in Smith (1985) is formulated for a general regression model.

The rest of this paper is organized as follows. In Section 2, the one sample case is

addressed and properties of estimators based on the sample extremes are derived. In Section 3, we present the results for the local maximum likelihood estimator for the regression problem under a general setup. Results of simulation experiments that are designed to investigate the finite sample properties are contained in Section 4. Technical details are given in the appendix.

## 2 Estimation based on the extreme order statistics for the one sample case

### 2.1 Naive estimator

Suppose an independent sample  $\{Y_1, \dots, Y_n\}$  is taken from model (3). If parameters  $A$  and  $B$  are estimated in advance,  $\mu$  and  $\sigma$  can be estimated simply by ordinary least-squares. A naive approach is then to use the two sample extremes,  $Y_{(1)}$  and  $Y_{(n)}$ , to estimate  $A$  and  $B$ , respectively, and then remove them from the sample and use the rest of the sample to estimate  $\mu$  and  $\sigma$ . We call such estimators naive. Why do the two sample extremes not perform well? It is not difficult to show that the two sample extremes are consistent, but their convergence rate is very slow. The proposition below gives asymptotic properties for these two statistics.

**Proposition 1.** *Let*

$$r_n = \{2 \log n\}^{1/2} - \frac{\log \log n + \log(4\pi)}{\{8 \log n\}^{1/2}}, \quad s_n = \frac{1}{\{2 \log n\}^{1/2}}.$$

*The following convergence results hold in distribution as  $n \rightarrow \infty$ :*

$$\begin{aligned} \frac{e^{\mu_0 + \sigma_0 r_n}}{\sigma_0 s_n} \left( \frac{1}{e^{\mu_0 + \sigma_0 r_n}} - \frac{Y_{(1)} - A_0}{B_0 - A_0} \right) &\rightarrow G_1, \\ \frac{e^{-\mu_0 + \sigma_0 r_n}}{\sigma_0 s_n} \left( \frac{1}{e^{-\mu_0 + \sigma_0 r_n}} - \frac{B_0 - Y_{(n)}}{B_0 - A_0} \right) &\rightarrow G_2, \end{aligned} \tag{4}$$

where  $\mu_0$ ,  $\sigma_0$ ,  $A_0$  and  $B_0$  are the true values of the parameters, and  $G_1$  and  $G_2$  are two independent random variables having the same distribution function  $F(t) = e^{-e^{-t}}$ .

*Proof.* In Appendix A.1. □

From this convergence result, it follows that, as  $n \rightarrow \infty$ ,

$$\begin{aligned} e^{\mu_0 + \sigma_0 r_n} (Y_{(1)} - A_0) &\rightarrow B_0 - A_0, \\ e^{-\mu_0 + \sigma_0 r_n} (B_0 - Y_{(n)}) &\rightarrow B_0 - A_0, \end{aligned} \tag{5}$$

in distribution, which gives the rate of convergence as  $e^{-\sigma_0 r_n}$ . Since, for any  $\alpha > 0$ ,  $e^{-\sigma_0 r_n} n^\alpha \rightarrow \infty$ , the rate of convergence is slower than  $n^{-\alpha}$  for any  $\alpha > 0$ . But it is still faster than  $1/\log n$  because  $e^{-\sigma_0 r_n} \log n \rightarrow 0$ . This proposition also tells us that there does not exist a constant sequence  $r_n^* \rightarrow \infty$  such that  $r_n^*(Y_{(1)} - A_0)$  or  $r_n^*(B_0 - Y_{(n)})$  converges to a non-degenerate distribution.

## 2.2 Bias adjusted estimators

Estimation based on the two sample extreme values can be improved by adjusting for their asymptotic biases. From (4) and (5), better estimators of  $A$  and  $B$  are obtained:

$$\begin{aligned}\hat{A}_{adj} &= Y_{(1)} - \frac{(1 - \gamma \hat{\sigma}^* s_n)(Y_{(n)} - Y_{(1)})}{\exp(\hat{\mu}^* + \hat{\sigma}^* r_n)}, \\ \hat{B}_{adj} &= Y_{(n)} + \frac{(1 - \gamma \hat{\sigma}^* s_n)(Y_{(n)} - Y_{(1)})}{\exp(-\hat{\mu}^* + \hat{\sigma}^* r_n)},\end{aligned}\tag{6}$$

where  $\hat{\mu}^*$  and  $\hat{\sigma}^*$  are two consistent estimates of  $\mu$  and  $\sigma$ , respectively, and  $\gamma \approx 0.577$  is the Euler-Mascheroni constant. Their sampling properties are given by the following convergence results:

$$\begin{aligned}\frac{e^{\mu_0 + \sigma_0 r_n}}{\sigma_0(B_0 - A_0)s_n}(\hat{A}_{adj} - A_0) &\rightarrow \gamma - G_1, \\ \frac{e^{\mu_0 + \sigma_0 r_n}}{\sigma_0(B_0 - A_0)s_n}(\hat{B}_{adj} - B_0) &\rightarrow \gamma - G_2,\end{aligned}$$

in distribution. By adjusting for the asymptotic biases of the two sample extremes, the estimators in (6) improve the rate of convergence from  $e^{-\sigma_0 r_n}$  to  $s_n e^{-\sigma_0 r_n}$ . Although this rate is also between  $1/\log n$  and  $n^{-\alpha}$  for any  $\alpha > 0$ , simulation results show that these estimators are much more efficient than the two sample extremes.

## 3 Maximum likelihood estimators

Since the estimators given in Section 2 do not possess the optimal convergence rate, we evaluate the method of maximum likelihood estimation. Generalizing model (1) in this section, we assume that an independent random sample  $\{(\mathbf{x}_i, Y_i), i = 1, \dots, n\}$  is taken from the model

$$\log\left(\frac{B - Y}{Y - A}\right) = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon,\tag{7}$$

where  $\mathbf{x}$  denotes a  $p$  dimensional covariate vector here and  $\boldsymbol{\beta}$  is an unknown  $p$  dimensional regression coefficient vector. For simplicity, let  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma, A, B)^T$  and let  $\boldsymbol{\theta}_0$  be the true value of  $\boldsymbol{\theta}$ . The likelihood function of  $\boldsymbol{\theta}$  based on the observed sample is

$$L_n(\boldsymbol{\theta}) = \frac{(B - A)^n I(A < Y_{(1)} < Y_{(n)} < B)}{(2\pi)^{\frac{n}{2}} \sigma^n \prod_{i=1}^n (B - Y_i)(Y_i - A)} \exp\left[-\frac{\sum_{i=1}^n \left\{\log\left(\frac{B - Y_i}{Y_i - A}\right) - \mathbf{x}_i^T \boldsymbol{\beta}\right\}^2}{2\sigma^2}\right],\tag{8}$$

where  $I(\cdot)$  is the indicator function. This likelihood function is unbounded and may become infinite along some paths; for example, let  $\boldsymbol{\beta} = 0$  and  $\sigma^2 = \sum_{i=1}^n [\log((B - Y_i)/(Y_i - A))]^2$ ; then  $\sigma^n \prod_{i=1}^n (B - Y_i)(Y_i - A)$  goes to 0 if  $A$  approaches  $Y_{(1)}$  from the left or  $B$  approaches  $Y_{(n)}$  from the right. So the likelihood function in (8) goes to infinity as  $\boldsymbol{\theta}$  goes to  $(0, +\infty, Y_{(1)}, Y_{(n)})$  along some paths. If  $\mu$  and  $\sigma$  are known, the likelihood function is bounded because it is

continuous and goes to 0 as  $A$  approaches  $Y_{(1)}$  or  $-\infty$ , or as  $B$  approaches  $Y_{(n)}$  or  $\infty$ . Since the likelihood function in (8) is unbounded, the global maximizer of the likelihood function is not a consistent estimator. In this section, following the idea of Smith (1985), we study the properties of local maximizer of the likelihood function, i.e., the solution to the likelihood equations. We prove the existence and consistency of the resultant estimator.

From calculations in Appendix A.2, the Fisher information matrix based on the sample is

$$\mathcal{I}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \begin{pmatrix} \frac{\mathbf{x}_i \mathbf{x}_i^T}{\sigma^2} & 0 & \frac{-1-c_i d}{\sigma^2(B-A)} \mathbf{x}_i^T & \frac{-1-\frac{d}{c_i}}{\sigma^2(B-A)} \mathbf{x}_i^T \\ 0 & \frac{2}{\sigma^2} & \frac{-2c_i d}{\sigma(B-A)} & \frac{2\frac{d}{c_i}}{\sigma(B-A)} \\ \frac{-1-c_i d}{\sigma^2(B-A)} \mathbf{x}_i & \frac{-2c_i d}{\sigma(B-A)} & \frac{c_i^2 d^4}{(B-A)^2} + \frac{1+2c_i d+c_i^2 d^4}{\sigma^2(B-A)^2} & \frac{-1}{(B-A)^2} + \frac{2+c_i d+\frac{d}{c_i}}{\sigma^2(B-A)^2} \\ \frac{-1-\frac{d}{c_i}}{\sigma^2(B-A)} \mathbf{x}_i & \frac{2\frac{d}{c_i}}{\sigma(B-A)} & \frac{-1}{(B-A)^2} + \frac{2+c_i d+\frac{d}{c_i}}{\sigma^2(B-A)^2} & \frac{\frac{d^4}{c_i^2}}{(B-A)^2} + \frac{1+2\frac{d}{c_i}+\frac{d^4}{c_i^2}}{\sigma^2(B-A)^2} \end{pmatrix},$$

where  $c_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$  and  $d = e^{\sigma^2/2}$ . If the Assumptions 1-3 in Appendix A.3 hold, then  $\mathcal{I}_n(\boldsymbol{\theta})/n$  converges to a positive-definite matrix, say  $\mathcal{I}(\boldsymbol{\theta})$ .

Denote the log-likelihood function by  $\ell_n(\boldsymbol{\theta})$ . The likelihood equations are

$$\begin{aligned} \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \frac{\left\{ \log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^T \boldsymbol{\beta} \right\}}{\sigma^2} \mathbf{x}_i = 0, \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{\left\{ \log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^T \boldsymbol{\beta} \right\}^2}{\sigma^3} = 0, \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial A} &= \frac{1}{B-A} \sum_{i=1}^n \frac{B-Y_i}{Y_i-A} - \sum_{i=1}^n \frac{\log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma^2(Y_i-A)} = 0, \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial B} &= -\frac{1}{B-A} \sum_{i=1}^n \frac{Y_i-A}{B-Y_i} - \sum_{i=1}^n \frac{\log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma^2(B-Y_i)} = 0. \end{aligned} \tag{9}$$

The following theorems present the properties of the local maximum likelihood estimator, the solution to (9). Proofs of these theorems are given in Appendix A.3.

**Theorem 1** (Existence). *If assumptions 1-3 in Appendix A.3. hold, then with probability approaching 1, there exists a sequence of solutions  $\boldsymbol{\theta}_n$  to the likelihood equations in (9) that is  $n^{1/2}$ -consistent for  $\boldsymbol{\theta}$ .*

**Theorem 2** (Uniqueness). *Assume assumptions 1-3 in Appendix A.3. hold. Let  $\delta$  be some fixed value and  $\delta_n = n^{-\alpha}$  for some  $\alpha > 0$ . Denote by  $S_\delta = \{\boldsymbol{\theta} : A \leq A_0 - \delta \text{ and } B \geq B_0 + \delta\}$  and  $T_{\delta,n} = \{\boldsymbol{\theta} : A_0 - \delta \leq A \leq A_0 + \delta_n, B_0 - \delta_n \leq B \leq B_0 + \delta \text{ and } \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| + |\sigma - \sigma_0| > \delta\}$ , where  $\|\cdot\|$  denote the Euclidean norm. Then, for any compact set  $K \subset \mathbb{R}^{p+3}$ ,*

$$\lim_{n \rightarrow \infty} Pr \left\{ \sup_{S_\delta \cap K} \ell_n(\boldsymbol{\theta}) < \ell_n(\boldsymbol{\theta}_0) \right\} = 1, \quad \lim_{n \rightarrow \infty} Pr \left\{ \sup_{T_{\delta,n} \cap K} \ell_n(\boldsymbol{\theta}) < \ell_n(\boldsymbol{\theta}_0) \right\} = 1.$$

**Theorem 3** (Asymptotic normality). *If assumptions 1-3 in Appendix A.3. hold, the  $n^{1/2}$ -consistent estimator  $\hat{\boldsymbol{\theta}}_n$  in Theorem 1 satisfies*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightarrow N \{0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\},$$

*in distribution.*

## 4 Numerical examples

In this section, simulation results are reported that examine the finite sample performance of the biased adjusted and local maximum likelihood estimators given in Sections 2 and 3. All the results are based on 1000 iterations of simulation.

Tables 1 and 2 give the relative mean square errors for model (3), the one sample case without covariates. The relative mean square errors are the ratios of the mean square errors of a given estimator calculated from simulated sample to that of the local maximum likelihood estimator defined in Section 3. So a value of relative mean square error greater than unity indicates that the given estimator is less efficient than the local maximum likelihood estimator, and vice-versa. Table 1 reports results when  $\mu$  and  $\sigma$  are assumed to be known while they are unknown in Table 2. A consistent solution to the likelihood equations always exists in our simulation studies if  $\mu$  and  $\sigma$  are known. When  $\mu$  and  $\sigma$  are unknown, a consistent solution to the likelihood equations occasionally did not exist for small sample sizes. From Table 2, one can see that this occurs rarely; in the worst case, consistent solutions were not found in 6 iterations out of 1000. When a solution to the likelihood equations was not found, the bias adjusted estimator was used instead.

It is seen in Tables 1 and 2 that all the relative mean square errors are greater than unity, which means both the naive and the bias adjusted estimators are dominated by the local maximum likelihood estimator. Furthermore, the relative performance of these two estimators deteriorates as the sample size grows. The bias adjusted estimator outperforms the naive estimator uniformly, and its performance relative to that of the naive estimator improves as the sample size increases. It is also observed that improvement of the local maximum likelihood estimator compared to the other estimators in Table 1 with known  $\mu$  and  $\sigma$  is more significant than in Table 2, in which  $\mu$  and  $\sigma$  are unknown.

For the regression model (1), the covariate values,  $x$ , were generated from a discrete uniform distribution on  $(1, 2, \dots, 10)$ . There was only one case in 1000 iterations where the consistent solution was not found when  $n = 50$ . When  $n$  is larger than 50, consistent solutions are always found in our studies. Table 3 gives the biases, standard errors, estimates of standard errors and the coverage probabilities of confidence intervals with a nominal level of 95%. The biases and the standard errors are calculated from the estimates based on the 1000 simulated samples, while the estimates of standard errors,  $\widehat{SE}$ , are calculated from the Hessian matrix of the likelihood function. The confidence intervals are constructed by  $\hat{\boldsymbol{\theta}} \pm \mathcal{Z}_{0.975} \widehat{SE}$ , where  $\hat{\boldsymbol{\theta}}$  is the local maximum likelihood estimator and  $\mathcal{Z}_{0.975}$  is the 97.5% normal quantile.

It is seen that both the biases and standard errors are small, indicating the consistency of the local likelihood estimator. Although it is evident that the standard errors are under-

estimated for small sample sizes and the coverage probabilities are lower than the nominal level, this situation ameliorates as the sample size increases.

## Acknowledgement

We thank Doctor Valerii Fedorov for suggesting to us the importance of model (1).

## Appendix A. Technical details

### Appendix A.1. Proof of Proposition 1

*Proof.* First, following the idea in Section 2.3 of Galambos (1978), for any  $t$ ,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \Pr \left[ \left\{ \frac{\log(B_0 - Y) - \log(Y - A_0) - \mu_0}{\sigma_0} \right\}_{(n)} < r_n + s_n t \right] = e^{-e^{-t}} \\ & = \lim_{n \rightarrow \infty} \Pr \left\{ \left( \frac{B_0 - Y}{Y - A_0} \right)_{(n)} < e^{\mu_0 + \sigma_0 r_n + \sigma_0 s_n t} \right\} \\ & = \lim_{n \rightarrow \infty} \Pr \left\{ \frac{B_0 - Y_{(1)}}{Y_{(1)} - A_0} < e^{\mu_0 + \sigma_0 r_n} + \sigma_0 s_n e^{\mu_0 + \sigma_0 r_n} (1 + v \sigma_0 s_n t) \right\}, \end{aligned}$$

where  $|v| \leq 1$ . Since  $v \sigma_0 s_n t \rightarrow 0$  as  $n \rightarrow \infty$ , it follows that

$$\lim_{n \rightarrow \infty} \Pr \left( \frac{B_0 - Y_{(1)}}{Y_{(1)} - A_0} < c_n + d_n t \right) = e^{-e^{-t}},$$

where  $c_n = e^{\mu_0 + \sigma_0 r_n}$  and  $d_n = \sigma_0 s_n e^{\mu_0 + \sigma_0 r_n}$ .

Then, for any  $t \neq 0$ ,

$$\begin{aligned} & \Pr \left( \frac{B_0 - Y_{(1)}}{Y_{(1)} - A_0} < c_n + d_n t \right) = \Pr \left( \frac{Y_{(1)} - A_0}{B_0 - A_0} > \frac{1}{1 + c_n + d_n t} \right) \\ & = \Pr \left[ \frac{Y_{(1)} - A_0}{B_0 - A_0} > \frac{c_n - 1}{c_n^2} - \left\{ \frac{d_n}{c_n^2} - \frac{(1 + d_n t)}{c_n^2 (1 + c_n + d_n t)} \right\} t \right]. \end{aligned}$$

It can be shown that  $[(1 + d_n t) / \{c_n^2 (1 + c_n + d_n t) t\}] / (d_n / c_n^2) \rightarrow 0$  and  $(1 / c_n^2) / (d_n / c_n^2) \rightarrow 0$  as  $n \rightarrow \infty$ . So from Lemma 2.2.2 in Galambos (1978),

$$\lim_{n \rightarrow \infty} \Pr \left\{ \frac{B_0 - Y_{(1)}}{Y_{(1)} - A_0} < c_n + d_n t \right\} = \lim_{n \rightarrow \infty} \Pr \left\{ \frac{Y_{(1)} - A_0}{B_0 - A_0} > \frac{1}{c_n} - \frac{d_n}{c_n^2} t \right\}.$$

When  $t = 0$ , the result can be verified by using the properties of the extreme order statistics of normal distribution directly. The second equation can be proved similarly.  $\square$

## Appendix A.2. Derivation of the Fisher information

The lemma below is useful in deriving the Fisher information.

**Lemma 1.** *From Lemma 2 of Stein (1981), we obtain that, if  $E|h'(Z)| < \infty$  for a normal random variable  $Z \sim N(\mu, \sigma^2)$  and some differentiable function  $h$ . Then*

$$E\{(Z - \mu)h(Z)\} = \sigma^2 E\{h'(Z)\}.$$

The log-likelihood function of model (7) based on one observation  $(\mathbf{x}, Y)$  is

$$\begin{aligned} \ell(\boldsymbol{\theta}, \mathbf{x}, Y) = & -\frac{\log(2\pi)}{2} - \log \sigma + \log(B - A) - \log(Y - A) \\ & - \log(B - Y) - \frac{\{\log(B - Y) - \log(Y - A) - \mathbf{x}^\top \boldsymbol{\beta}\}^2}{2\sigma^2} \end{aligned}$$

for  $Y \in (A, B)$  and 0 otherwise. By direct calculation,

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}, \mathbf{x}, Y)}{\partial A} &= \frac{1}{Y - A + \frac{(Y-A)^2}{B-Y}} - \frac{\{\log(B - Y) - \log(Y - A) - \mathbf{x}^\top \boldsymbol{\beta}\}}{\sigma^2(Y - A)}, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{x}, Y)}{\partial A^2} &= \frac{1 + 2\frac{Y-A}{B-Y}}{\left\{Y - A + \frac{(Y-A)^2}{B-Y}\right\}^2} - \frac{\{\log(B - Y) - \log(Y - A) - \mathbf{x}^\top \boldsymbol{\beta}\}}{\sigma^2(Y - A)^2} - \frac{1}{\sigma^2(Y - A)^2} \\ &= \frac{e^{2Z} + 2e^Z}{(B - A)^2} - \frac{(Z - \mathbf{x}^\top \boldsymbol{\beta})(1 + e^Z)^2}{\sigma^2(B - A)^2} - \frac{(1 + e^Z)^2}{\sigma^2(B - A)^2}. \end{aligned}$$

Then from LEMMA 1,

$$\begin{aligned} E \left\{ \frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{x}, Y)}{\partial A^2} \right\} &= -E \left\{ \frac{e^{2Z}}{(B - A)^2} + \frac{1 + 2e^Z + e^{2Z}}{\sigma^2(B - A)^2} \right\} \\ &= -\frac{e^{2\mathbf{x}^\top \boldsymbol{\beta} + 2\sigma^2}}{(B - A)^2} - \frac{1 + 2e^{\mathbf{x}^\top \boldsymbol{\beta} + \frac{\sigma^2}{2}} + e^{2\mathbf{x}^\top \boldsymbol{\beta} + 2\sigma^2}}{\sigma^2(B - A)^2}. \end{aligned}$$

Other elements of the Fisher information can be derived similarly.

## Appendix A.3. Proof of Theorems for the Regression Model

The following assumptions are required in this section.

**Assumption 1.**  $\sup_i |\mathbf{x}_i| < \infty$ .

**Assumption 2.** *The following quantities converge as  $n \rightarrow \infty$ ,  $n^{-1} \sum_{i=1}^n \mathbf{x}_i$ ,  $n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$ ,  $n^{-1} \sum_{i=1}^n c_i$ ,  $n^{-1} \sum_{i=1}^n c_i^{-1}$ ,  $n^{-1} \sum_{i=1}^n c_i \mathbf{x}_i$ ,  $n^{-1} \sum_{i=1}^n c_i^{-1} \mathbf{x}_i$ ,  $n^{-1} \sum_{i=1}^n c_i^2$  and  $n^{-1} \sum_{i=1}^n c_i^{-2}$ .*

**Assumption 3.**  $(\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)^\top$  is full rank.

The proof of Theorem 1 begins with some lemmas.

**Lemma 2.** For constant sequences  $v_n \downarrow v$  and  $w_n \uparrow w$  as  $n \rightarrow \infty$ , let  $\xi_{v_n} \in (v_{n+1}, v_n)$  and  $\xi_{w_n} \in (w_n, w_{n+1})$ . If a continuous function sequence  $f_n(\cdot) > 0$ , which is decreasing in  $n$ , satisfies  $n^{1+\alpha} f_n(\xi_{v_n}) \rightarrow 0$  and  $n^{1+\alpha} f_n(\xi_{w_n}) \rightarrow 0$  for  $\alpha > 0$  as  $n \rightarrow \infty$ , then

$$\limsup_n \int_{v_n}^{w_n} f_n(x) dx < \infty.$$

*Proof.* Let  $S_n = \int_{v_n}^{w_n} f_n(x) dx$ . Then

$$\begin{aligned} S_n - S_{n-1} &= \int_{v_n}^{w_n} f_n(x) dx - \int_{v_{n-1}}^{w_{n-1}} f_{n-1}(x) dx \\ &\leq (v_{n-1} - v_n) f_{n-1}(\xi_{v_{n-1}}) + (w_n - w_{n-1}) f_{n-1}(\xi_{w_{n-1}}) \\ &= \frac{(v_{n-1} - v_n) n^{1+\alpha} f_{n-1}(\xi_{v_{n-1}}) + (w_n - w_{n-1}) n^{1+\alpha} f_{n-1}(\xi_{w_{n-1}})}{n^{1+\alpha}} = o\left(\frac{1}{n^{1+\alpha}}\right). \end{aligned}$$

So  $\limsup_n S_n = \limsup_n \sum_{i=1}^n (S_n - S_{n-1})$  is finite.  $\square$

**Lemma 3.** For any  $\alpha > 0$ , let  $\delta_n = n^{-\alpha}$ . Then for any  $k_1 \geq 0$  and  $k_2 > 0$ , there exists a constant  $M$  such that

$$\lim_{n \rightarrow \infty} Pr \left\{ \frac{1}{n} \sum_{i=1}^n \frac{|\log(B - Y_i)|^{k_1}}{(Y_i - A)^{k_2}} < M \right\} = 1, \quad \lim_{n \rightarrow \infty} Pr \left\{ \frac{1}{n} \sum_{i=1}^n \frac{|\log(Y_i - A)|^{k_1}}{(B - Y_i)^{k_2}} < M \right\} = 1 \quad (10)$$

uniformly in  $A$  and  $B$  such that  $|A - A_0| < \delta_n$  and  $|B - B_0| < \delta_n$ .

*Proof.* We give the details of proof for the first quantity in (10). The proof for the other one is similar.

$$\begin{aligned} \frac{\log^{k_1}(B - Y_i)}{(Y_i - A)^{k_2}} &= \frac{|\log(B - B_0 + B_0 - Y_i)|^{k_1}}{(Y_i - A_0 + A_0 - A)^{k_2}} I(Y_i - A_0 > 2\delta_n, B_0 - Y_i > 2\delta_n) + o_p(1) \\ &= \frac{|\log(B - B_0 + B_0 - Y_i)|^{k_1}}{(Y_i - A_0 + A_0 - A)^{k_2}} I(Y_i - A_0 > 2\delta_n, 1 - 2\delta_n > B_0 - Y_i > 2\delta_n) \\ &\quad + \frac{|\log(B - B_0 + B_0 - Y_i)|^{k_1}}{(Y_i - A_0 + A_0 - A)^{k_2}} I(Y_i - A_0 > 2\delta_n, B_0 - Y_i > 1) + o_p(1) \\ &< \frac{I(Y_i - A_0 > 2\delta_n, 1 - 2\delta_n > B_0 - Y_i > 2\delta_n)}{(B_0 - Y_i - \delta_n)^{k_1} (Y_i - A_0 - \delta_n)^{k_2}} \\ &\quad + \frac{(B_0 - Y_i + \delta_n)^{k_1}}{(Y_i - A_0 - \delta_n)^{k_2}} I(Y_i - A_0 > 2\delta_n, B_0 - Y_i > 1) + o_p(1) \\ &< \frac{I(B_0 - 2\delta_n > Y_i > A_0 + 2\delta_n)}{(B_0 - Y_i - \delta_n)^{k_1} (Y_i - A_0 - \delta_n)^{k_2}} \\ &\quad + \frac{(B_0 - A_0 + 1)^{k_1}}{(Y_i - A_0 - \delta_n)^{k_2}} I(B_0 - 1 > Y_i > A_0 + 2\delta_n) + o_p(1) \\ &= C_{in1} + C_{in2} + o_p(1). \end{aligned}$$

Note that

$$\begin{aligned}
& (2\pi)^{\frac{1}{2}} E(C_{in1}) \\
&= \frac{1}{\sigma_0} \int_{A_0+2\delta_n}^{B_0-2\delta_n} \frac{1}{(B_0-y-\delta_n)^{k_1}(y-A_0-\delta_n)^{k_2}} \times \\
&\quad \frac{B_0-A_0}{(B_0-y)(y-A_0)} \exp \left[ -\frac{\left\{ \log \left( \frac{B_0-y}{y-A_0} \right) - \mathbf{x}_i^T \boldsymbol{\beta}_0 \right\}^2}{2\sigma_0^2} \right] dy \\
&\leq \frac{1}{\sigma_0} \int_{A_0+2\delta_n}^{B_0-2\delta_n} \frac{1}{(B_0-y-\delta_n)^{k_1+1}(y-A_0-\delta_n)^{k_2+1}} \exp \left[ -\frac{\frac{1}{4} \left\{ \log \left( \frac{B_0-y}{y-A_0} \right) \right\}^2 - (\mathbf{x}_i^T \boldsymbol{\beta}_0)^2}{2\sigma_0^2} \right] dy \\
&= \frac{\exp \left\{ \frac{(\mathbf{x}_i^T \boldsymbol{\beta}_0)^2}{2\sigma_0^2} \right\}}{\sigma_0} \int_{A_0+2\delta_n}^{B_0-2\delta_n} \frac{1}{(B_0-y-\delta_n)^{k_1+1}(y-A_0-\delta_n)^{k_2+1}} \exp \left[ -\frac{\left\{ \log \left( \frac{B_0-y}{y-A_0} \right) \right\}^2}{8\sigma_0^2} \right] dy.
\end{aligned}$$

From Lemma 2,  $\limsup_n E(C_{in1})$  is bounded by a finite constant, say  $C_1$ . Similarly, it can also be shown that  $\limsup_n E(C_{in2})$  is bounded by a finite constant, say  $C_2$ . So using the formula  $X_n = E(X_n) + O_P\{\text{var}(X_n)^{1/2}\}$ , we have

$$\frac{1}{n} \sum_{i=1}^n \frac{|\log(B-Y_i)|^{k_1}}{(Y_i-A)^{k_2}} = \frac{1}{n} \sum_{i=1}^n E(C_{in1}) + \frac{1}{n} \sum_{i=1}^n E(C_{in2}) + O_P\left(n^{-\frac{1}{2}}\right) + o_P(1).$$

Thus any  $M$  that is greater than  $C_1 + C_2$  satisfies the requirement.  $\square$

**Lemma 4.** *If Assumption 1- 3 hold, then  $-n^{-1}\partial^2\ell_n(\boldsymbol{\theta})/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T) \rightarrow \mathcal{I}(\boldsymbol{\theta}_0)$  in probability uniformly over  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_n$ .*

*Proof.* The first element of  $\partial^2\ell_n(\boldsymbol{\theta})/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T)$  is

$$\frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial A^2} = \sum_{i=1}^n \left[ \frac{1 - \frac{1}{\sigma^2}}{(Y_i - A)^2} - \frac{1}{(B - A)^2} - \frac{\left\{ \log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^T \boldsymbol{\beta} \right\}}{\sigma^2(Y_i - A)^2} \right].$$

So it is straightforward to get

$$\begin{aligned}
\frac{1}{n} \left| \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial A^2} - \frac{\partial^2\ell_n(\boldsymbol{\theta}_0)}{\partial A^2} \right| &\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{(B-A)^2} - \frac{1}{(B_0-A_0)^2} \right| + \frac{1}{n} \sum_{i=1}^n \left| \frac{1 + \frac{1}{\sigma^2}}{(Y_i-A)^2} - \frac{1 + \sigma_0^2}{(Y_i-A_0)^2} \right| \\
&\quad + \frac{1}{n} \sum_{i=1}^n \left| \frac{\left\{ \log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^T \boldsymbol{\beta} \right\}}{\sigma^2(Y_i-A)^2} - \frac{\left\{ \log \left( \frac{B_0-Y_i}{Y_i-A_0} \right) - \mathbf{x}_i^T \boldsymbol{\beta}_0 \right\}}{\sigma_0^2(Y_i-A_0)^2} \right| \\
&= \Delta_1 + \Delta_2 + \Delta_3.
\end{aligned}$$

$\Delta_1$  goes to 0 as  $\delta_n$  goes to 0. By straightforward but tedious calculation, we obtain

$$\begin{aligned}
\Delta_3 &\leq \frac{1}{n} \sum_{i=1}^n \left| \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right| \left| \frac{1}{\sigma^2(Y_i - A)^2} - \frac{1}{\sigma_0^2(Y_i - A_0)^2} \right| \\
&\quad + \frac{1}{n} \sum_{i=1}^n \left| \frac{\log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma_0^2(Y_i - A_0)^2} - \frac{\log \left( \frac{B_0 - Y_i}{Y_i - A_0} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_0}{\sigma_0^2(Y_i - A_0)^2} \right| \\
&\leq \left| \frac{1}{\sigma^2} - \frac{1}{\sigma_0^2} \right| \times \frac{1}{n} \sum_{i=1}^n \frac{\left| \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right|}{(Y_i - A)^2} \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{\left| \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right|}{\sigma_0^2} \times \left| \frac{1}{(Y_i - A)^2} - \frac{1}{(Y_i - A_0)^2} \right| \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{\left| \log \left( \frac{B - Y_i}{B_0 - Y_i} \right) \right|}{\sigma_0^2(Y_i - A_0)^2} + \frac{1}{n} \sum_{i=1}^n \frac{\left| \log \left( \frac{Y_i - A}{Y_i - A_0} \right) \right|}{\sigma_0^2(Y_i - A_0)^2} + \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{\sigma_0^2(Y_i - A_0)^2} \\
&\leq \frac{1}{n} \sum_{i=1}^n \frac{\left| \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right|}{(Y_i - A)^2} \times \left| \frac{1}{\sigma^2} - \frac{1}{\sigma_0^2} \right| + \frac{4B|A - A_0|}{\sigma_0^2} \times \frac{1}{n} \sum_{i=1}^n \frac{\left| \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right|}{(Y_i - A)^2(Y_i - A_0)^2} \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{|B - B_0|}{\sigma_0^2(B^* - Y_i)(Y_i - A_0)^2} + \frac{1}{n} \sum_{i=1}^n \frac{|A - A_0|}{\sigma_0^2(Y_i - A^*)(Y_i - A_0)^2} + \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{\sigma_0^2(Y_i - A_0)^2} \\
&= \Delta_{3.1} + \Delta_{3.2} + \Delta_{3.3} + \Delta_{3.4} + \Delta_{3.5},
\end{aligned}$$

where  $A^*$  is between  $A$  and  $A_0$  and  $B^*$  is between  $B$  and  $B_0$ . Now we look into each term in the last equation above.

$$\Delta_{3.2} \leq \frac{2B|A - A_0|}{\sigma_0^2} \times \frac{1}{n} \sum_{i=1}^n \left| \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right| \left\{ \frac{1}{(Y_i - A)^4} + \frac{1}{(Y_i - A_0)^2} \right\}. \quad (11)$$

The right hand side term in (11) goes to 0 in probability uniformly since the second factor is bound with probability tending to 1 by Lemma 3 and the boundedness of  $x_i$ . Similarly,  $\Delta_{3.1}$ ,  $\Delta_{3.3}$ ,  $\Delta_{3.4}$  and  $\Delta_{3.5}$  can be shown to go to 0 in probability uniformly which implies  $\Delta_3$  goes to 0 in probability uniformly. Similarly but more easily,  $\Delta_1$  and  $\Delta_2$  can be shown to converge to 0 in probability uniformly, which implies  $n^{-1} |\partial^2 \ell_n(\boldsymbol{\theta}) / \partial A^2 - \partial^2 \ell_n(\boldsymbol{\theta}_0) / \partial A^2| \rightarrow 0$  in probability uniformly. By similar arguments, other components of  $\partial^2 \ell_n(\boldsymbol{\theta}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top)$  can be shown to have the same property. This implies that  $-n^{-1} \partial^2 \ell_n(\boldsymbol{\theta}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top) \rightarrow \mathcal{I}(\boldsymbol{\theta}_0)$  in probability uniformly over  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_n$ .  $\square$

The following lemma is the Lemma 5 of Smith (1985). We state it for integrity and skip the proof.

**Lemma 5.** *Let  $h$  be a continuously differentiable real-valued function of  $p + 1$  real variables and let  $H$  denote the gradient vector of  $h$ . Suppose that the scalar product of  $\mathbf{u}$  and  $H(\mathbf{u})$  is negative whenever  $\|\mathbf{u}\| = 1$ . Then  $h$  has a local maximum, at which  $H = 0$ , for some  $\mathbf{u}$  with  $\|\mathbf{u}\| < 1$ .*

of Theorem 1. It suffices to show for any  $\epsilon$ , there exists a constant  $c$  such that

$$\Pr \left\{ \mathbf{u}^\top \frac{\partial \ell_n(\boldsymbol{\theta}_0 + n^{-1/2} c \mathbf{u})}{\partial \boldsymbol{\theta}} < 0 \right\} > 1 - \epsilon \quad (12)$$

for any vector  $\mathbf{u}$  such that  $\|\mathbf{u}\| = 1$ . Using Taylor's expansion,

$$\begin{aligned} \frac{\partial \ell_n(\boldsymbol{\theta}_0 + n^{-1/2} c \mathbf{u})}{\partial \boldsymbol{\theta}} &= \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + c n^{-1/2} \frac{\partial^2 \ell_n(\boldsymbol{\theta}_0 + n^{-1/2} c \mathbf{u}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \mathbf{u} \\ &= \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - c n^{1/2} \mathcal{I}(\boldsymbol{\theta}_0) \mathbf{u} + n^{1/2} \epsilon_{n, \mathbf{u}}, \end{aligned}$$

where  $\mathbf{u}^*$  is a vector satisfying  $\|\mathbf{u}^*\| \leq 1$  and, by Lemma 3,  $\epsilon_{n, \mathbf{u}} \rightarrow 0$  in probability uniformly over  $\|\mathbf{u}\| \leq 1$  as  $n \rightarrow \infty$ . It follows that

$$n^{-1/2} \mathbf{u}^\top \frac{\partial \ell_n(\boldsymbol{\theta}_0 + n^{-1/2} \mathbf{u})}{\partial \boldsymbol{\theta}} = n^{-1/2} \mathbf{u}^\top \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - c \mathbf{u}^\top \mathcal{I}(\boldsymbol{\theta}_0) \mathbf{u} + \mathbf{u}^\top \epsilon_{n, \mathbf{u}}. \quad (13)$$

Note that  $n^{-1/2} \mathbf{u}^\top \partial \ell_n(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta}$  is  $O_P(1)$ . So the second term dominates the first term in (13) for large enough  $c$ . This proves equation (12) and the result follows from Lemma 5.  $\square$

of Theorem 2, part 1. For any  $\boldsymbol{\theta}_1 \in S$ ,  $E \ell_n(\boldsymbol{\theta}_1) < \infty$ , so  $E[\ell_n(\boldsymbol{\theta}_1) - \ell_n(\boldsymbol{\theta}_0)] < 0$  by Jensen's inequality. This implies the existence of  $\xi_{\boldsymbol{\theta}_1}$  such that

$$\lim_{n \rightarrow \infty} \Pr \{ \ell_n(\boldsymbol{\theta}_1) - \ell_n(\boldsymbol{\theta}_0) < -\xi_{\boldsymbol{\theta}_1} \} = 1.$$

For  $\boldsymbol{\theta}$  and  $\eta$  such that  $|\boldsymbol{\theta} - \boldsymbol{\theta}_1| < \eta < |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0| - \delta$ ,

$$\begin{aligned} |\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_1)| &\leq |\log \sigma - \log \sigma_1| + \frac{1}{n} \sum_{i=1}^n \left| \log \left( \frac{1}{B - Y_i} + \frac{1}{Y_i - A} \right) - \log \left( \frac{1}{B_1 - Y_i} + \frac{1}{Y_i - A_1} \right) \right| \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left| \frac{\left\{ \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right\}^2}{\sigma^2} - \frac{\left\{ \log \left( \frac{B_1 - Y_i}{Y_i - A_1} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2}{\sigma_1^2} \right| \\ &= \Delta_4 + \Delta_5 + \Delta_6. \end{aligned}$$

$\Delta_4$  can be made smaller than  $\xi_{\boldsymbol{\theta}_1}/4$  by choosing  $\eta$  small enough. By the mean value theorem,

$$\begin{aligned} \Delta_5 &= \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{B^* - Y_i} \frac{Y_i - A^*}{B^* - A^*} (B - B_1) + \frac{1}{Y_i - A^*} \frac{B^* - Y_i}{B^* - A^*} (A - A_1) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \left\{ \frac{B_0 - A_1 + \eta |B - B_1|}{B_0 - A_0} \frac{1}{B_0 - Y_i} + \frac{B_1 - A_0 + \eta |A - A_1|}{B_0 - A_0} \frac{1}{Y_i - A_0} \right\}, \end{aligned}$$

for some  $A^*$  between  $A_0$  and  $A_1$  and  $B^*$  between  $B_0$  and  $B_1$ . So  $E(\Delta_5)$  can be made arbitrarily small by choosing small enough  $\eta$ , which implies

$$\lim_{n \rightarrow \infty} \Pr \left( \Delta_5 < \frac{\xi_{\boldsymbol{\theta}_1}}{4} \right) = 1$$

for small enough  $\eta$ .

$$\begin{aligned}
\Delta_6 &\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{\left\{ \log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right\}^2}{\sigma^2} - \frac{\left\{ \log \left( \frac{B_1-Y_i}{Y_i-A_1} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2}{\sigma_1^2} \right| \\
&\quad + \frac{1}{n} \sum_{i=1}^n \left\{ \log \left( \frac{B_1-Y_i}{Y_i-A_1} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2 \left| \frac{1}{\sigma^2} - \frac{1}{\sigma_1^2} \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{\log \left( \frac{B-Y_i}{Y_i-A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} + \log \left( \frac{B_1-Y_i}{Y_i-A_1} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1}{\sigma^2} \right| \\
&\quad \times \left\{ \frac{|A-A_0|}{A_0-Y_i} + \frac{|B-B_0|}{B_0-Y_i} + |\mathbf{x}_i^\top \boldsymbol{\beta} - \mathbf{x}_i^\top \boldsymbol{\beta}_1| \right\} \\
&\quad + \frac{1}{n} \sum_{i=1}^n \left\{ \log \left( \frac{B_1-Y_i}{Y_i-A_1} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2 \left| \frac{1}{\sigma^2} - \frac{1}{\sigma_1^2} \right|.
\end{aligned}$$

So, for small enough  $\eta$ , we obtain

$$\lim_{n \rightarrow \infty} \Pr \left( \Delta_6 < \frac{\xi_{\boldsymbol{\theta}_1}}{4} \right) = 1.$$

Combining results for  $\Delta_4$ ,  $\Delta_5$  and  $\Delta_6$ ,

$$\lim_{n \rightarrow \infty} \Pr \left\{ \sup_{|\boldsymbol{\theta} - \boldsymbol{\theta}_1| < \eta} \ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_1) < -\frac{\xi_{\boldsymbol{\theta}_1}}{4} \right\} = 1.$$

For any compact set  $K$ ,  $S_\delta \cap K$  can be covered by a finite number of neighborhoods of points in  $S_\delta$ , so it follows that

$$\lim_{n \rightarrow \infty} \Pr \left\{ \sup_{S_\delta \cap K} \ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_1) < -\xi_m \right\} = 1.$$

□

of *Theorem 2, part 2*. First, if  $A_0$  and  $B_0$  are known, model (7) can be transformed to a linear model with normal random error with unknown mean and variance. It follows that

$$\lim_{n \rightarrow \infty} \Pr \left\{ \sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \delta, |\sigma - \sigma_0| > \delta} \ell_n(\boldsymbol{\beta}, \sigma, A_0, B_0) - \ell_n(\boldsymbol{\theta}_0) < -\xi \right\} = 1. \quad (14)$$

For  $\boldsymbol{\beta}_1$ ,  $\sigma_1$ ,  $\eta$  and  $(\boldsymbol{\beta}, \sigma, A, B) \in T$  such that  $(\boldsymbol{\beta}_1, \sigma_1, A, B) \in T$ ,  $\|\boldsymbol{\beta} - \boldsymbol{\beta}_1\| < \eta$ ,  $|\sigma - \sigma_1| < \eta$

and  $\delta < \eta$ ,

$$\begin{aligned}
& |\ell_n(\boldsymbol{\beta}, \sigma, A, B) - \ell_n(\boldsymbol{\beta}_1, \sigma_1, A_0, B_0)| \\
& \leq |\log \sigma - \log \sigma_1| + |\log(B - A) - \log(B_0 - A_0)| \\
& \quad + \frac{1}{n} \sum_{i=1}^n |\log(B - Y_i) - \log(B_0 - Y_i)| + \frac{1}{n} \sum_{i=1}^n |\log(Y_i - A) - \log(Y_i - A_0)| \\
& \quad + \frac{1}{2n} \sum_{i=1}^n \left| \frac{\left\{ \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right\}^2}{\sigma^2} - \frac{\left\{ \log \left( \frac{B_0 - Y_i}{Y_i - A_0} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2}{\sigma_1^2} \right| \\
& = \Delta_7 + \Delta_8 + \Delta_9 + \Delta_{10} + \Delta_{11}.
\end{aligned} \tag{15}$$

The terms  $\Delta_7$  and  $\Delta_8$  can be made smaller than  $\xi/8$  by choosing  $\eta$  small enough. By the mean value theorem,

$$\Delta_9 = \frac{1}{n} \sum_{i=1}^n \left| \frac{B - B_0}{B^* - Y_i} \right| \leq \frac{|B - B_0|}{n} \sum_{i=1}^n \frac{1}{\min(B, B_0) - Y_i}$$

with probability tending to 1. If  $B \geq B_0$ ,  $n^{-1} \sum_{i=1}^n 1/|\min(B, B_0) - Y_i| \leq n^{-1} \sum_{i=1}^n 1/(B_0 - Y_i)$  which goes to the limit of  $(1 + n^{-1} \sum_{i=1}^n e^{-\mathbf{x}_i^\top \boldsymbol{\beta} + \sigma^2/2}) / (B_0 - A_0)$  in probability. If  $B_0 - \delta_n < B < B_0$ , Lemma 3 provides that there exists some constant  $M^*$  such that

$$\lim_{n \rightarrow \infty} \Pr \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{|B - Y_i|} < M^* \right) = 1$$

for small enough  $\eta$ . This implies that for small enough  $\eta$ ,

$$\lim_{n \rightarrow \infty} \Pr \left( \Delta_9 < \frac{\xi}{8} \right) = 1. \tag{16}$$

The same result can be found for  $\Delta_{10}$  using similar arguments.

$$\begin{aligned}
\Delta_{11} & \leq \frac{1}{2n} \sum_{i=1}^n \left| \frac{\left\{ \log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} \right\}^2}{\sigma^2} - \frac{\left\{ \log \left( \frac{B_0 - Y_i}{Y_i - A_0} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2}{\sigma^2} \right| \\
& \quad + \frac{1}{2n} \sum_{i=1}^n \left\{ \log \left( \frac{B_0 - Y_i}{Y_i - A_0} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2 \left| \frac{1}{\sigma^2} - \frac{1}{\sigma_1^2} \right| \\
& \leq \frac{1}{2n} \sum_{i=1}^n \left| \frac{\log \left( \frac{B - Y_i}{Y_i - A} \right) - \mathbf{x}_i^\top \boldsymbol{\beta} + \log \left( \frac{B_0 - Y_i}{Y_i - A_0} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1}{\sigma^2} \right| \times |\mathbf{x}_i^\top \boldsymbol{\beta} - \mathbf{x}_i^\top \boldsymbol{\beta}_1| \\
& \quad + \frac{1}{2n} \sum_{i=1}^n \left\{ \log \left( \frac{B_0 - Y_i}{Y_i - A_0} \right) - \mathbf{x}_i^\top \boldsymbol{\beta}_1 \right\}^2 \left| \frac{1}{\sigma^2} - \frac{1}{\sigma_1^2} \right|.
\end{aligned}$$

So we obtain, for small enough  $\eta$ ,

$$\lim_{n \rightarrow \infty} \Pr \left( \Delta_{11} < \frac{\xi}{8} \right) = 1. \quad (17)$$

Combining (14), (15), (16) and (17), we have

$$\lim_{n \rightarrow \infty} \Pr \left\{ \sup \ell_n(a, b, \sigma, A, B) - \ell_n(\boldsymbol{\theta}_0) < -\frac{3\xi}{8} \right\} = 1,$$

where the supremum is taken over all  $\boldsymbol{\theta}$  satisfying  $(\boldsymbol{\beta}_1, \sigma_1, A, B) \in T$ ,  $\|\boldsymbol{\beta} - \boldsymbol{\beta}_1\| < \eta$  and  $|\sigma - \sigma_1| < \eta$  for fixed  $\boldsymbol{\beta}_1$  and  $\sigma_1$ . This result can be extended directly to any finite set of values of  $\boldsymbol{\beta}_1$  and  $\sigma_1$ , and then to any compact sets of values of  $\boldsymbol{\beta}_1$  and  $\sigma_1$ .  $\square$

of Theorem 3. By Taylor expansion,

$$0 = \frac{\partial \ell_n(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \ell_n(\hat{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where  $\hat{\boldsymbol{\theta}}^*$  is between  $\boldsymbol{\theta}_0$  and  $\hat{\boldsymbol{\theta}}_n$ . From Lemma 4,  $n^{-1} \partial^2 \ell_n(\hat{\boldsymbol{\theta}}^*) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T) \rightarrow -\mathcal{I}(\boldsymbol{\theta}_0)$  in probability. So

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \{\mathcal{I}(\boldsymbol{\theta}_0)\}^{-1} n^{-1/2} \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_P(1). \quad (18)$$

Note  $n^{-1/2} \partial \ell_n(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta} = n^{-1/2} \sum_{i=1}^n \partial \ell(\boldsymbol{\theta}_0, \mathbf{x}_i, Y_i) / \partial \boldsymbol{\theta}$  is summation of independent random vectors and its variance converges to  $\mathcal{I}(\boldsymbol{\theta}_0)$ . Also we have for  $t > 0$ ,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n E \left[ \left\| \frac{\partial \ell(\boldsymbol{\theta}_0, \mathbf{x}_i, Y_i)}{\partial \boldsymbol{\theta}} \right\|^2 I \left\{ \left\| \frac{\partial \ell(\boldsymbol{\theta}_0, \mathbf{x}_i, Y_i)}{\partial \boldsymbol{\theta}} \right\| > n^{1/2} \epsilon \right\} \right] \\ & \leq \frac{1}{n} \frac{1}{(n^{1/2} \epsilon)^t} \sum_{i=1}^n E \left[ \left\| \frac{\partial \ell(\boldsymbol{\theta}_0, \mathbf{x}_i, Y_i)}{\partial \boldsymbol{\theta}} \right\|^{2+t} I \left\{ \left\| \frac{\partial \ell(\boldsymbol{\theta}_0, \mathbf{x}_i, Y_i)}{\partial \boldsymbol{\theta}} \right\| > n^{1/2} \epsilon \right\} \right] \\ & \leq \frac{1}{n} \frac{1}{(n^{1/2} \epsilon)^t} \sum_{i=1}^n E \left[ \left\| \frac{\partial \ell(\boldsymbol{\theta}_0, \mathbf{x}_i, Y_i)}{\partial \boldsymbol{\theta}} \right\|^{2+t} \right] \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

By the multivariate central limit theorem (cf. Rao, 1973; Serfling, 1980),

$$n^{-1/2} \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \rightarrow N(0, \mathcal{I}(\boldsymbol{\theta}_0)) \quad (19)$$

in distribution. Combining (18), (19) and applying Slutsky's theorem, the result follows.  $\square$

# Acknowledgements

## References

- Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)* **45**, 394–403.
- Cheng, R. C. H. and Iles, T. C. (1987). Corrected maximum likelihood in non-regular problems. *Journal of the Royal Statistical Society. Series B (Methodological)* **49**, 95–101.
- Cheng, R. C. H. and Traylor, L. (1995). Non-regular maximum likelihood problems. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 3–44.
- Cohen, A. C. (1951). Estimating parameters of logarithmic-normal distributions by maximum likelihood. *Journal of the American Statistical Association* **46**, 206–212.
- Cohen, A. C. and Whitten, B. J. (1980). Estimation in the three-parameter lognormal distribution. *Journal of the American Statistical Association* **75**, 399–404.
- DeLean, A., Munson, P. J., and Rodbard, D. (1978). Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *American Journal of Physiology: Endocrinology and Metabolism* **235**, E97–E102.
- Dragalin, V., Hsuan, F., and Padmanabhan, S. K. (2007). Adaptive designs for dose-finding studies based on sigmoid  $e_{\max}$  model. *Journal of Biopharmaceutical Statistics* **17**, 1051–1070.
- Galambos, J. (1978). *The asymptotic theory of extreme order statistics*. John Wiley & Sons, New York.
- Giesbrecht, F. and Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter lognormal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)* **38**, 257–264.
- Harter, H. L. and Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association* **61**, 842–851.
- Hill, B. M. (1963). The three-parameter lognormal distribution and bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association* **58**, 72–84.
- Holford, N. H. and Sheiner, L. B. (1981). Understanding the dose-effect relationship: Clinical application of pharmacokinetic-pharmacodynamic models. *Clinical Pharmacokinetics* **6**, 429–453.
- Leonov, S. and Miller, S. (2009). An adaptive optimal design for the  $e_{\max}$  model and its application in clinical trials. *Journal of Biopharmaceutical Statistics* **19**, 360–385.

- MacDougall, J. (2006). Analysis of dose-response studies  $e_{\max}$ . In N. Ting, ed., *Dose Finding in Drug Development*, 127–145. Springer, New York.
- Nix, B. and Wild, D. (2001). Calibration curve-fitting. In D. Wild, ed., *The Immunoassay Handbook*, 198–210. Nature Publishing Group, New York.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. John Wiley & Sons, 2nd edn.
- Ratkowsky, D. A. and Reedy, T. J. (1986). Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. *Biometrics* **42**, 575–582.
- Sebaugh, J. L. (2011). Guidelines for accurate  $ec_{50}/ic_{50}$  estimation. *Pharmaceutical Statistics* **10**, 128–134.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.
- Smith, R. L. (1994). Nonregular regression. *Biometrika* **81**, 173–183.
- Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* **9**, 1135–1151.
- Vølund, A. (1978). Application of the four-parameter logistic model to bioassay: Comparison with slope ratio and parallel line models. *Biometrics* **34**, 357–365.
- Weiss, L. and Wolfowitz, J. (1973). Maximum likelihood estimation of a translation parameter of a truncated distribution. *The Annals of Statistics* **1**, 944–947.
- Woodroffe, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. *The Annals of Mathematical Statistics* **43**, 113–122.
- Woodroffe, M. (1974). Maximum likelihood estimation of translation parameter of truncated distribution ii. *The Annals of Statistics* **2**, 474–488.

Table 1: Relative mean square errors for the one sample case when  $\mu$  and  $\sigma$  are known. In this table, Adjusted is the bias adjusted estimator relative to the local maximum likelihood estimator; Naive is the naive estimator relative to the local maximum likelihood estimator.

Parameter		$\theta_0 = (1, 0.5, 0, 10)$		$\theta_0 = (-1, 0.5, 0, 5)$	
		Adjusted	Naive	Adjusted	Naive
$n = 50$	$A$	7.12	20.65	3.82	36.41
	$B$	3.62	35.03	6.64	19.29
$n = 100$	$A$	9.85	33.20	5.43	64.44
	$B$	5.13	62.55	9.15	31.13
$n = 400$	$A$	19.44	88.55	11.06	199.12
	$B$	10.69	191.09	17.94	81.85

Table 2: Relative mean square errors for the one sample case when  $\mu$  and  $\sigma$  are unknown. In this table, Adjusted is the bias adjusted estimator relative to the local maximum likelihood estimator; Naive is the naive estimator relative to the local maximum likelihood estimator; NS is the number of cases out of 1000 iterations that consistent solutions to the likelihood equation are not found.

Parameter		$\theta_0 = (1, 0.5, 0, 10)$		$\theta_0 = (-1, 0.5, 0, 5)$	
		Adjusted	Naive	Adjusted	Naive
$n = 50$		NS= 6		NS= 3	
	$\mu$	1.22	1.38	1.30	1.45
	$\sigma$	1.63	2.21	1.67	2.26
	$A$	1.37	1.93	1.47	1.84
	$B$	1.38	1.75	1.27	1.81
$n = 100$		NS= 2		NS= 0	
	$\mu$	1.47	1.73	1.66	1.92
	$\sigma$	2.68	4.30	2.91	4.67
	$A$	1.81	2.63	1.90	2.43
	$B$	1.66	2.15	1.77	2.60
$n = 400$		NS= 0		NS= 0	
	$\mu$	2.01	2.62	2.66	3.46
	$\sigma$	4.08	8.53	6.01	12.57
	$A$	2.72	4.53	3.10	4.38
	$B$	2.14	3.05	3.42	5.73

Table 3: Biases ( $\times 10^3$ ), standard errors ( $\times 10^3$ ), estimates of standard errors ( $\times 10^3$ ) and coverage probabilities ( $\times 10^2$ ) for the regression model. In this table, SE is standard errors;  $\widehat{SE}$  is estimates of standard errors; CP is coverage probabilities.

		$\boldsymbol{\theta}_0 = (5, -1, 0.5, 0, 10)$				
		$a$	$b$	$\sigma$	$A$	$B$
$n = 50$	Bias	234	40	53	42	15
	SE	270	46	65	53	19
	SEE	248	42	61	42	16
	CP	90.3	90.2	92.5	84.2	84.3
$n = 100$	Bias	142	24	35	25	9
	SE	170	29	44	31	11
	SEE	170	29	43	29	11
	CP	93.8	93.6	93.3	89.2	89.6
$n = 400$	Bias	67	11	17	12	4
	SE	83	14	22	14	5
	SEE	82	14	21	14	5
	CP	94.5	95	94.5	93.3	94.4