A BALLOONED BETA-LOGISTIC MODEL

A Thesis presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
Min Yi
Dr. Nancy Flournoy, Dissertation Supervisor
May 2015
The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

**A BALLOONED BETA-LOGISTIC MODEL**

presented by Min Yi,
a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

________________________________________
Dr. Nancy Flournoy

________________________________________
Dr. Wade V. Welshons

________________________________________
Dr. Jianguo Sun

________________________________________
Dr. Subharup Guha

________________________________________
Dr. Hongyuan Cao
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Nancy Flournoy, for her excellent guidance, caring, patience, and introducing me this challenging and interesting topic. Without her continuous encouragement and inspiration, this work will never been possible.

I truly appreciate my committee members: Dr. Tony Sun, Dr. Subharup Guha, Dr. Hongyuan Cao and Dr. Wade V. Welshons for their insightful comments and suggestion on this work.

I would especially like to thank Dr. Tony Sun for providing me endless help and guidance during my PhD study.

I am deeply grateful to Dr. Ram Tiwari, Dr. Li Zhu and Dr. Maggie Chen for offering me internship opportunities in government and industry.

Finally, I would like to thank my parents and my wife, Yarui Liu. They were always supporting me and encouraging me with their best wishes.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .......................................................... ii

LIST OF TABLES ....................................................................... v

LIST OF FIGURES ................................................................... vi

ABSTRACT .............................................................................. viii

CHAPTER

1 Introduction ................................................................. 1

2 The Ballooned Beta-Logistic Model .............................. 7

3 Parameter Estimation ................................................... 10

3.1 Estimate Response Boundaries Using the Extreme Order Statistics 11

3.2 Least Square Estimates under BBL model ................. 13

3.3 Maximum Likelihood Estimates under BBL Model .... 16

3.4 Fisher Information ........................................................ 18

3.5 Maximum Likelihood Estimates of Slope and $EC_{50}$ 19

3.6 Finding Maximum Likelihood Estimates ................. 20

3.7 Comparison of Estimators ............................................ 21

3.8 Technical Details .......................................................... 23

3.8.1 The Hessian Matrix of a Ballooned Beta-logistic Distributed Random Variable 23

3.8.2 Proof of Theorem 3.3.2 and 3.3.3 ............................ 30

4 Illustration from Assay Experiment ............................ 37
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1  Performance of $\hat{L}$ and $\hat{U}$ under the BBL model</td>
<td>24</td>
</tr>
<tr>
<td>3.2  Performance of $\hat{S}$ and $\hat{EC}_{50}$ under the BBL model</td>
<td>25</td>
</tr>
<tr>
<td>4.1  Parameter Estimates in Exploring the Need for $\beta_2$</td>
<td>41</td>
</tr>
<tr>
<td>4.2  Boundary Estimates for Each Plate under the 4PL and BBL Models</td>
<td>44</td>
</tr>
<tr>
<td>4.3  Reference Failure Detection under the 4PL and BBL Models</td>
<td>45</td>
</tr>
<tr>
<td>4.4  Simultaneous Multiple Comparisons of Slopes and $EC_{50}$ values from ELISA Plates</td>
<td>48</td>
</tr>
<tr>
<td>4.5  Boundary Estimates from the ELISA study for BBL, BLL and 4PL models</td>
<td>49</td>
</tr>
<tr>
<td>4.6  Estimates of Selected Distributional Characteristics</td>
<td>50</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Simulated data from 4PL (1.1) and BBL models (2.1), with $\alpha_1 = -2.5$, $\alpha_2 = 2$, $\beta = 2.2$ in (2.1) and $p = \beta - \alpha_1$, $q = \alpha_2$, $\sigma = 0.25$ in (1.1). The two models have same mean response curve.</td>
<td>4</td>
</tr>
<tr>
<td>3.1 Data are generated under the BBL model. No plate effects are considered and plates are assumed independent. Model parameters are $\alpha_1 = 1$, $\alpha_2 = 6$, $\beta_1 = 1$ and $\beta_2 = -3$.</td>
<td>22</td>
</tr>
<tr>
<td>4.1 Anti-F ELISA Immunoassay Plate Layout</td>
<td>39</td>
</tr>
<tr>
<td>4.2 Responses from the Anti-F IgG ELISA study. Dash curve depicts the expected response with $g(x)' = (1, x)$ and $\phi(x) = 1$; Solid curve has $g(x)' = \phi(x)' = (1, x)$.</td>
<td>51</td>
</tr>
<tr>
<td>4.3 Expected responses and transformed expected response for each plate. Figure (a) shows the expected response for each plate under assuming that all plates have same boundaries; Figure (b) shows the expected response for each plate considering each plate have different boundaries.</td>
<td>52</td>
</tr>
<tr>
<td>4.4 95% bootstrapped prediction interval of responses. The dashed curve is the expected response function with $g(x)' = \phi(x)' = (1, x)$.</td>
<td>53</td>
</tr>
</tbody>
</table>
4.5 A series of confidence ellipsoids for $10^\text{slope}$ and $EC_{50}$ values under assumption that all plates have same boundaries. . . . . . . . . . . . 54
ABSTRACT

The beta distribution is a simple and flexible model in which responses are naturally confined to the finite interval, $(0, 1)$. The parameters of the distribution can be related to covariates such as dose and gender through a regression model. The Ballooned Beta-logistic model, with expected responses equal to the Four Parameter Logistic model, is introduced. It expands the response boundaries of the beta regression model from $(0, 1)$ to $(L, U)$, where $L$ and $U$ are unknown parameters. Under the Ballooned Beta-logistic model, expected responses follow a logistic function, but it differs from the classical Four Parameter Logistic model, which has normal additive normal errors, with positive probability of response from $-\infty$ to $\infty$. In contrast, the Ballooned Beta-logistic model naturally has skewed responses with smaller response variances at more extreme covariate values and symmetric responses with relative large variance at central values of the covariate. These features are common in bioassay data at different concentrations. The asymptotic normality of maximum likelihood estimators is obtained even though the support of this non-regular regression model depends on unknown parameters.

We find maximum likelihood estimates of boundaries converge faster to $L$ and $U$ than do extreme values at the minimum and maximum concentrations. We also find that maximum likelihood estimators perform better than least squares estimators when the covariate range is not sufficiently wide. Given multiple enzyme-linked immunosorbent assay (ELISA) data from different plates, the motivating question in a validation study was whether all plates had equivalent performance. A step-wise procedure is applied to measure equivalence of boundaries, slope and $EC_{50}$ values.
First, we establish suitability criteria for estimates of $L$ and $U$ under the Ballooned Beta-logistic model, after which plates with boundary estimates outside these limits would be considered as "reference failures". Second, we use a bivariate normal approximation to evaluate the equivalence of Hill slopes and the dose giving, half maximal responses, the $EC_{50}$ values, among plates considering $L$ and $U$ to be nuisance parameters, after accepting the boundary equivalences. A series of confidence ellipsoids, an indicator of laboratories inhomogeneity, are drawn to detect plates with outlying slopes and $EC_{50}$s. The maximum likelihood estimates of parameters are obtained using a combination of a grid search with the Newton-Raphson method. Moreover, different non-linear models compared in terms of their $EC_{10}$, $EC_{50}$, and $EC_{90}$ values and the bootstrap method is applied to draw 95% bootstrap predictive intervals for responses over all concentrations.
Chapter 1

Introduction

A dose-response study measures the change in effect at different doses, or chemical concentrations, after a certain exposure time. Motivation for dose-response studies focuses on determining safe, hazardous, and effective dose levels for drugs, pollutants and other substances. To model dose-response relationships that are naturally sigmoid shaped with continuous responses, specifically, to explain the binding of oxygen to hemoglobin, Hill et al. (1910) introduced the $E_{\text{max}}$ model, which is also known as the four parameter logistic model (4PL). The 4PL model is widely used in bioassay, immunoassay, genetic, nutrition and agriculture studies:

$$Y = \eta(x) + \epsilon, \text{ with } \epsilon \sim N(0, \sigma^2), \quad (1.1)$$

where the mean function is
\[ \eta(x) = E[y|x] = B + (A - B) \frac{1}{1 + \exp(R + Sx)} = A + (B - A) \frac{1}{1 + \exp(-(R + Sx))}, \]

(1.2)

with parameters \( A, B, R, S \) and covariate \( x = \log(u) \), where \( u \) is the concentration.

Note \( \eta(x) \to A \) as \( x \to -\infty \) when \( S > 0 \) or as \( x \to \infty \) when \( S < 0 \); \( \eta(x) \to B \) as \( x \to \infty \) when \( S > 0 \) or as \( x \to -\infty \) when \( S < 0 \); \( \eta(x) = B + (A - B)/(1 + \exp(R)) = A + (B - A)/(1 + \exp(R)) \) when \( s = 0 \). The rightmost second term in (1.2) can be written as

\[ \frac{B - A}{1 + e^{-(R+Sx)}} = \frac{B - A}{1 + e^{-S(R/S+x)}} = \frac{B - A}{1 + (u/e^{-R/S} - S)}, \]

(1.3)

where \( S \) is the so called Hill slope and \( e^{-R/S} \) is the \( EC_{50} \) Holford and Sheiner (1981).

Michaelis and Menten (1913) studied a simplified version of model (1.1,1.2) with \( A = 0 \) and \( S = 1 \). Wagner (1968) first used the \( E_{\text{max}} \) model to explain the relationship between drug concentrations and responses. Applications of the \( E_{\text{max}} \) model are discussed by DeLean et al. (1978), Volund (1978), Holford and Sheiner (1981), Ratkowsky and Reedy (1986), Finke et al. (1989), Gahl et al. (1991), Ernst et al. (1997), Triantafilis et al. (2000), Menon and Bhandarkar (2004), Macdougall (2006), Dragalin et al. (2007), Vedenov and Pesti (2008), Sebaugh (2011) and many others.

Two shortcomings exist with the 4PL model. First, the parameters \( A \) and \( B \) are the minimum and maximum, respectively, of \( E(Y|X) \) and not bounds on the response \( Y \). Second, the response variances are constant. However, in many dose-response studies, response are likely to have smaller variance at the extreme doses than at central ones. See, for example, the allocations given in Chapter 4 and Leonov and Miller (2009). Leonov and Miller (2009) relaxed the constant variance assumption
by letting it depend on a covariate; but they left the range of possible responses unbounded. Figure 1.1 compares simulated data under a BBL model (described Chapter 2) and the 4PL model which has the same expected response function. One feature of the BBL model is that the distribution can be symmetric with relative large variant at central values of the covariate and skewed with smaller variance at more extreme values. Alternatively, the variances can be monotone increasing or decreasing depending on parameter values. These features are common in bioassay data.

To address the two disadvantages of the 4PL model with additive normal errors, Wang et al. (2013) developed a new bounded log-linear (BLL) regression model. They set a transformed response equal to a linear predictive function with an additive error $\epsilon$:

$$Y = U + (L - U) \frac{1}{1 + e^{C + D_x + \epsilon}}, \quad \epsilon \sim N(0, \sigma^2),$$

(1.4)

where $U$ and $L$ are two unknown bounds on the response random variable $Y$. The difference between (1.1,1.2) and (1.4) is that the classical 4PL model has error additive to the mean function, while the BLL model has error added after the predictor function is linearized. Even though model 1.4 has a constant error term for a transformed response, untransformed responses at central concentrations are more scattered than those at more extreme concentrations.

Ferrari and Cribari-Neto (2004) modeled rates and proportions using a beta regression function. Tamhane et al. (2002) described regression for ordinal data using a beta model for quality improvement. A beta regression model with logistic mean function was proposed by Wu et al. (2005), but they left the response variable in the beta distribution bounded between 0 and 1. So bounds not hold in many situations, and this motivated us to develop a new model that retains the good properties of the
Figure 1.1: Simulated data from 4PL (1.1) and BBL models (2.1), with $\alpha_1 = -2.5$, $\alpha_2 = 2$, $\beta = 2.2$ in (2.1) and $p = \beta - \alpha_1$, $q = \alpha_2$, $\sigma = 0.25$ in (1.1). The two models have same mean response curve.
beta regression model but has two unknown boundaries.

In our model, the support of the random variable $Y$ depends on unknown boundaries $L$ and $U$. Therefore, asymptotic normality of the maximum likelihood estimates (MLEs) does not follow from standard arguments. Smith (1985) derived the properties of MLEs for a board class of non-regular regression models which include a single unknown boundary parameter. His proof is based on a key requirement that $c_n(y - L)$ converges to a non-degenerate distribution as $y \to L$ where $c_n$ is some sequence of constants and $L$ is the lower bound of response $Y$. Because the BBL model has two unknown boundaries, we take a different approach to characterizing the MLEs. Harter and Moore (1966) proposed using solutions to the maximum likelihood equations in place of maximizing the likelihood function which might provide an infinite estimate. These are called local MLEs. Wang et al. (2013) provided an alternative to Smith’s proof of the existence of a consistent local MLE. In this paper, we follow the work of Smith (1985), Smith (1994) and Wang et al. (2013) in showing that the solutions to the likelihood equations provide good estimates of the unknown parameters.

Sebaugh (2011) investigated the importance of the covariate range in estimation quality. A comparison of different parameter estimates is provided in section 3. When the expected response function has a clear pattern of sigmoid shape, the MLEs of boundaries under the BBL model have slightly smaller bias and standard deviation than least square estimates (LSEs). However, when the expected response function doesn’t display a sigmoid shape over the covariate range used, the MLEs of the boundaries under the BBL have much smaller bias and standard deviation than LSEs. The LSEs under the BBL model is equivalent to the MLEs and LSEs for the 4PL
model. We also evaluated the performance of the extreme order statistics.

The rest of this paper is organized as follows. In Chapter 2, the new Ballooned Beta-logistic (BBL) model with two unknown bounds is introduced and the asymptotic distributions of its minimum and maximum order statistics are given. In Chapter 3, we characterize the solution to the maximum likelihood equations in the BBL model and compare MLEs, LSEs and extreme order statistics between two models with different covariate ranges. In Chapter 4, we analyze a real enzyme-linked immunosorbent assay (ELISA) dataset and compare the performance of the new BBL model with that of the 4PL and BLL models.
Chapter 2

The Ballooned Beta-Logistic Model

The probability density function of a standard beta distribution is

\[ f_W(w) = B(a, b) w^{a-1} (1 - w)^{b-1} \]

for \( 0 \leq w \leq 1, \ a \geq 0, \ \text{and} \ b \geq 0; \) where \( B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} \, dt = \Gamma(a)\Gamma(b)/\Gamma(a+b) \) is the beta function and \( \Gamma(p) = \int_0^\infty e^{-t} t^{p-1} \, dt \) is the gamma function. The mean and variance, respectively, of the beta distribution are

\[ \text{E}[W] = a/(a+b) \] and \[ \text{Var}[W] = ab/[(a+b)^2(a+b+1)]^{-1}. \]

A beta regression model with logistic mean function, which is bounded between (0,1), was introduced by Wu et al. (2005). The parameters \( a \) and \( b \) in the beta density are set to functions of covariates as \( \ln(a) = \alpha'g(x) \) and \( \ln(b) = \beta'\phi(x) \) so \( a \) and \( b \) are positive regardless of the value of the regression coefficients; \( \alpha \) and \( \beta \) are vectors:
\( \alpha' = (\alpha_1, \ldots, \alpha_{ma}) \) and \( \beta' = (\beta_1, \ldots, \beta_{mb}) \); and the functions \( g(x) \) and \( \phi(x) \) are vector valued functions of the covariate \( x \). For example, it may be that \( g(x)' = (1, x) \) and \( \phi(x) = 1 \) with \( ma = 2 \) and \( mb = 1 \).

Note one can write the mean function as

\[
E[W|x] = \frac{1}{1 + \exp(\beta'\phi(x) - \alpha'g(x))},
\]

To generalize this model we introduce a new random variable \( Y \) having two arbitrary unknown real valued boundaries, \( L \) and \( U \) with \( L < U \), through the transformation \( Y = L + (U - L)W \). Now, \( E(Y|x) = L + (U - L)E(W|x) \). We also allow the possibility of plate effects so that one may investigate the homogeneity of data from different laboratories. Let \( Y_{ij} \) be the response for the \( i \)th concentration on the \( j \)th plate, \( i = 1, \ldots, I \), and \( j = 1, \ldots, J \). Then the general form of the BBL model is

\[
f(y_{ij}) = \frac{\Gamma(a_{ij} + b_{ij})}{\Gamma(a_{ij})\Gamma(b_{ij})} \frac{1}{U_j - L_j} \left( \frac{y_{ij} - L_j}{U_j - L_j} \right)^{a_{ij}-1} \left( \frac{U_j - y_{ij}}{U_j - L_j} \right)^{b_{ij}-1},
\]

where \( a_{ij} = \exp(\alpha_j'g(x_i)) \) and \( b_{ij} = \exp(\beta_j'\phi(x_i)) \); \( g(x) \) and \( \phi(x) \) are vector valued functions of the concentration, \( u = \exp(x) \). For simplicity, we use \( \alpha \) and \( \beta \) to denote arbitrary parameters \( \alpha_j \) and \( \beta_j \), respectively.

Wu et al. (2005) considered the special case of a single covariate effect on \( a_{ij} \), namely, \( g(x)' = (1, x) \) and \( \phi(x) = 1 \). As will be shown in Chapter 4, this model did not fit our motivating dataset well and so we consider covariate effects also on \( b_{ij} \). Specifically, we focus on a model in which \( g(x)' = (1, x) \) and \( \phi(x)' = (1, x) \). The
resulting expected response function of model (2.1) is

$$\eta(x) = E_Y[Y|x] = L + (U - L) \frac{1}{1 + \exp((\beta_1 + \beta_2 x) - (\alpha_1 + \alpha_2 x))}$$

$$= L + (U - L) \frac{1}{1 + \left[ u/\exp\left(-\frac{\beta_1 - \alpha_1}{\beta_2 - \alpha_2}\right) \right]^{\beta_2 - \alpha_2}}, \quad (2.2)$$

which has the same logistic shape as the mean function of the 4PL model in (1.1,1.2). Note $\eta(x) \to L$ or $U$ as $x \to \pm \infty$. Matching terms in equation (1.3) and (2.2), the Hill slope and the $EC_{50}$ for BBL model are seen, respectively, to be

$$S = \alpha_2 - \beta_2 \text{ and } EC_{50} = \exp\left(-\frac{\beta_1 - \alpha_1}{\beta_2 - \alpha_2}\right).$$

and these equations imply also that $L = A$ and $U = B$. 
Chapter 3

Parameter Estimation

This section characterizes extreme order statistics as estimates for boundaries, least square estimates (LSEs) and maximum likelihood estimates (MLEs) of the BBL (2.1) and 4PL (1.1,1.2) models. Without loss of generality, the BBL model discussed in this section has a vector of six parameters ($\alpha_1, \alpha_2, \beta_1, \beta_2, L, U$), but only four unique normal equations; the 4PL model has parameter vector $\theta = (S, EC_{50}, L, U)$. However, the LSEs of $S$ and $EC_{50}$ in the BBL model, which are the functions of $\alpha_1, \alpha_2, \beta_1$ and $\beta_2$, are estimable, and they are equivalent to the LSEs of 4PL model or any other model with the same mean function (see Section 3.2).

Introduction and related inference of extreme values, LSEs and MLEs are shown below. Also, details of our approach to finding MLEs for BBL model are described in Section 3.6. A simulation study comparing extreme values, MLEs and LSEs under the BBL and 4PL models is described in Section 3.7.
3.1 Estimate Response Boundaries Using the Extreme Order Statistics

Suppose an independent sample \{Y_1, Y_2, \ldots, Y_n\} is obtained a single plate under model (2.1). If parameters \(L\) and \(U\) were estimated by a previous experiment and can be considered known, a transformation of \(Y\) will have a beta distribution and parameters in \(a\) and \(b\) can be estimated using the Newton-Raphson method. When \(L\) and \(U\) are unknown, one might consider estimating them using extreme order statistics: \(Y_{(1)} = \min(Y_1, \ldots, Y_n)\) and \(Y_{(n)} = \max(Y_1, \ldots, Y_n)\). These sample extreme values don’t perform very well as estimates of \(L\) and \(U\) because, although they are consistent, they have a slow convergence rate. This can be seen in Theorem 3.1.1. Define

\[
\gamma_1 = \left( \frac{\Gamma(a)\Gamma(b)\, b}{\Gamma(a + b)\, n} \right)^{1/b} (U - L) \quad \text{and} \quad \gamma_1 = \left( \frac{\Gamma(a)\Gamma(b)\, a}{\Gamma(a + b)\, n} \right)^{1/a} (U - L).
\]

**Theorem 3.1.1.** The limiting distributions of \(Y_{(1)}\) and \(Y_{(n)}\), respectively, are given by

\[
\gamma_1^{-1}(Y_{(n)} - U) \xrightarrow{L} \text{exp}\{-(y)^b\} \quad \text{as} \quad n \to \infty;
\]

\[
\gamma_2^{-1}(L - Y_{(1)}) \xrightarrow{L} \text{exp}\{-y^a\} \quad \text{as} \quad n \to \infty.
\]

These results are consistent with those found for extreme order statistics under the BLL model Wang et al. (2013).

*Proof of Theorem 3.1.1*

Define probability density function and cumulative distribution function of random variable \(y\) as \(f(y)\) and \(F(y)\), respectively. Also define \(y_\infty = \sup\{y : F(y) < 1\}\).
Consider an arbitrary $a = a_{ij}$ and $b = b_{ij}$. Then in the Ballooned Beta-logistic model, $y_\infty = U$, the upper bound of $Y$. When $y \to U$, Ferguson (1996)

$$\lim_{y \to U} \frac{f(y)}{\zeta_1(U - y)^{b-1}} \to 1, \text{ where } \zeta_1 = \frac{\Gamma(a + b)}{\Gamma(a) \Gamma(b)} \left( \frac{1}{U - L} \right)^b,$$

and

$$1 - F(y) = \zeta_1 \int_y^U (U - t)^{b-1} dt = \frac{1}{b} (U - y)^b.$$

Hence when $y \to U$, $f(y)$ and $\zeta_1(U - y)^{b-1}$ are asymptotically equivalent. Condition (b) of Theorem 14 in Ferguson (1996) holds, and so the result

$$1 - F(1 - \gamma_1) = \frac{1}{n}$$

yields $\gamma_1^b = b/(\zeta_1 n) = \Gamma(a) \Gamma(b)/\Gamma(a + b)(U - L)^b$; the explicit expression of $\gamma_1$ is

$$\gamma_1 = \left( \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b) n} \right)^{1/b} (U - L).$$

Hence,

$$\gamma_1^{-1}(Y(n) - U) \xrightarrow{n \to \infty} G_{2,b} = \exp \left\{ \frac{-y}{b} \right\}.$$

To get the extreme value distribution of the minima, let $T = -y$ and substitute $y$ in the distribution function. The density of $T$ is

$$f_T(t) = \frac{\Gamma(a + b)}{\Gamma(a) \Gamma(b)} \frac{1}{U - L} \left( \frac{-t - L}{U - L} \right)^{a-1} \left( \frac{U + t}{U - L} \right),$$

where $t \in [-U, -L]$. $Y(1)$ can be expressed by $T$ through $Y(1) = -\max(T_1, \ldots, T_n)$. Define $t_\infty = \sup\{ t : F(t) < 1 \}$; then $t_\infty = -L$. When $t \to -L, f(t)$ is asymptotically
equivalent with $\zeta_2(-t - L)^{a-1}$, where $\zeta_2 = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{1}{U-L}\right)^a$. Thus,

$$1 - F(t) = \zeta_2 \int_t^{-L} (-p - L)^{a-1} dp = \zeta_2 \frac{1}{a} (-t - L)^a.$$ 

Condition (b) of Theorem 14 in Ferguson (1996) still holds, with $\gamma = a$ and $t_0 = -L$, and the equation

$$1 - F(\gamma_2) = \frac{1}{n}$$

yields $\gamma_2^a = a(\zeta_2 n)^{-1}$. The explicit expression is $\zeta_2 = \left(\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \frac{a}{n}\right)^{1/a} (U - L)$.

Hence, we have

$$\gamma_2^{-1}(t(n) - (-L)) \xrightarrow{L} G_{2,a} = exp\{-(-t)^a\}$$

$$\gamma_2^{-1}(L - Y_{(1)}) \xrightarrow{L} exp(-y^a).$$

### 3.2 Least Square Estimates under BBL model

The method of least squares is always applied to find estimates for linear or nonlinear regression models. The main goal is to minimize the sum of squared residuals, which are the difference between observed value and the fitted value. The LS for BBL model is shown in (3.2).

$$LS_{BBL} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \left( y_{ij} - L - \frac{U - L}{1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i)} \right)^2$$

(3.2)

where concentrate $i = 1, \ldots, I$ and replicates $j = 1, \ldots, J$. 
The first derivatives of (3.2) with respect to different parameters are

\[ \frac{\partial LS_{BLL}}{\partial U} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \ast 2 (y_{ij} - \eta(x_i, \theta)) \]

\[ \left( -\frac{1}{1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i)} \right) \]

\[ \frac{\partial LS_{BLL}}{\partial L} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \ast 2 (y_{ij} - \eta(x_i, \theta)) \]

\[ \left( -1 + \frac{1}{1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i)} \right) \]

\[ \frac{\partial LS_{BLL}}{\partial \alpha_1} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \ast 2 (y_{ij} - \eta(x_i, \theta)) \]

\[ \left( -(U - L) \frac{exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i)}{(1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i))^2} \right) \]

\[ \frac{\partial LS_{BLL}}{\partial \beta_1} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \ast 2 (y_{ij} - \eta(x_i, \theta)) \]

\[ \left( (U - L) \frac{exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i)}{(1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i))^2} \right) \]

\[ \frac{\partial LS_{BLL}}{\partial \alpha_2} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \ast 2 (y_{ij} - \eta(x_i, \theta)) \]

\[ \left( -(U - L) \frac{exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i) x_i}{(1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i))^2} \right) \]

\[ \frac{\partial LS_{BLL}}{\partial \beta_2} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i \ast 2 (y_{ij} - \eta(x_i, \theta)) \]

\[ \left( (U - L) \frac{exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i) x_i}{(1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i))^2} \right) \]

where

\[ \eta(x, \theta) = L - \frac{U - L}{1 + exp(\beta_1 + \beta_2 x_i - \alpha_1 - \alpha_2 x_i)} \]

Since \( \alpha_1, \beta_1 \) and \( \alpha_2, \beta_2 \) are the intercept and first order coefficient of the covariate
effect in parameter $a$ and $b$, respectively. The first order derivatives of $LS_{BBL}$ with respect to $\alpha_1$, $\beta_1$ are same; and the first order derivatives of $LS_{BBL}$ with respect to $\alpha_2$, $\beta_2$ are same. Hence, $\alpha_1$, $\alpha_2$, $\beta_1$ and $\beta_2$ are not identifiable. However, when parameters slope and $EC_{50}$ are considered in BBL model, where slope and $EC_{50}$ are functions of $\alpha$’s and $\beta$’s, those two parameters are identifiable. In this case, the LSE of slope, $EC_{50}$, $L$ and $U$ under the BBL model are same as the LSE of slope, $EC_{50}$, $A$ and $B$ under the 4PL model, since BBL and 4PL models have exact the same expected response functions.

One note worth mentioning is that the LSEs and MLEs are equivalent under the 4PL model. The least square for the 4PL model can be expressed as

$$LS_{4PL} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left( y_{ij} - A - \frac{B - A}{1 + (u/EC_{50})^{-S}} \right)^2$$  \hspace{1cm} (3.3)

The density of 4PL random variable is

$$f(y_{ij}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{(y_{ij} - A - \frac{B - A}{1 + (u/EC_{50})^{-S}})^2}{2\sigma^2}\right)$$

The corresponding likelihood function for all observations is

$$L(\theta', x) = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{(y_{ij} - A - \frac{B - A}{1 + (u/EC_{50})^{-S}})^2}{2\sigma^2}\right),$$  \hspace{1cm} (3.4)

where $\theta' = (S, EC_{50}, A, B)$. Taking logarithm of (3.4),

$$\ell(\theta', x) = IJ \times log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left( y_{ij} - A - \frac{B - A}{1 + (u/EC_{50})^{-S}} \right)^2$$  \hspace{1cm} (3.5)
Since the first derivatives of (3.3) and (3.5) with respect to $S, EC_{50}, A$ and $B$ are same, the least square estimates and maximum likelihood estimates under the 4PL model are same. Equivalent estimates can simplify the estimates comparison in section 3.7.

### 3.3 Maximum Likelihood Estimates under BBL Model

Assuming independence conditional on concentration, the likelihood for plate $j$ is

$$
L(\theta_j, x, y) = \prod_{i=1}^{I} \prod_{k=1}^{K} f(y_{ijk}|\theta_j, x_i),
$$

where $\theta_j$ includes all model parameters for plate $j$, $\theta'_j = (\alpha_j, \beta_j, L_j, U_j)$ with $\alpha'_j = (\alpha_{1j}, \ldots, \alpha_{maj})$, $\beta'_j = (\beta_{1j}, \ldots, \beta_{mbj})$, and all its possible values belong to a compact set. Assuming responses for each plate are independent conditional on concentration, the likelihood for all plates is as

$$
L(\theta, x, y) = \prod_{j=1}^{J} L(\theta_j, x, y), \text{ where } \theta' = (\theta_1, \ldots, \theta_J). \quad (3.6)
$$

Identities of other relationships between the $\theta_j$ may be specified. The maximum likelihood estimators are $\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta, x, y)$.

There are three assumptions required to support asymptotic properties of MLEs. Related theorems and proofs are given for an arbitrary plate, so $J = 1$.

**Assumption 1:** $\sup ||x|| < \infty$, where $|| \cdot ||$ is the Euclidean norm.

**Assumption 2:** The following terms converge as $n \to \infty$: $n^{-1}a, n^{-1}b, n^{-1}ag(x), n^{-1}a\phi(x), n^{-1}bg(x), n^{-1}b\phi(x), n^{-1}abg(x)\phi(x), a(a + b - 1)/bn, (a + b - 1)/(an)$.

**Assumption 3:** The vectors $g(x)$ and $\phi(x)$ are full rank.
Theorem 3.3.1. (Existence) If assumptions 1-3 hold, then with probability approaching 1, there exists a sequence of solutions $\hat{\theta}_n$ to the likelihood equations of (3.6) that is $n^{1/2}$-consistent for $\theta$.

The proof of this theorem is given by Wang et al. (2013).

Theorem 3.3.2. (Uniqueness) Let assumptions 1-3 hold and let $\delta$ be some fixed value and $\delta_n = n^{-\alpha}$ for some $\alpha > 0$. Denote by $S_\delta = \{ \theta : L \leq L_0 - \delta$ and $U \geq U_0 + \delta \}$ and $T_{\delta,n} = \{ \theta : L_0 - \delta \leq L \leq L_0 + \delta_n, U_0 - \delta_n \leq U \leq U_0 + \delta$ and $\|a-a_0\| + \|b-b_0\| > \delta \}$. Then for any compact set $K \in \mathbb{R}^{p+2}$,

$$\lim_{n \to \infty} \Pr \{ \sup_{S_\delta \cap K} l_n(\theta) < l_n(\theta_0) \} = 1;$$

$$\lim_{n \to \infty} \Pr \{ \sup_{T_{\delta,n} \cap K} l_n(\theta) < l_n(\theta_0) \} = 1.$$

Theorem 3.3.3. (Asymptotic Normality) If assumptions 1-3 hold, then the asymptotic distribution of $\hat{\theta}$ satisfies

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N \left\{ 0, M^{-1}(\theta_0) \right\}, \quad (3.7)$$

where $\theta_0$ is a vector containing true values of parameters and $M(\theta)$ is the Fisher information matrix. Estimate of $M^{-1}(\theta_0)$ can be computed from $M^{-1}(\hat{\theta})$ where $\hat{\theta}$ is the MLEs of $\theta_0$ in section 3.6.

The proofs of Theorems 3.3.2 and 3.3.3 are provided in Section 3.8.
3.4 Fisher Information

The information matrix for a single response on plate \( j \) at dose \( x \) with respect to \( \theta' = (\alpha, \beta, L, U) \) can be expressed as

\[
\mu(\theta_j, x) = -\mathbb{E}\left[ \frac{\partial^2}{\partial \theta \partial \theta^T} \ln \mathcal{L}(y_i | \theta_j, x) \right]
\]

\[
= \begin{pmatrix}
\mu_{11j} & \mu_{21j} & \frac{a_j b_j}{(a_j - 1)(U_j - L_j)} g(x) & \frac{a_j}{U_j - L_j} g(x) \\
\mu_{21j} & \mu_{22j} & -\frac{b_j}{U_j - L_j} \phi(x) & -\frac{a_j b_j}{(b_j - 1)(U_j - L_j)} \phi(x) \\
\frac{a_j b_j}{(a_j - 1)(U_j - L_j)} g^T(x) & -\frac{b_j}{U_j - L_j} \phi^T(x) & \frac{b_j}{a_j - 2} \frac{a_j + b_j - 1}{(U_j - L_j)^2} g(x) & \frac{a_j + b_j - 1}{(U_j - L_j)^2} g(x) \\
\frac{a_j}{U_j - L_j} g^T(x) & -\frac{a_j b_j}{(b_j - 1)(U_j - L_j)} \phi^T(x) & \frac{a_j + b_j - 1}{(U_j - L_j)^2} \phi(x) & \frac{a_j}{b_j - 2} \frac{a_j + b_j - 1}{(U_j - L_j)^2} \phi(x)
\end{pmatrix}
\]

(3.8)

where we note the left upper \((ma + mb) \times (ma + mb)\) submatrix of \( \mu(\theta, x) \) by \( \mu_{\text{beta}}(\theta_j) \) because it is the information matrix for a single response from the standard beta regression model Wu et al. (2005):

\[
\mu_{\text{beta}}(\theta_j) = \begin{pmatrix}
\{\psi'(a_j) - \psi'(a_j + b_j)\} a_j^2 g(x) g^T(x) & -\psi'(a_j + b_j) a_j b_j g(x) \phi^T(x) \\
-\psi'(a_j + b_j) a_j b_j \phi(x) g^T(x) & \{\psi'(b_j) - \psi'(a_j + b_j)\} b_j^2 \phi(x) \phi^T(x)
\end{pmatrix},
\]

where \( \psi' \) is digamma function.

Assuming plates are independent, under the full model with \( I \) concentration levels at \( x_1, \ldots, x_I \) and \( J \) plate effects, the total information \( M(\theta, x) \) can be reached by

\[
M(\theta, x) = K \sum_{i=1}^{I} \sum_{j=1}^{J} \mu(\theta_j, x_i),
\]

(3.9)

where \( K \) is the number of replicates at each concentration.
As an example, for a BBL model with \( \alpha_j' = (\alpha_1^j, \alpha_2^j) \), \( \beta_j' = (\beta_1^j, \beta_2^j) \) and \( g(x) = \phi(x) = (1, x)' \), reverse dimensions of \( \mu(\theta_j, x) \), \( \mu_{\text{beta}}(\theta_j) \), and \( M(\theta) \) are \( 4 \times 4 \), \( 6 \times 6 \) and \( 6 \times 6 \), respectively.

### 3.5 Maximum Likelihood Estimates of Slope and \( EC_{50} \)

The properties of MLEs of slope and \( EC_{50} \) are discussed as follow,

**Corollary 3.5.1.** For a special case of BBL model with \( \theta' = (\alpha, \beta, L, U) \) where \( \alpha' = (\alpha_1, \alpha_2) \) and \( \beta' = (\beta_1, \beta_2) \). Given the asymptotic normality of \( \hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{L}, \hat{U}) \), the joint distribution of \( (\hat{S}, \hat{I}, \hat{L}, \hat{U}) \) can be obtained by Cramer’s theorem, also known as Delta method:

\[
\sqrt{n}(g(\hat{\theta}) - g(\theta)) \rightarrow N(0, \hat{g}(\theta) \Sigma \hat{g}(\theta)')
\] (3.10)

with

\[
g(\theta)' = (S, EC_{50}, L, U)' = \left( \alpha_2 - \beta_2, \exp\left(-\frac{\beta_1 - \alpha_1}{\beta_2 - \alpha_2}\right), L, U \right)
\]

and

\[
\hat{g}(\theta) = \begin{pmatrix}
0 & 1 & 0 & -1 & 0 & 0 \\
EC_{50}/S & -EC_{50}/S & EC_{50}/S & EC_{50}/S & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix},
\]

where \( EC_{50} = \exp(- (\beta_1 - \alpha_1)/(\beta_2 - \alpha_2)) \).
The marginal distribution of \( \hat{S} \) and \( \hat{EC}_{50} \) is

\[
\begin{pmatrix}
\hat{S} \\
\hat{EC}_{50}
\end{pmatrix} \sim N
\begin{pmatrix}
S \\
EC_{50}
\end{pmatrix}, \Sigma^{*}_{11}
\]

(3.11)

where the covariate matrix of \( \Sigma^{*}_{11} \) is the upper left 2x2 submatrix of \( \dot{g}(\theta)\Sigma\ddot{g}(\theta)' \) in (3.10).

When there are no plates effects, above theorems also hold for responses from all plates combined.

### 3.6 Finding Maximum Likelihood Estimates

As for parameter estimation, the Newton-Raphson method is widely used for non-linear models. However, using this method with a large number of plates requires a high dimensional Hessian matrix. In this paper, we combined a grid with the Newton-Raphson method to estimate parameters. For example, assuming there is no plate effect and plates are independent, estimates of parameters can be found as follows: A grid of possible pairs \((L, U)\) is formed as described in Appendix B. For each pair \((L, U)\) the Newton-Raphson method is applied to find estimates of the remaining parameters. The MLEs selected is the vector of estimates yielding the maximum of the likelihood function. For details, see Appendix B.
3.7 Comparison of Estimators

In this section, different kinds of parameter estimates are compared under the BBL and 4PL models when the expected response functions of the BBL and 4PL models are the same. First, we compared the extreme values, MLEs and LSEs for upper and lower bounds under the BBL and 4PL models. Second, we compared the MLEs and LSEs of slope and $EC_{50}$ under the BBL and 4PL models.

In Section 3.2, we proved that the LSEs for the BBL and 4PL models are equivalent. And also, the LSEs and MLEs for the 4PL model are same. Hence, comparison between BBL and 4PL models can reduce to the comparison of MLEs and LSEs under the BBL model.

Data were randomly generated under the BBL parameters with $\alpha_1 = 4$, $\alpha_2 = 6$, $\beta_1 = 1$ and $\beta_2 = -3$. Two scenarios are considered. In first scenario, seven different levels of covariate $x = \log_2(u)$ are evenly allocated between -0.5 and 0.5. In the second scenario, the upper limit of covariate is reduced to 0.1. Figure 3.1 shows responses from the two scenarios. At extreme covariate values in Figure 3.1(a), variation among responses is much smaller than in the middle range of covariate. In Figure 3.1(b), upper limit of covariate is truncated. We investigate the performance of estimates when covariate responses are not clustered up against $U$.

For each scenario, a set of MLEs and LSEs for the BBL model was obtained simultaneously. Different numbers of plates, namely, 30, 50, 100, are considered to have 1 replication at each concentration. The bias and variance of the estimates are computed based on 500 simulations. Table 3.1 compares boundary estimates in both scenarios. In scenario 1, MLEs and extreme values have slightly smaller bias and standard deviation than LSEs when number of plates is 30; the difference among
Figure 3.1: Data are generated under the BBL model. No plate effects are considered and plates are assumed independent. Model parameters are $\alpha_1 = 1$, $\alpha_2 = 6$, $\beta_1 = 1$ and $\beta_2 = -3$. 
three estimators decrease when the number of plates is gets large. In scenario 2, in which the sigmoid dose-response pattern isn’t clear, the LSEs of $U$ has larger bias and standard deviation than MLE’s: 15.081 and 31.430 as compared to -0.178 and 0.098, respectively. As the number of plates increases, the difference between LSEs and MLEs of $U$ decreases, but LSEs still have larger bias and standard deviation. Since the lower limit of the covariate range is -0.5 in both scenarios, estimates of $L$ are similar for both scenarios.

Table 3.2 compares estimates for slope and $EC_{50}$ in these two scenarios. In scenario 1, MLEs of slope have smaller bias and standard deviation than LSEs when number of plate is 30, 50; but similar performance when number of replicate is 100. As for the estimate of $EC_{50}$, MLEs and LSEs have similar performance. In scenario 2, MLEs of slope and $EC_{50}$ have much smaller bias and standard deviation than LSEs.

Simulation results show that when data demonstrate an apparent sigmoid shape with heterogeneous variance, MLEs and LSEs of boundaries perform similarly in terms of bias and standard deviation when the number of plates is large. But when responses don’t reach $U$, the LSEs of aren’t good. Analogous results pertain to the lower limit.

### 3.8 Technical Details

#### 3.8.1 The Hessian Matrix of a Ballooned Beta-logistic Distributed Random Variable

Adapting useful notations from Ferguson (1996), define a vector $x' = (x_1, x_2, ..., x_d)$, where $d$ is the dimension. Let $t(x)$ be a function of $x$. Then if $t : R^d \rightarrow R$, the first
Table 3.1: Performance of $\hat{L}$ and $\hat{U}$ under the BBL model

$\alpha_1 = 1, \alpha_2 = 6, \beta_1 = 1, \beta_2 = -3, L = 0$ and $U = 5,$
with $S = 9, EC_{50} = 1$ and covariate $\in [-0.5, 0.5].$

<table>
<thead>
<tr>
<th># of Plate</th>
<th>Estimate</th>
<th>Bias of $\hat{L}$</th>
<th>SD of $\hat{L}$</th>
<th>Bias of $\hat{U}$</th>
<th>SD of $\hat{U}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>MLEs</td>
<td>-0.001</td>
<td>0.012</td>
<td>0.001</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>LSEs</td>
<td>0.013</td>
<td>0.028</td>
<td>-0.011</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Extrs</td>
<td>0.000</td>
<td>0.012</td>
<td>0.001</td>
<td>0.012</td>
</tr>
<tr>
<td>50</td>
<td>MLEs</td>
<td>-0.001</td>
<td>0.004</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>LSEs</td>
<td>0.012</td>
<td>0.026</td>
<td>0.009</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>Extrs</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>100</td>
<td>LSEs</td>
<td>0.010</td>
<td>0.017</td>
<td>-0.004</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Extrs</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.001</td>
</tr>
</tbody>
</table>

$\alpha_1 = 1, \alpha_2 = 6, \beta_1 = 1, \beta_2 = -3, L = 0$ and $U = 5,$
with $S = 9, EC_{50} = 1$ and covariate $\in [-0.5, 0.1].$

<table>
<thead>
<tr>
<th># of Plate</th>
<th>Estimate</th>
<th>Bias of $\hat{L}$</th>
<th>SD of $\hat{L}$</th>
<th>Bias of $\hat{U}$</th>
<th>SD of $\hat{U}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>MLEs</td>
<td>-0.002</td>
<td>0.012</td>
<td>-0.178</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>LSEs</td>
<td>0.168</td>
<td>0.184</td>
<td>15.081</td>
<td>31.430</td>
</tr>
<tr>
<td></td>
<td>Extrs</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.178</td>
<td>0.094</td>
</tr>
<tr>
<td>50</td>
<td>LSEs</td>
<td>-0.000</td>
<td>0.006</td>
<td>-0.163</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>Extrs</td>
<td>0.044</td>
<td>0.030</td>
<td>-0.186</td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>MLEs</td>
<td>-0.000</td>
<td>0.001</td>
<td>-0.162</td>
<td>0.052</td>
</tr>
<tr>
<td>100</td>
<td>LSEs</td>
<td>0.008</td>
<td>0.030</td>
<td>-0.083</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>Extrs</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.071</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Note: Extrs indicate extreme value estimates. LSEs for the BBL model are equivalent to the LSEs and MLEs for the 4PL model.
Table 3.2: Performance of \( \hat{S} \) and \( \hat{EC}_{50} \) under the BBL model

\[ \alpha_1 = 1, \alpha_2 = 6, \beta_1 = 1, \beta_2 = -3, L = 0 \text{ and } U = 5, \]

with \( S = 9, EC_{50} = 1 \) and covariate \( \in [-0.5, 0.5] \).

<table>
<thead>
<tr>
<th># of Plate</th>
<th>Estimate</th>
<th>Bias of ( \hat{S} )</th>
<th>SD of ( \hat{S} )</th>
<th>Bias of ( \hat{EC}_{50} )</th>
<th>SD of ( \hat{EC}_{50} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>MLEs</td>
<td>-0.063</td>
<td>0.372</td>
<td>0.003</td>
<td>0.012</td>
</tr>
<tr>
<td>50</td>
<td>LSEs</td>
<td>0.174</td>
<td>1.081</td>
<td>0.003</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>MLEs</td>
<td>-0.034</td>
<td>0.137</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td>100</td>
<td>LSEs</td>
<td>0.079</td>
<td>0.837</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>MLEs</td>
<td>-0.027</td>
<td>0.102</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: LSEs for the BBL model are equivalent to the LSEs and MLEs for the 4PL model.
derivative of $t$ is a row vector is

$$
\dot{t}(x) = \frac{d}{dx} t(x) = \left( \frac{\partial}{\partial x_1} t(x), \ldots, \frac{\partial}{\partial x_d} t(x) \right),
$$

and the second derivative of $f : R^d \rightarrow R$ can be written as

$$
\ddot{t}(x) = \frac{d}{dx} \dot{t}(x)^T = \begin{pmatrix}
\frac{\partial^2}{(\partial x_1)^2} t(x) & \ldots & \frac{\partial^2}{\partial x_1 \partial x_d} t(x) \\
\vdots & \ddots & \vdots \\
\frac{\partial^2}{\partial x_d \partial x_1} t(x) & \ldots & \frac{\partial^2}{(\partial x_d)^2} t(x)
\end{pmatrix}.
$$

For a random variable $W$ from the beta distribution, the mean is $E[w] = a/(a+b)$, where $a$ and $b$ are modeled as functions of the covariates of interest $x$. Consider $\alpha' = (\alpha_1, \ldots, \alpha_m) \in R^{ma}$, $g(x)' = (x_1, x_2, \ldots, x_m) \in R^{ma}$; let $a(x) = exp\{\alpha' g(x)\}$ and $b(x) = exp\{\beta' \phi(x)\}$.

The first derivative about $a(x)$ with respect to $\alpha_j$, $j$-th element in $\alpha$ is

$$
\partial a(x)/\partial \alpha_j = \exp \left\{ \sum_{i=1}^{ma} \alpha_i x_i \right\} x_j = a(x)x_j,
$$

where $x_j$ is the $j$-th element in vector $x$. Then $\dot{a}(x)$ can be expressed as

$$
da(x)/d\alpha = (\partial a(x)/\partial \alpha_1, \ldots, \partial a(x)/\partial \alpha_m) = (a(x)x_1 , a(x)x_2 , \ldots , a(x)x_m) = a(x)f(x)^T.
$$
The Hessian matrix is

\[
\ddot{a}(x) = \frac{d}{d\alpha} \dot{a}(x)^T = \begin{pmatrix}
\frac{\partial^2}{(\partial\alpha_1)^2} a(x) & \cdots & \frac{\partial^2}{\partial\alpha_1 \partial\alpha_m} a(x) \\
\vdots & \ddots & \vdots \\
\frac{\partial^2}{\partial\alpha_m \partial\alpha_1} a(x) & \cdots & \frac{\partial^2}{(\partial\alpha_m)^2} a(x)
\end{pmatrix} = a(x)g(x)g(x)^T.
\]

First and second derivatives of \(b(x)\) can be expressed in a similar way by using \(\dot{b}(x)\) and \(\ddot{b}(x)\).

We use \(a\) and \(b\) for arbitrary \(a_j(x)\) and \(b_j(x)\) for simplicity. The log-likelihood function for one observation of \(y\) at \(x\) under the BBL model is

\[
\ell(\theta, x, y) = \log(\Gamma(a + b)) - \log(\Gamma(a)) - \log(\Gamma(b)) - (a + b - 1)\log(U - L) \\
+ (a - 1)\log(y - L) + (b - 1)\log(U - y).
\]
By direct calculation,

\[
\frac{\partial \ell(y|\theta)}{\partial a} = \log \left( \frac{y - L}{U - L} \right) - \psi(a) + \psi(a + b);
\]

\[
\frac{\partial \ell(y|\theta)}{\partial \alpha} = \frac{\partial \ell(y|\theta)}{\partial a} \frac{\partial a}{\partial \alpha} = \left\{ \log \left( \frac{y - L}{U - L} \right) - \psi(a) + \psi(a + b) \right\} a(x)f(x)^T;
\]

\[
\frac{\partial^2 \ell(y|\theta)}{\partial \alpha^2} = \frac{\partial^2 \ell(y|\theta)}{\partial a^2} \left( \frac{\partial a}{\partial \alpha} \right)^2 + \frac{\partial \ell(y|\theta)}{\partial a} \frac{\partial^2 a}{\partial \alpha^2}
\]
\[
= \left\{ \psi'(a + b) - \psi'(a) \right\} \{a f(x)f(x)^T\}^2
\]
\[
+ \left\{ \log \left( \frac{y - L}{U - L} \right) - \psi(a) + \psi(a + b) \right\} a f(x)f(x)^T;
\]

\[
\frac{\partial \ell(y|\theta)}{\partial b} = \log \left( \frac{U - y}{U - L} \right) - \psi(b) + \psi(a + b);
\]

\[
\frac{\partial \ell(y|\theta)}{\partial \beta} = \frac{\partial \ell(y|\theta)}{\partial b} \frac{\partial b}{\partial \beta} = \left\{ \log \left( \frac{U - y}{U - L} \right) - \psi(b) + \psi(a + b) \right\} b \phi(x)^T;
\]

\[
\frac{\partial^2 \ell(y|\theta)}{\partial \beta^2} = \frac{\partial^2 \ell(y|\theta a)}{\partial b^2} \left( \frac{\partial b}{\partial \beta} \right)^2 + \frac{\partial \ell(y|\theta)}{\partial b} \frac{\partial^2 b}{\partial \beta^2}
\]
\[
= \left\{ \psi'(a + b) - \psi'(b) \right\} \{b \phi(x) \phi(x)^T\}^2
\]
\[
+ \left\{ \log \left( \frac{U - y}{U - L} \right) - \psi(b) + \psi(a + b) \right\} b \phi(x) \phi(x)^T;
\]

\[
\frac{\partial^2 \ell(y|\theta)}{\partial \alpha \partial \beta} = \frac{\partial}{\partial \beta} \left( \frac{\partial \ell(y|\theta)}{\partial \alpha} \right) = \frac{\partial}{\partial \beta} \left( \frac{\partial \ell(y|\theta)}{\partial a} \frac{\partial a}{\partial \alpha} \right)
\]
\[
= \frac{\partial^2 \ell(y|\theta)}{\partial a \partial b} \frac{\partial a}{\partial \beta} \frac{\partial b}{\partial \alpha} + \frac{\partial \ell(y|\theta)}{\partial a} \frac{\partial^2 a}{\partial \alpha \partial \beta}
\]
\[
= \frac{\partial^2 \ell(y|\Theta)}{\partial a \partial b} \frac{\partial a}{\partial \beta} \frac{\partial b}{\partial \alpha} = \psi'(a + b)ab f(x)^T \phi(x);
\]
\[
\frac{\partial \ell(y|\theta)}{\partial L} = \frac{a + b - 1}{U - L} - \frac{a - 1}{y - L};
\]
\[
\frac{\partial \ell(y|\theta)}{\partial U} = \frac{b - 1}{U - y} - \frac{a + b - 1}{U - L};
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial L^2} = \frac{a + b - 1}{(U - L)^2} - \frac{a - 1}{(y - L)^2};
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial U^2} = \frac{a + b - 1}{(U - L)^2} - \frac{b - 1}{(U - y)^2};
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial \alpha \partial L} = \frac{\partial}{\partial L} \left( \log \left( \frac{y - L}{U - L} \right) - \psi(a) + \psi(a + b) \right) a f(x)^T
\]
\[
= \left( \frac{1}{U - L} + \frac{1}{L - y} \right) a f(x);
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial \alpha \partial U} = \frac{\partial}{\partial U} \left( \log \left( \frac{y - L}{U - L} \right) - \psi(a) + \psi(a + b) \right) a f(x)^T
\]
\[
= \frac{1}{L - U} a f(x);
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial \beta \partial L} = \frac{\partial}{\partial L} \left( \log \left( \frac{U - y}{U - L} \right) - \psi(b) + \psi(a + b) \right) b \phi(x)^T
\]
\[
= \frac{1}{U - L} b \phi(x);
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial \beta \partial U} = \frac{\partial}{\partial U} \left( \log \left( \frac{U - y}{U - L} \right) - \psi(b) + \psi(a + b) \right) b \phi(x)^T
\]
\[
= \left( \frac{1}{L - U} + \frac{1}{U - y} \right) b \phi(x);
\]
\[
\frac{\partial^2 \ell(y|\theta)}{\partial L \partial U} = \frac{1 - a - b}{(U - L)^2}.
\]

The expectation of minus each second derivative term yields the information matrix
3.8.2 Proof of Theorem 3.3.2 and 3.3.3

We include four lemmas for completeness. Our Lemmas 1 and 2 are the Lemmas 2 and 3 in Wang et al. (2013) and our Lemma 4 is Lemma 5 in Smith (1985). For simplicity, let $a_i, b_i, L_j, U_j; y$ denote an arbitrary $a_{ij}, b_{ij}, L_j$ and $U_j; y$ denotes an arbitrary $y_{ij}$.

**Lemma 1:** For constant sequences $v_n \downarrow v$ and $w_n \uparrow w$ as $n \to \infty$, 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 $n^{1+\alpha} f_n(\xi_{v_n}) \to 0$ and $n^{1+\alpha} f_n(\xi_{w_n}) \to 0$ for $\alpha > 0$ as $n \to \infty$, then

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

**Lemma 2:** For any $\alpha > 0$, let $\delta_n = n^{-\alpha}$. Then for any $k_1 \geq 0$ and $k_2 > 0$, there exists a constant $Q$ such that

$$\lim_{n \to \infty} \Pr \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{|\log(U - y_i)|^{k_1}}{(y_i - L)^{k_2}} < Q \right\} = 1,$$

uniformly in $L$ and $U$ such that $|L - L_0| < \delta_n$ and $|U - U_0| < \delta_n$.

**Lemma 3:** If Assumptions 1-3 hold, then $-n^{-1} \partial^2 \ell_n(\theta)/\partial \theta \partial \theta^T \xrightarrow{L} M(\theta_0)$ uniformly over $||\theta - \theta_0|| < \delta$. 

(3.8).
Proof: From Section 3.5.1, we have

\[
\frac{\partial^2 \ell_n(y|\theta)}{\partial \alpha^2} = \frac{\partial^2 \ell_n(y|\theta)}{\partial a^2} \left( \frac{\partial a}{\partial \alpha} \right)^2 + \frac{\partial \ell_n(y|\theta)}{\partial a} \frac{\partial^2 a}{\partial \alpha^2} = \sum_{i=1}^{n} \{\psi'(a+b) - \psi'(a)\} \{ag(x)g(x)^T\}^2 \\
+ \sum_{i=1}^{n} \log \left( \frac{y_i - L}{U - L} \right) - \psi(a) + \psi(a + b) \}ag(x)g(x)^T.
\]

It follows that

\[
\frac{1}{n} \left| \frac{\partial^2 \ell_n(\theta_n)}{\partial \alpha^2_n} - \frac{\partial^2 \ell_n(\theta_0)}{\partial \alpha_0^2} \right| \\
\leq \frac{1}{n} \sum_{i=1}^{n} \left| \{\psi'(a+b) - \psi'(a)\} \{ag(x)g(x)^T\}^2 - \{\psi'(a_0 + b) - \psi'(a_0)\} \{a_0g(x)g(x)^T\} \right| \\
+ \frac{1}{n} \sum_{i=1}^{n} \left| \log \left( \frac{y_i - L}{U - L} \right) a - \log \left( \frac{y_i - L_0}{U_0 - L_0} \right) a_0 \right| g(x)g(x)^T.
\]

The second term can be expressed as

\[
\frac{1}{n} \sum_{i=1}^{n} \left| \log \left( \frac{y_i - L}{U - L} \right) a - \log \left( \frac{y_i - L_0}{U_0 - L_0} \right) a_0 \right| g(x)g(x)^T \\
\leq \frac{1}{n} \sum_{i=1}^{n} \left| \log \left( \frac{y_i - L}{B - L} \right) (a - a_0) \right| + \frac{1}{n} \sum_{i=1}^{n} \left| \log \left( \frac{y_i - L}{U_0 - L_0} \right) - \log \left( \frac{y_i - L_0}{U_0 - L_0} \right) \right| a_0 \\
= \frac{1}{n} \sum_{i=1}^{n} \left| \log \left( \frac{y_i - L}{B - L} \right) (a - a_0) \right| + \frac{1}{n} \sum_{i=1}^{n} \left| \frac{1}{y_i - L^*}(L - L_0) \right| a_0.
\]

The right most term in (3.13) goes to 0 with small enough $\delta$ and $\delta_n$. Also, the first term in (3.13) converges to 0 in probability, which implies $n^{-1}|\partial^2 \ell_n(\theta_n)/\partial \alpha_n^2 -$
\[ \frac{\partial^2 \ell_n(\theta_n)}{\partial \alpha^2} \rightarrow 0 \] in probability uniformly. Similarly, other elements in information matrix have the same property, such as 
\[ \frac{\partial^2 \ell_n(\theta)}{\partial \beta \partial \beta^T} \rightarrow \frac{\partial^2 \ell_n(\theta_0)}{\partial \beta_0 \partial \beta_0^T}, \]
\[ \frac{\partial^2 \ell_n(\theta)}{\partial U \partial U^T} \rightarrow \frac{\partial^2 \ell_n(\theta_0)}{\partial U_0 \partial U_0^T} \] and 
\[ \frac{\partial^2 \ell_n(\theta)}{\partial L \partial L^T} \rightarrow \frac{\partial^2 \ell_n(\theta_0)}{\partial L_0 \partial L_0^T} \] in probability. □

**Lemma 4:** 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 \( u \) and \( H(u) \) is negative whenever \( \| u \| = 1 \). Then \( h \) has a local maximum at which \( H = 0 \), for some \( u \) with \( \| u \| < 1 \).

**Proof of Theorem 3.3.2:**

For any \( \theta_1 \in U \), \( E[\ell_n(\theta_1)] < \infty \), so \( E[\ell_n(\theta_1) - \ell_n(\theta_0)] < 0 \) by Jensen’s inequality. This implies there exist \( \xi_{\theta_1} \) such that

\[ \lim_{n \to \infty} P \{ \ell_n(\theta_1) - \ell_n(\theta_0) < -\xi_{\theta_1} \} = 1. \]

The BLL model has

\[ \ell(\theta) = \log(L(\theta)) = n \log(a + b) - n \log(a) - n \log(b) - (na + nb - n) \log(U - L) \]
\[ + \sum_{i=1}^{n} \log(y_i - L) + \sum_{i=1}^{n} \log(U - y_i). \]

For \( |\theta - \theta_1| < \eta < |\theta_1 - \theta_0| < \delta \),
The terms $\Delta_1, \Delta_2, \Delta_3$ can be made smaller than $\xi_{\theta_1}$ by making $\eta$ small enough.

Now to prove $\Delta_4$ is very small, let $f((L, U))^T = \log(y_i - L)(U - y_i)$. Then

$$
\dot{f} \left( \frac{L}{U} \right) = \left( \frac{\partial f \left( \frac{L}{U} \right)}{\partial L}, \frac{\partial f \left( \frac{L}{U} \right)}{\partial U} \right) = \left( \frac{-(U - y_i)}{(y_i - L)(U - y_i)}, \frac{(y_i - L)}{(y_i - L)(U - y_i)} \right).
$$

Let $L^* \in (L, L_1)$ and $U^* \in (U, U_1)$. Then by the mean value theorem

$$
f \left( \frac{L}{U} \right) - f \left( \frac{L_1}{U_1} \right) = \int_0^1 \dot{f} \left( \frac{L + \lambda \left( \frac{L_1 - L}{U_1 - U} \right)}{U + \lambda \left( \frac{L_1 - L}{U_1 - U} \right)} \right) d\lambda \left( \frac{L_1 - L}{U_1 - U} \right)
\dot{f} \left( \frac{L^*}{U^*} \right) * \left( \frac{L_1 - L}{U_1 - U} \right
= \left( \frac{-(U^* - y_i)}{(y_i - L^*)(U^* - y_i)}, \frac{(y_i - L^*)}{(y_i - L^*)(U^* - y_i)} \right) \left( \frac{L_1 - L}{U_1 - U} \right)
= \left| -\frac{L_1 - L}{y_i - L^*} + \frac{U_1 - U}{U^* - y_i} \right| \cdot
$$
From (3.14), it follows that

\[ \Delta_4 = \frac{1}{n} \sum_{i=1}^{N} \left| \frac{L_1 - L}{y_i} - \frac{U_1 - U}{U^* - y_i} \right| \]

\[ \leq \frac{1}{n} \sum_{i=1}^{N} \left( -\frac{\eta}{y_i - L_0} |L_1 - L| + \frac{\eta}{U_0 - y_i} |U_1 - U| \right) \]  

(3.15)

Now \( E[\Delta_4] \) can be made arbitrary small by choosing \( \eta \) small enough, which implies

\[ \lim_{n \to \infty} Pr \left( \Delta_4 < \frac{\xi_{\theta_1}}{5} \right) = 1. \]

Combining (3.14) and (3.15) yields

\[ \lim_{n \to \infty} Pr \left\{ \sup_{S_{\delta} \cap K} l_n(\theta) - l_n(\theta_0) < -\frac{\xi_{\theta_1}}{5} \right\} = 1 \]

for any compact set \( K \). \( S_{\delta} \cap K \) can be covered by a finite number of neighborhoods of points in \( S_{\delta} \), where \( S_{\delta} = \{ \theta : L \leq L_0 - \delta \text{ and } U \geq U_0 + \delta \} \), Hence,

\[ \lim_{n \to \infty} Pr \left\{ \sup_{S_{\delta} \cap K} l_n(\theta) - l_n(\theta_0) < -\xi_m \right\} = 1. \]

If \( U_0 \) and \( L_0 \) are known, the extended beta model can be transformed to the standard beta distribution, so it follows that

\[ \lim_{n \to \infty} Pr \left\{ \sup_{a-a_0 \parallel > \delta, \parallel b-b_0 \parallel > \delta} l_n(a, b, L_0, U_0) - l_n(\theta_0) < \xi \right\} = 1. \]
Since \((a, b, L, U) \in \theta\), for \(a_1\) and \(b_1\), \((a_1, b_1, U, L_0) \in \theta\). For \(|a - a_1| < \eta\) and \(|b - b_1| < \eta\),

\[
\frac{1}{n} |l_n(a, b, U, L) - l_n(a_1, b_1, U_0, L_0)| = \frac{1}{n} \left| n \log \frac{\Gamma(a + b)}{\Gamma(a_1 + b_1)} + n \log \frac{\Gamma(a_1) \Gamma(b_1)}{\Gamma(a) \Gamma(b)} \right|
\]

\[
+ n |(a_1 + b_1 - 1) \log(U_1 - L_1) - n(a + b - 1) \log(U - L)|
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} |\log(y_i - L) - \sum_{i=1}^{N} (y_i - L_0) + \sum_{i=1}^{N} \log(U - y_i) - \sum_{i=1}^{N} \log(U_0 - y_i)|
\]

\[
\leq \Delta_5 + \frac{1}{n} \sum_{i=1}^{N} |\log(y_i - L) - \log(y_i - L_0)| + \frac{1}{n} \sum_{i=1}^{N} |\log(U - y_i) - \log(U_0 - y_i)|
\]

\[
= \Delta_5 + \Delta_6 + \Delta_7.
\]

\(\Delta_5\) can be made smaller than \(\xi/4\) by choosing \(\eta\) small enough.

For \(\Delta_7\), we have \(\partial \log(U - y_i)/\partial U = (U - y_i)^{-1}\) by the mean-value theorem. So if \(U > U_0\),

\[
\sum_{i=1}^{N} |\log(U - y_i) - \log(U_0 - y_i)| = \sum_{i=1}^{N} \frac{|U - U_0|}{U^* - y_i}
\]

\[
\leq |U - U_0| \sum_{i=1}^{N} \frac{1}{\min(U, U_0) - y_i}
\]

\[
\leq \sum_{i=1}^{N} \frac{1}{U_0 - y_i}.
\]

If \(U_0 - \delta_n < U < U_0\), from Lemma 2, there exist some constant \(M^*\), such that

\[
\lim_{n \to \infty} \Pr \left\{ \frac{1}{n} \sum_{i=1}^{N} \frac{1}{|U - y_i|} < M^* \right\} = 1 \text{ for small } \eta.
\]
For some small enough \( \eta \), \( \lim_{n \to \infty} Pr\{\Delta_5 < \frac{\xi}{4}\} \), so we have

\[
\lim_{n \to \infty} Pr \left\{ \sup l_n(a, b, L, U) - l_n(\theta_0) < -\frac{\xi}{4} \right\} = 1.
\]

Now Theorem 2 is proved. \( \square \)

**Proof of Theorem 3.3.3**

Define \( \hat{l}_n = \sum_{i=1}^N \ell(\theta, x_i) \). By the mean value theorem,

\[
\hat{l}_n(\theta) = \hat{l}_n(\theta_0) + \int_0^1 \hat{l}_n(\theta_0 + \lambda(\theta - \theta_0)) d\lambda(\theta - \theta_0).
\]

Replace \( \theta \) with \( \hat{\theta}_n \), where \( \hat{\theta}_n \) is a solution of likelihood equations. Then

\[
\hat{l}_n(\hat{\theta}_n) = \hat{l}_n(\theta_0) + \int_0^1 \hat{l}_n(\theta_0 + \lambda(\hat{\theta}_n - \theta_0)) d\lambda(\hat{\theta}_n - \theta_0) = \hat{l}_n(\theta_0) + \hat{l}_n(\hat{\theta}_n^\ast)(\hat{\theta}_n - \theta_0) = 0,
\]

where \( \hat{\theta}_n^\ast \) is between \( \theta_0 \) and \( \hat{\theta}_n \). From Lemma 4, \(-n^{-1} \partial^2 l_n(\theta)/(\partial \theta \partial \theta^T) \to I(\theta_0)\).

Hence,

\[
\frac{1}{\sqrt{n}} \hat{l}_n(\theta_0) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \Psi(\theta_0, x_i) \right) \to N(0, I(\theta_0))
\]

and

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \to I^{-1/2}(\theta_0)Z \to N(0, I^{-1}(\theta_0))
\]

in distribution. \( \square \)
Chapter 4

Illustration from Assay Experiment

An assay is an analytic procedure in laboratory medicine, pharmacology, environmental biology for qualitatively assessing or quantitatively measuring the presence or amount or the functional activity of a target entity. Depending on the substrate on which the assay principle is applied, assay has three different types: bioassay, ligand binding assay and immunoassay. Bioassays are typically conducted to measure the potency or effects of a drug or material by utilizing the reaction caused by its application to experimental subjects that are living. Immunoassay is a biochemical test that measures the presence or concentrations of a macromolecule in a solution through the use of an antibody or immunoglobulin. Nowadays, assays are used in many science studies, such as measurement of the pharmacological activity of new substances, investigation of the function of endogenous mediators, determination of drug toxicity and so on.

For assay study, it is informative to establish a relationship between dose and the magnitude of the response produced by the dose. The relationship can be used to
study the potency of a dose from the response it produces. The estimate of potency is always relative to a standard preparation of a stimulus, which may be a convenient working standard adopted in a laboratory or laboratories. A test preparation of the stimulus, having an unknown potency, is assayed to find the mean response to a selected drug. Next we find the dose of the standard preparation which produces the same mean response. The ratio of the two equally effective doses is an estimate of the potency of the test preparation relative to that of the standard.

An ideal situation to measure dose-response relationship and estimate potency is that the test and standard preparations are identical in their biologically active ingredient and differ only in degree of dilution by inactive materials to which they are subjected. From this point of view, more precious dose-response relation and estimate of potency can be obtained if all participated laboratories have identical experimental conditions and results. Although assay is regarded as a recent development, the essence of quantal response techniques were used by many people in early years. Emmens et al. and Finney (1947) were the pioneers who first consider the statistical aspects of bioassay. Coward (1938) and Gaddum (1948) considered the biological aspects of the assay.

In immunoassay research, the method of enzyme-linked immunosorbent assay (ELISA) is used to identify substances through color changes that are caused by antibodies effects. The ELISAs are typically performed in 96-well polystyrene plates, which will passively bind antibodies and proteins. Color changes are related to the binding strength between antibodies and proteins; see Figure 4.1. To assure the assay quality, the Food and Drug Administration 2010 report (FDA, 2010) mentions the necessity of the assay’s reliability, and suggests to use appropriate statistical analyses
support data validation. The report of International Conference on Harmonisation of the Technique Requirements for Registration of Pharmaceuticals for Human Use (ICH, 2010) covers several aspects of validating assay data, such as specificity, accuracy, precision etc., but no specific statistical method is mentioned. Hence, one of primary objectives is to establish a statistical method to support assay validation.

In this section, we analyze data from an Anti-F IgG ELISA study about a F protein nanoparticle vaccine. In this study, a total of 736 absorbances were measured at optical densities (OD) of 450-630 nanometers (nm) at 8 different concentrations in ELISA units (EU) from 46 plates. Each plate was collected from a different laboratory. There are two replicates at each concentration for each plate. The observations from three plates in which data were recorded incorrectly were removed from the dataset and we only used the remaining 688 observations. This data can be found in Wang
et al. (2013) and is shown in Figure 4.2. Also shown in this Figure are maximum likelihood estimates of expected response function under two different ballooned beta-logistic models. More details are discussed in Section 4.1.

The initial motivating problem of assay validation was raised by Dr. Eloi Kpamegan, who is Executive Director of Clinical & Nonclinical Biostatistics at Novavax. In validation studies, the primary objective is to establish suitability criteria. Previously, this was done by fitting the classical 4PL model and obtaining estimates of $A$ and $B$ for each plate. He defined two boundaries at the minimum and the maximum concentration levels. Let $SC_1 = \hat{\mu}_A + 2 * s_1$ and $SC_2 = \hat{\mu}_B - 2 * s_1$, where $\hat{\mu}_A$, and $\hat{\mu}_B$ denote the mean of the individual plate’s least square estimates of $A$ and $B$, respectively; $s_A$ and $s_B$ denote the sample standard deviations at the minimum and maximum concentrations. Then future plates are considered suitable if the responses less than or equal to $SC_1$ at the minimum concentration, $x_{(1)}$, and responses are larger than or equal to $SC_2$ at the maximum concentration, $x_{(I)}$.

4.1 Model Selection with the BBL Family

Before assay validation, a proper model must be selected. In the general BLL model, the parameters $a$ and $b$ are expanded as $a = exp(\alpha' g(x))$ and $b = exp(\beta' h(x))$. Wu et al. (2005) mentioned that a covariate in the expansion of $b$ has less effect than does a covariate in the expansion of $a$, and they did not consider covariate effects in $b$. The dashed line in Figure 4.2 depicts the MLEs of expected responses assuming $g(x)' = (1, x)$ and $h(x) = 1$. It doesn’t go through responses at lower concentrations, but it fits well at the higher concentrations. The solid line is the MLE of expected
Table 4.1: Parameter Estimates in Exploring the Need for $\beta_2$

<table>
<thead>
<tr>
<th>Models</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_2$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{L}$</th>
<th>$\hat{U}$</th>
<th>log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>($\alpha_1, \alpha_2, \beta_1, \beta_2, L, U$)</td>
<td>2.699</td>
<td>1.602</td>
<td>2.962</td>
<td>-6.727</td>
<td>0.044</td>
<td>3.954</td>
<td>394.189</td>
</tr>
<tr>
<td>($\alpha_1, \alpha_2, \beta_1, \beta_2 = 0, L, U$)</td>
<td>1.597</td>
<td>7.215</td>
<td>1.545</td>
<td>NA</td>
<td>0.052</td>
<td>3.986</td>
<td>132.442</td>
</tr>
</tbody>
</table>

Note: NA indicates the parameter is not in the model.

responses with $g(x)' = h(x)' = (1, x)$, which fits the data considerably better.

The impact of having a covariate effect on $b$ is studied using the generalized likelihood ratio test. The hypotheses are $H_0: \beta_2 = 0$ versus $H_a: \beta_2 \neq 0$. Under null hypothesis, $g(x) = (1, x)$, $\phi(x) = 1$ and $\Theta_0 = \{\alpha_{1j}, \alpha_{2j}, \beta_{1j}, \beta_{2j} = 0, L_j, U_j, j = 1, \ldots, 46\}$; while under alternative hypothesis, $g(x) = \phi(x) = (1, x)$ and $\Theta = \{\alpha_{1j}, \alpha_{2j}, \beta_{1j}, \beta_{2j}, L_j, U_j, j = 1, \ldots, 46\}$. The likelihood under $H_0$ or $H_a$ is the product of (2.1) for all observations. Under $H_0$, $-2\log\{\lambda(Y)\} \sim \chi^2_{\alpha, \nu}$, where $\nu$ is the number of parameters in $\Theta$ minus the number of parameters in $\Theta_0$. Assuming no plate effect, the likelihood ratio test statistic is

$$
\lambda(Y) = \frac{\max_{\theta \in \Theta} \mathcal{L}(\theta; y)}{\max_{\theta \in \Theta_0} \mathcal{L}(\theta; y)},
$$

where likelihood is $\mathcal{L}(\theta; x, y) = \prod_{j=1}^{46} \prod_{i=1}^{8} \prod_{k=1}^{2} f(y_{ijk}|\theta, x_i)$.

Table 4.1 shows the maximum likelihood estimators from the two models. The critical value of 95% quantile of $\chi^2$ distribution with 1 degree of freedom is $\chi^2_{0.95, 1} = 3.841$, much smaller than $-2\log(\lambda(y)) = 523.49$ indicating $\beta_2$ should be kept in the model.
4.2 Assay Validation

4.2.1 Suitability Criteria

Following Dr. Kpamagan, suitability criteria under the BBL model can be established by using estimates of $L$ and $U$. Reference failures can be defined as plates having responses larger than $\hat{\mu}_L + 2 \cdot s_L$ at $x(1)$ or responses smaller than $\hat{\mu}_U - 2 \cdot s_U$ at $x(I)$, where $\hat{\mu}_L$ and $\hat{\mu}_U$ are the mean of the individual plate estimates of $L$ and $U$, respectively; $s_L$ and $s_U$ are the sample standard deviation of $L$ and $U$ at the minimum and maximum concentrations, respectively. Boundary estimates under the BBL and 4PL models are shown in Table 4.2. Table 4.3(a) shows these lower and upper suitability bounds for our ELISA dataset under the 4PL and BBL models. The lower bound under the 4PL model is about twice as large as the lower bound under the BBL model; conversely, the upper bound under the 4PL model is smaller than under the BBL model. Table 4.3(b) lists the information in the five plates that have failures. Their estimated asymptotes $\hat{A}$ and $\hat{B}$ under the 4PL model and $\hat{L}$ and $\hat{U}$ under the BBL model.

There are two plates falling above the $SC_1$ among the 43 plates under the BBL model and only one falling above $SC_1$ under the 4PL model. Both BBL and 4PL models detect four reference failures among the upper asymptotes. From Table 4.3, the BBL model is more sensitive than the 4PL model to detecting reference failures.

An alternative approach to establishing a suitability criteria is to evaluate (2.1) at the MLEs and integrate to obtain estimates of the 97.5th percentile at $x_1$ and the 2.5th percentile at $x_I$. Future plates having responses $y^*$ such that $Pr(y < y^*|x_1) > 0.975$; or plates having responses $y^*$ such that $Pr(y < y^*|x_I) < 0.025$ can be considered
reference failures. Percentiles are given for the five failed plates in Table 3(b). None of the five plates would be considered failures if suitability was based on the percentiles. Estimates of the percentiles are computed under assumption that all plates have same boundary, slope and $EC_{50}$ values.

4.2.2 Likelihood Ratio Test for Testing Boundary Difference

In this subsection, the BBL model is used to analyze the unequalness of difference among response boundaries among laboratories. Consider the likelihood ratio test $H_0 : L_1 = \cdots = L_J = L_0$ and $U_1 = \cdots = U_J = U_0$ with $\Theta_0 = (\alpha_j, \beta_j, L_0, U_0, j = 1, \ldots, J)$; and $H_a$: there is at least one plate has different boundaries than others with $\Theta = (\alpha_j, \beta_j, L_j, U_j, j = 1, \ldots, J)$. Figure 4.3 shows the expected response function for each plate under $H_0$ and $H_a$, respectively. Under $H_a$, the predicted response boundaries for each plate cluster tightly at the minimum and maximum concentrations.

Under the null hypothesis, in our ELISA dataset, the number of parameters is 174. Under $H_a$, the number of unknown parameters is 258. For this hypothesis test, $-2 \times \log(\lambda) = 70.19$ with $2(J - 1) = 84$ degree of freedom. Since $70.19 < \chi^2_{0.95,84} = 106.39$, $H_0$ can not be rejected. Thus this assessment indicates that no plate has boundaries that are extreme compared to others. This is consistent with our assessment of no plate failures based on predicted percentiles.

We also looked for outlying responses over the entire range of concentrations using the bootstrap method. Detection limits were defined as limits of the bootstrap prediction interval that covers 95% of the predictions from a model. The upper bound of the bootstrapped prediction limit is the 97.5% quantile of all bootstrapped
Table 4.2: Boundary Estimates for Each Plate under the 4PL and BBL Models

<table>
<thead>
<tr>
<th>Plate ID</th>
<th>$\hat{A}$</th>
<th>$\hat{L}$</th>
<th>$\hat{B}$</th>
<th>$\hat{U}$</th>
<th>Plate ID</th>
<th>$\hat{A}$</th>
<th>$\hat{L}$</th>
<th>$\hat{B}$</th>
<th>$\hat{U}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.134</td>
<td>0.016</td>
<td>4.044</td>
<td>3.951</td>
<td>26</td>
<td>0.102</td>
<td>0.035</td>
<td>4.029</td>
<td>3.949</td>
</tr>
<tr>
<td>2</td>
<td>0.052</td>
<td>0.005</td>
<td>4.029</td>
<td>3.946</td>
<td>27</td>
<td>0.132</td>
<td>0.043</td>
<td>4.050</td>
<td>3.952</td>
</tr>
<tr>
<td>3</td>
<td>0.133</td>
<td>0.045</td>
<td>4.032</td>
<td>3.952</td>
<td>28</td>
<td>0.128</td>
<td>0.033</td>
<td>4.062</td>
<td>3.947</td>
</tr>
<tr>
<td>4</td>
<td>0.153</td>
<td>0.021</td>
<td>4.041</td>
<td>3.948</td>
<td>29</td>
<td>0.163</td>
<td>0.010</td>
<td>4.047</td>
<td>3.952</td>
</tr>
<tr>
<td>5</td>
<td>0.171</td>
<td>0.065</td>
<td>4.029</td>
<td>3.954</td>
<td>30</td>
<td>0.116</td>
<td>0.045</td>
<td>4.007</td>
<td>3.948</td>
</tr>
<tr>
<td>6</td>
<td>0.153</td>
<td>0.062</td>
<td>4.011</td>
<td>3.946</td>
<td>31</td>
<td>0.123</td>
<td>0.066</td>
<td>4.012</td>
<td>3.954</td>
</tr>
<tr>
<td>7</td>
<td>0.144</td>
<td>0.055</td>
<td>4.013</td>
<td>3.952</td>
<td>32</td>
<td>0.149</td>
<td>0.007</td>
<td>4.021</td>
<td>3.951</td>
</tr>
<tr>
<td>8</td>
<td>0.153</td>
<td>0.049</td>
<td>4.036</td>
<td>3.953</td>
<td>33</td>
<td>0.131</td>
<td>0.027</td>
<td>4.026</td>
<td>3.948</td>
</tr>
<tr>
<td>9</td>
<td>0.092</td>
<td>0.042</td>
<td>4.048</td>
<td>3.951</td>
<td>34</td>
<td>0.106</td>
<td>0.030</td>
<td>4.022</td>
<td>3.935</td>
</tr>
<tr>
<td>10</td>
<td>0.121</td>
<td>0.036</td>
<td>4.030</td>
<td>3.948</td>
<td>35</td>
<td>0.094</td>
<td>0.016</td>
<td>4.046</td>
<td>3.951</td>
</tr>
<tr>
<td>11</td>
<td>0.142</td>
<td>0.036</td>
<td>4.047</td>
<td>3.952</td>
<td>36</td>
<td>0.095</td>
<td>0.047</td>
<td>4.005</td>
<td>3.969</td>
</tr>
<tr>
<td>12</td>
<td>0.101</td>
<td>0.038</td>
<td>4.054</td>
<td>3.949</td>
<td>37</td>
<td>0.130</td>
<td>0.039</td>
<td>4.035</td>
<td>3.954</td>
</tr>
<tr>
<td>13</td>
<td>0.187</td>
<td>0.050</td>
<td>4.024</td>
<td>3.954</td>
<td>38</td>
<td>0.090</td>
<td>0.024</td>
<td>4.055</td>
<td>3.951</td>
</tr>
<tr>
<td>14</td>
<td>0.144</td>
<td>0.035</td>
<td>4.028</td>
<td>3.948</td>
<td>39</td>
<td>0.110</td>
<td>0.037</td>
<td>4.048</td>
<td>3.954</td>
</tr>
<tr>
<td>15</td>
<td>0.212</td>
<td>0.133</td>
<td>4.012</td>
<td>3.951</td>
<td>40</td>
<td>0.150</td>
<td>0.044</td>
<td>3.927</td>
<td>3.869</td>
</tr>
<tr>
<td>16</td>
<td>0.173</td>
<td>0.035</td>
<td>4.033</td>
<td>3.947</td>
<td>41</td>
<td>0.136</td>
<td>0.017</td>
<td>3.932</td>
<td>3.873</td>
</tr>
<tr>
<td>17</td>
<td>0.120</td>
<td>0.031</td>
<td>4.072</td>
<td>3.950</td>
<td>42</td>
<td>0.137</td>
<td>0.039</td>
<td>3.914</td>
<td>3.853</td>
</tr>
<tr>
<td>18</td>
<td>0.109</td>
<td>0.022</td>
<td>4.061</td>
<td>3.950</td>
<td>43</td>
<td>0.142</td>
<td>0.109</td>
<td>3.886</td>
<td>3.862</td>
</tr>
<tr>
<td>19</td>
<td>0.118</td>
<td>0.043</td>
<td>4.047</td>
<td>3.947</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.119</td>
<td>0.034</td>
<td>4.049</td>
<td>3.953</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.134</td>
<td>0.043</td>
<td>4.048</td>
<td>3.949</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.136</td>
<td>0.010</td>
<td>4.064</td>
<td>3.954</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.133</td>
<td>0.043</td>
<td>4.017</td>
<td>3.949</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.149</td>
<td>0.039</td>
<td>4.030</td>
<td>3.955</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.177</td>
<td>0.055</td>
<td>4.024</td>
<td>3.951</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $\hat{L}$ and $\hat{U}$ are boundary estimates under the BBL model; $\hat{A}$ and $\hat{B}$ are boundary estimates under the 4PL model.
Table 4.3: Reference Failure Detection under the 4PL and BBL Models

a. Suitability Criteria

<table>
<thead>
<tr>
<th>Lower Suitability Bounds</th>
<th>Upper Suitability Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>4PL</td>
<td>BBL</td>
</tr>
<tr>
<td>$\hat{\mu}_A + 2s_A$</td>
<td>$\hat{\mu}_L + 2s_L$</td>
</tr>
<tr>
<td>0.197</td>
<td>0.087</td>
</tr>
</tbody>
</table>

b. Plate Suitability Failures

<table>
<thead>
<tr>
<th>Plate ID</th>
<th>Lower Suitability</th>
<th>Upper Suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4PL A A.%</td>
<td>BBL L L.%</td>
</tr>
<tr>
<td>15</td>
<td>0.212*</td>
<td>0.709</td>
</tr>
<tr>
<td>40</td>
<td>0.150</td>
<td>0.555</td>
</tr>
<tr>
<td>41</td>
<td>0.136</td>
<td>0.520</td>
</tr>
<tr>
<td>42</td>
<td>0.137</td>
<td>0.522</td>
</tr>
<tr>
<td>43</td>
<td>0.142</td>
<td>0.537</td>
</tr>
</tbody>
</table>

1. $\hat{A}$ and $\hat{B}$ are the minimum and maximum asymptotes under the 4PL model; $\hat{L}$ and $\hat{U}$ are the lower and upper boundary estimates under the BBL model.
2. % denotes the percentiles of each $\hat{A}$, $\hat{B}$, $\hat{L}$ and $\hat{U}$ under the 4PL or BBL models evaluated at the MLEs.
3. * indicates failures as determined by plates having responses exceeding the average predicted estimate +/- 2 sample standard deviations at the minimum or maximum concentrations. Plates which are not listed had no failed responses.
points. Similarly, the lower bound of the 95% bootstrapped prediction limit is the 2.5% quantile of all bootstrapped points. Details of building bootstrap prediction limits are given in Appendix A.

Figure 4.4 shows the 95% bootstrapped prediction limits of responses. A plate having responses outside the prediction limits could be considered inhomogenous with other plates. Comparing the values of response with the prediction limits, the existence of plates containing outliers could be easily identified. In Figure 4.4, we show that all responses fall in the prediction limits.

4.3 Simultaneous Multiple Comparisons of Slope and $EC_{50}$ Estimates

We compared slope and $EC_{50}$ estimates for each plate based on assuming equal boundaries which is supported by the likelihood ratio test. Estimates of $(\alpha_j, \beta_j, L_0, U_0)$, for each plate $j = 1, \ldots, J$, are obtained by maximizing the likelihood under assumption of equal boundaries. A multivariate version of Tukey’s method was applied to compare, simultaneously, slopes and $EC_{50}$ of each possible pair of plates. A simultaneous 95% confidence intervals of $S_j - S_{j^*}, EC_{50,j} - EC_{50,j^*}$, for $j \neq j^*$, were constructed. We conclude that $j$ and $j^*$ are significantly different if neither confidence interval of $S_j - S_{j^*}, EC_{50,j} - EC_{50,j^*}$ contains zero. Details of building multivariate confidence interval are included in Appendix C.

Table 4.4 shows the pairwise comparison results. Plates ranked by total number of significant differences are shown and the total number of significant differences for each column is given in the bottom row. Plate 36 has 39/42 significant differences with
other plates. However, even thorough simultaneous multiple comparisons can indicate that a given plate differences from others, the number of significant comparisons does not provide enough clear evidence to indicate plate inhomogeneity.

A series of confidence ellipsoids of $S$ and $EC_{50}$ is shown in Figure 4.5. Those points lying outside an ellipsoid indicating a plate that is significantly different from other plates. Most plates have slope and $EC_{50}$ estimates clustered within the 99% confidence ellipsoid. If we define outliers to be those plates whose slope and EC50 estimate falls outside the 99% ellipsoid, plate 2 and plate 36 are outlying plates.

4.4 A Bootstrap Comparison with Three Models

Comparing BBL with BLL and 4PL, the BBL (2.1) and the BLL (1.4) models both have smaller variances at more extreme exposure levels and have relative large response variances at central exposure levels. Even though the 4PL model (1.1) has unbounded variance, we include this model in our comparisons because of its wide use in many fields and because it has same mean function as the BBL model. In assay studies, an effective concentration is the concentration or amount of drug that produces an expected therapeutic response or desired effect that is some fixed fraction of the response range. It is commonly used as a measure of an expected potency. For example, the $EC_{50}$ is the concentration of a drug or antibody which produces expected responses halfway between the baseline and the maximum after a specific exposure time. Some distributional characteristic such as $EC_{10}$, $EC_{50}$ and $EC_{90}$ are estimated under the three different models. Parameters $A$ and $B$ in 4PL model are asymptotes of $E(y|x)$ and so don’t compare directly with $L$ and $U$, which are the
Table 4.4: Simultaneous Multiple Comparisons of Slopes and $EC_{50}$ values from ELISA Plates

<table>
<thead>
<tr>
<th>Plate</th>
<th>36</th>
<th>28</th>
<th>15</th>
<th>41</th>
<th>30</th>
<th>8</th>
<th>14</th>
<th>19</th>
<th>24</th>
<th>5</th>
<th>16</th>
<th>35</th>
<th>12</th>
<th>21</th>
<th>42</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>23</td>
<td>20</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Note: Order of plates in column is ranked by the number of significant comparisons
Table 4.5: Boundary Estimates from the ELISA study for BBL, BLL and 4PL models

<table>
<thead>
<tr>
<th>Models</th>
<th>Estimates</th>
<th>Bias*</th>
<th>SD*</th>
<th>Estimates</th>
<th>Bias*</th>
<th>SD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBL</td>
<td>$\hat{L} = 0.045$</td>
<td>0.001</td>
<td>0.004</td>
<td>$\hat{U} = 3.953$</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>BLL</td>
<td>$\hat{L} = 0.050$</td>
<td>-0.018</td>
<td>0.002</td>
<td>$\hat{U} = 3.963$</td>
<td>0.036</td>
<td>0.002</td>
</tr>
<tr>
<td>4LP</td>
<td>$\hat{A} = 0.147$</td>
<td>-0.001</td>
<td>0.002</td>
<td>$\hat{B} = 4.023$</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: Bias* and SD* are estimated bias and standard deviation from bootstrap.

lower and the upper boundaries on the actual responses in BBL and BLL model. Table 4.5 shows the estimates and bootstrapped bias and standard deviation of these estimates.

Assuming no plate effects, the estimates of $\hat{L}$ and $\hat{U}$ under the BBL and BLL models are similar. They are (0.045, 3.953) and (0.050, 3.963), respectively. Estimates of the expected response for 4PL model are (0.147, 4.023). The BBL model and the BLL model have similar estimates of the boundaries, but latter has relative large bootstrapped bias of estimates of boundaries for both $\hat{L}$ and $\hat{U}$. Those bootstrapped variances of estimates in three models are all small.

The BLL model and the BBL model produce similar values of $\hat{E}C_{10}$ and $\hat{E}C_{90}$, which are less than that use the 4PL model. The $\hat{E}C_{50}$ among three models are different but not that much. The bias of estimate in BLL is much larger than that of other two models. Under BLL model, the bias of $\hat{E}C_{10}$ and $\hat{E}C_{90}$ are −0.007 and 0.006, respectively. The bias of all these $\hat{E}C$’s under the BBL model and 4PL model are less than that of BLL.
Table 4.6: Estimates of Selected Distributional Characteristics

<table>
<thead>
<tr>
<th>Models</th>
<th>$\hat{EC}<em>{10}$ Bias* of $\hat{EC}</em>{10}$ SD* of $\hat{EC}_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBL</td>
<td>-0.214 0.002 0.014</td>
</tr>
<tr>
<td>BLL</td>
<td>-0.226 -0.007 0.034</td>
</tr>
<tr>
<td>4LP</td>
<td>-0.189 0.001 0.021</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>$\hat{EC}<em>{50}$ Bias* of $\hat{EC}</em>{50}$ SD* of $\hat{EC}_{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBL</td>
<td>0.032 -0.002 0.048</td>
</tr>
<tr>
<td>BLL</td>
<td>0.013 0.006 0.067</td>
</tr>
<tr>
<td>4LP</td>
<td>0.062 0.001 0.038</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>$\hat{EC}<em>{90}$ Bias* of $\hat{EC}</em>{90}$ SD* of $\hat{EC}_{90}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBL</td>
<td>0.276 -0.002 0.020</td>
</tr>
<tr>
<td>BLL</td>
<td>0.253 0.006 0.017</td>
</tr>
<tr>
<td>4LP</td>
<td>0.311 0.001 0.027</td>
</tr>
</tbody>
</table>

Note: Bias* and SD* are estimated bias and standard deviation from bootstrap.
Figure 4.2: Responses from the Anti-F IgG ELISA study. Dash curve depicts the expected response with $g(x)' = (1, x)$ and $\phi(x) = 1$; Solid curve has $g(x)' = \phi(x)' = (1, x)$.
Figure 4.3: Expected responses and transformed expected response for each plate. Figure (a) shows the expected response for each plate under assuming that all plates have same boundaries; Figure (b) shows the expected response for each plate considering each plate have different boundaries.
Figure 4.4: 95% bootstrapped prediction interval of responses. The dashed curve is the expected response function with $g(x)' = \phi(x)' = (1, x)$.
Figure 4.5: A series of confidence ellipsoids for 10*slope and $EC_{50}$ values under assumption that all plates have same boundaries.
Chapter 5

Summary and Concluding Remarks

Here, we summarize our main findings and point out some directions for future research.

1. In this paper, we developed a Ballooned Beta-Logistic (BBL) model, a nonlinear regression model with inhomogeneous and skewed responses variance. This new non-regular regression model can be parameterized to have the same expected response function as the four parameter logistic regression model, but with true response boundaries instead of lower and upper expected response asymptotes. Compared with the bounded log-linear regression model, the BBL model contains the parameter of slope and $EC_{50}$, which are more easily explained due to their biological interpretation. We have illustrated that the smallest and largest observations are not good estimators of the two unknown boundary parameters. However, we provided that the maximum likelihood estimates for boundaries and other parameters are consistent, asymptotically efficient and asymptotically normal. These normality results permit many questions of inference to
be addressed straightforwardly, and we illustrate some applications with our motivating data set.

2. Restricted Newton-Raphson is a standard method used to find MLEs for non-linear models. However, when multiple plates are involved, this method depends on a complex Hessian matrix with high dimension. We applied an alternative approach to find MLEs for parameters in the BBL model. We found the Newton-Raphson method over a grid of boundary parameters works well. This approach can be applied to any model which has unknown lower and upper boundaries on the responses. Given a pair of possible boundaries, with distinct \((L_j, U_j)\) or common \((L_0, U_0)\), on the grid, the remaining parameters can be estimated using the Newton-Raphson method for each plate \(j\) separately. The MLEs are one set of estimates (boundaries and remaining parameters) which reaches the maximum likelihood over the boundary grid.

3. With one covariate in each prediction function, \(a(x)\) and \(b(x)\), the BBL model has six unknown parameters, which is close to the number of observations from each plate in our motivating study, namely, 8. This may cause estimates to have large bias. However, summarized precision measures comparing BBL with 4PL and BLL models reveal that the BBL model inherits the advantages of the 4PL and BLL models. We also found that the bias of estimates of boundaries, and of the \(EC_{10}\), \(EC_{50}\) and \(EC_{90}\) are all small.

4. As in the 4PL model, the slope and \(EC_{50}\) can be expressed as functions of parameters in the BLL model. We compared \((S_j, EC_{50,j})\) for \(j = 1, \ldots, J\) in the BBL model rather than comparing parameters \((\alpha_j; \beta_j)\) for \(j = 1, \ldots, J\).
advantage of making inference on the slope and \( EC_{50} \) is that these two quantities have real toxicologic and biological interpretations. In addition, using slope and \( EC_{50} \) reduces the dimension of parameters from 6 to 4. Also, based on proven asymptotic normality of the parameters, the asymptotic normality of slope and \( EC_{50} \) was obtained using the Delta method.

5. When the expected response function doesn’t have a clear sigmoid pattern, simulated MLEs of boundaries have smaller bias and standard deviation than do the LSEs for the BBL model. LSEs for BBL model are the same as the MLEs and LSEs under the 4PL model. When the sigmoid pattern is apparent, the performance of LSEs and MLEs under the BBL model are much more similar. The MLE also permits estimation of a heteroscedastic variance. Hence we recommend the MLE approach.

6. For our motivating study, three different approaches are used to detect reference failures: suitability criteria, percentile estimation and likelihood ratio testing. The method using observed percentile of boundary estimates is more conservative than using the classical suitability criteria. Also, the likelihood ratio test shows consistent results with the observed percentiles. Five plates are found to be failures using the suitability criteria, but no differences between plates boundaries are found estimating percentile and using the likelihood ratio test.

7. We investigated differences of slope and \( EC_{50} \) between plates utilizing the asymptotic normality properties of MLEs assuming plates have same boundaries. First we considered methods for multiple comparisons, such as Tukey’s HSD method, Tukey’s range test, Bonferroni adjustment, Benjamini-Hochberg
method, etc. Tukey’s range test, which compares the difference between minimum and maximum of ordered observations, can be used to test for differences among plates given a single measurement. If there is no significant difference between minimum and maximum observations, it is reasonable to conclude that other plates not statistically different. However, existing methods are limited for multivariate statistics. Since the slope and $EC_{50}$ have different scales of measure, we couldn’t create a satisfactory summary statistic. Hence, we made all possible multivariate Turkey-type comparisons of slope and $EC_{50}$ between pairwise plates. Either of the 95% confidence intervals failing to cover zero indicates between-plate variability.

8. Larger numbers of significant differences do suggest that a plate difference is biologically important. However, multiple comparisons did not provide a powerful tool for identifying plates that are significantly differently from others. Therefore, using asymptotic normality of MLEs, we constructed confidence ellipsoids of the slope and $EC_{50}$ to show which plates are outliers. This approach is simple, straightforward and was extremely successful in identifying outlying plates in our motivating study.

9. We provided several methods to validate this ELISA bioassay dataset. Using suitability criteria can detect five reference failures; Estimating percentiles and the likelihood ratio test, which are more conservative, did not detect any failures. After dropping any failed plates detected, this assay data is valid and a future step is to estimate the potency of Anti-F IgG and to construct relevant inferential procedures under the BBL model.
Appendix A

An algorithm for generating prediction confidence bands

Denote the mean function evaluated at estimates of parameters by $\eta(x_j; \hat{\theta})$, where $j$ indexes the concentration level. Residuals between observed and predicted responses at plate $i$ and level $j$ are denoted by $r_i(x_j) = y_{ij} - \eta(x_j; \hat{\theta})$ Davison (1997); Efron and Tibshirani (1994).

For $r = 1, \ldots, R$,

1. Compute $r_i(x_j)$ from an original dataset.

2. Create a bootstrap sample response $y_{ij}^*$ at the $ith$ plate and $jth$ concentration by $y_{ij}^* = \eta(x_j; \hat{\theta}) + \epsilon_{ij}^*$, where $\epsilon_{ij}$ can be generated from empirical CDF of $r_i(x_j)$.

3. Estimate the MLE from the bootstrapped sample, $\hat{\theta}^*$, and then compute $\eta(x_j; \hat{\theta}^*)$ corresponding to a new observation at $x_j = x_{j^+}$. Then

4. Define $G$ as the size of bootstrapped sample. For $g = 1, \ldots, G$, 59
(a) Sample $\epsilon_{ij;rg}^*$ from $r_i(x_j)$ at each $x_j$;

(b) Set $y_{ij;+,rg} = \eta(x_j; \hat{\theta}^*)_+ + \epsilon_{ij;rg}^*$;

(c) Compute the bootstrap prediction error $d_{+,rg}^* = y_{ij;+,rg}^* - \eta(x_j; \hat{\theta})$.

For each $j$, order the RG values of $d_{+,rg}^*$ to obtain $d_{+(1)}^* \leq \cdots \leq d_{+(RG)}^*$. Then calculate the $(1 - \alpha)\%$ prediction limits:

the $(1 - \alpha)\%$ lower prediction limit is $y_{j;\alpha,L} = \eta(x_j; \hat{\theta}) + d_{j,+(RG+1)\alpha/2}$;

the $(1 - \alpha)\%$ upper prediction limit is $y_{j;\alpha,U} = \eta(x_j; \hat{\theta}) + d_{j,+(RG+1)(1-\alpha/2)}$,

where $\alpha$ is the nominal probability that responses will fall outside the prediction interval.
Appendix B

Methods for finding maximum likelihood estimators under the Ballooned Beta-logistic model

Maximum likelihood estimates of parameters in a beta regression model Wu et al. (2005) can be iterative computed by Newton-Raphson method. However, the Newton-Raphson method doesn’t guarantee that the MLE of two boundaries are restricted. The ballooned beta regression model requires that $L \in (-\infty, \min(Y))$ and $U \in (\max(Y), \infty)$, where $Y$ indicates response under this model. Hence, a grid-Newton-Raphson method combines the grid search and Newton-Raphson method is used to find MLE of parameters. Details are as follows:

Two constants $l$ and $u$ are arbitrary numbers with $l < \min(Y)$ and $u > \max(Y)$. To reduce computation complexity, we used $l = 0$ and $u = 4$.

1. Consider $L \in (l, \min(Y))$ and $U \in (\max(Y), u)$ by steps of length 0.001.

2. Collect all $T$ combinations of $L$ and $U$ on the grid.
3. For $t = 1, \ldots, T$,

(a) Transform the responses at dose $x$ as $w = \frac{Y - L_t}{U_t - L_t}$. Given $(L_t, U_t, x)$, the random variable $w$ follows beta distribution with parameter $a_t$ and $b_t$, where $\ln(a_t) = \alpha_1 t + \alpha_2 t * \log_2(x)/10$ and $\ln(b_t) = \beta_1 t + \beta_2 t * \log_2(x)/10$, with $x > 0$.

(b) Define $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2)$ and update $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n + [F_t'(\boldsymbol{\theta}_n)]^{-1} F_t(\boldsymbol{\theta}_n)$, where $F_t(\boldsymbol{\theta}_n)$ is the vector of likelihood equations derived from beta regression model Wu et al. (2005), and $F_t'(\boldsymbol{\theta}_n)$ is the derivative of $F_t(\boldsymbol{\theta}_n)$ with respect to all parameters in $\boldsymbol{\theta}_n$.

4. Compute maximum likelihood estimates of other parameters for all $T$ combinations of $L$ and $U$; find the set of estimators which produces the largest likelihood.
Appendix C

Simultaneous confidence procedures for multiple comparisons of mean vectors in multivariate normal populations

Building simultaneous confidence intervals for multiple comparisons among mean vectors is based on the study of Seo et al. (1995). Let $M = (\mu_1, \ldots, \mu_k)'$ be the matrix of $k$ $p$-dimensional mean vectors corresponding to the $k$ treatments. Let $\hat{M} = (\hat{\mu}_1, \ldots, \hat{\mu}_k)$ be the estimator of $M$ such that vector $\hat{M}$ is distributed as $N_{kp}(\text{vec}(M), V \otimes \Sigma)$, where $\text{vec}(.)$ denotes the column vector formed by stacking the columns of the matrix under each other and $V : k \times k$ and $\Sigma : p \times p$ are a known and an unknown positive definite matrices, respectively. Let $S$ be an unbiased estimator of $\Sigma$ such that $\nu S$ is independent of $\hat{M}$ and is distributed as a Wishart distribution $W_p(\Sigma, \nu)$. Then the usual simultaneous confidence intervals for multiple
comparisons can be written as the form

\[ a'Mb \in \left[ a'\hat{M}b \pm t(b'Vb)^{1/2}(a'Sa)^{1/2} \right], \quad \forall a \in \mathcal{R}^p, \forall b \in \mathcal{B}^k, \quad (C.1) \]

where \( \mathcal{R}^p \) is the set of any non-zero real p-dimensional vectors and \( \mathcal{B}^k \) is a subset that consists of \( r \) vectors in the \( k \)-dimensional space. Typically, value of \( t \) is hard to compute. Seo et al. (1995) proposed a modified second approximation procedure of \( t \) in (C.1).

Put \( z_i = (b_i'Vb_i)^{-1/2}(\hat{M} - M)b_i, i = 1, \ldots, r \), where \( b_i \)'s are given vectors. Then \( z_i \) has the p-dimensional normal distribution with mean vector 0 and covariance matrix \( \Sigma \). The first approximation to \( t^2 \) by \( t^2_1 \) satisfying \( \sum_{i=1}^{r} Pr\{z_i'S^{-1/2}z_i > t^2_1\} = \alpha \). Such \( t^2_1 \) can be determined by using the fact that \( z_i'S^{-1/2}z_i \) is the Hotelling \( T^2 \)-statistics with \( \nu \) degree of freedom; that is,

\[ t^2_1 = \frac{\nu p}{\nu - p + 1} F_{p,\nu-p+1}\left(\frac{\alpha}{r}\right), \]

where \( F_{p,\nu-p+1}(\alpha/r) \) is the upper \( \alpha/r \) percentile of \( F \) distribution with d.f. \( p \) and \( \nu - p + 1 \). The modified second approximation to \( t^2 \) is defined by \( t^2_M \) satisfying

\[ \sum_{i=1}^{r} Pr\{z_i'S^{-1/2}z_i > t^2_M\} = \alpha + \beta, \]

where \( \beta = \sum_{i<j} Pr\{z_i'S^{-1/2}z_i > t^2_1, z_j'S^{-1/2}z_j > t^2_1\} \). Hence, the modified second approximation \( t^2_m \) is

\[ t^2_M = \frac{\nu p}{\nu - p + 1} F_{p,\nu-p+1}\left(\frac{\alpha + \beta}{r}\right). \]
In our ELISA bioassay data, there are total 43 plates. And for each plate there are two replicates at each of 8 doses. Hence, the total number of observation is 688. Under assumption that all plates have same boundaries, the total number of parameters need to be estimated is 174, and the degree of freedom is 514.
Bibliography


VITA

Min Yi was born in Shanxi, China on April 4, 1987. After graduating with a Bachelor of Biology degree from the Shanxi University, he was recommended for admission without examination to Graduate School in Shanxi University and graduated with a Master of Science degree in Genetics. In 2010, he entered the University of Missouri and began research with Professor Nancy Flournoy in Feb 2012. He finished his Master’s degree in Statistics in 2014. Also, he participated in two summer intern programs at US Food and Drug Administration, Silver Spring, Maryland and Amgen Inc., Thousand Oaks, California in 2013 and 2014. He will join Amgen Inc. as a biostatistics manager in the summer of 2015.