Title
Prediction Intervals in Generalized Linear Mixed Models

Author
Yang, Cheng-Hsueh

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Prediction Intervals in Generalized Linear Mixed Models

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Cheng-Hsueh Yang

March 2013

Dissertation Committee:
Dr. Daniel R. Jeske, Chairperson
Dr. James M. Flegal
Dr. Xinping Cui

The Dissertation of Cheng-Hsueh Yang is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. My sincere gratitude goes to all those people who have made this dissertation possible:

First and foremost, I would like to thank to my advisor of this dissertation, Dr. Daniel R. Jeske for the valuable guidance and advice. He inspired me greatly to work in this dissertation. His willingness to motivate me contributed tremendously to my dissertation. I also would like to thank him for showing me some example that related to the topic of my dissertation.

Dr. James Flegal and Dr. Xinping Cui are dissertation committee members, for participating in my doctoral committee and the distinguished teaching of classes that build up my theoretical statistics and probability knowledge.

DEDICATIONS

Most importantly, none of this would have been possible without the love and patience of my family. My immediate family, to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my family. Especially, I would like to thank my wife Tzu-Chun for standing beside me throughout my career and writing this dissertation. She has been my inspiration and motivation for continuing to improve my knowledge and move my career forward.

Many friends have helped me through these difficult years. Their support and care helped me overcome setbacks and stay focused on my graduate studies. I greatly value their friendship and I deeply appreciate their belief in me. Especially, I am highly grateful to my best friend, Roberto Crackel, who helped me learn a lot about American culture. Thanks to Roberto Crackel for sharing in my happy moments during the process of writing this dissertation and providing encouragement when it seemed too difficult to complete. I would have probably given up without his support. Furthermore, following through on his **PHILOSOPHY**, he gave me a lot of opportunities to demonstrate my ability to answer his intractable questions during my final defense.

ABSTRACT OF THE DISSERTATION

Prediction Intervals in Generalized Linear Mixed Models

by

Cheng-Hsueh Yang

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, March 2013
Dr. Daniel R. Jeske, Chairperson

Three methods for constructing prediction intervals in a generalized linear mixed model (GLMM) are the methods based on pseudo-likelihood, Laplace, and Quadrature approximations. All three of these methods are available in the SAS procedure GLIMMIX. The pseudo-likelihood method involves approximate linearization of the GLMM into a linear mixed model (LMM) framework, and the other two methods utilize approximate conditional mean squared error (MSE) formulas for the empirical best predictor (eBP). For constructing a prediction interval, we propose two new generalized methods based on a mean squared error (MSE) approximation of the empirical best linear predictor (eBLP) and the empirical best predictor (eBP). Following the approach by Harville and Kackar (1984) for a linear mixed model (LMM), we decompose the prediction error into two terms for the purpose of deriving the MSE

approximation. Unlike in the LMM case, however, closed form expressions for the two terms in the subsequent MSE approximation are not available.

In terms of the BLP based interval, we approximate these two terms using the Taylor series expansion. In terms of the BP based interval, we confront the computational challenge by proposing a Monte Carlo algorithm for evaluating the plug-in estimators of these two terms. Furthermore, two terms from the prediction error are shown to be uncorrelated using the eBP as the predictor rather than the eBLP. The last proposed method is in regards to a highest posterior density (HPD) from a Bayesian view, using the information from the posterior distribution of random factors given the response vector. As discussed in this dissertation, a HPD interval is not as general as the BP and BLP based prediction interval because the prior distribution is limited.

For the Poisson and the Bernoulli GLMMs, simulation studies show that the methodology for our three proposed prediction intervals improves the coverage probability over the three existing methods available in GLIMMIX. Moreover, our results show that with bootstrap adjustments, our proposed BP and BLP based prediction interval achieve coverage probabilities satisfactorily close to the nominal level. However, it is intractable to derive a HPD under the negative binomial GLMMs because of the prior distribution for dispersion parameter. As we have mentioned, two proposed generalized methods are still applied to the negative binomial GLMM. The simulation results are as the same as the Poisson and Bernoulli GLMM.

TABLE OF CONTENTS

LIST OF TABLES

1. Introduction

For several years, in the linear mixed models (LMMs), much of the literature has extensively focused on point prediction such as the best predictor (BP), the best linear predictor (BLP) and the best linear unbiased predictor (BLUP). McCulloch, Searle and Neuhaus (2008) is a good reference to this literature. The BLUP can be derived using multiple methods. One common method is to use the joint maximum likelihood approach from Henderson (1950). However, in most cases, the BP, the BLP and the BLUP contain unknown parameters. Hence, the empirical best predictor (eBP), the empirical best linear predictor (eBLP) and the empirical best linear unbiased predictor (eBLUP) were proposed, replacing the unknown parameters with corresponding estimators. Kackar and Harville (1984) derived a mean squared error (MSE) approximation of the eBLUP using a Taylor series expansion. Jeske and Harville (1988) used the eBLUP and its MSE approximation to construct a prediction interval in LMMs. In addition, the SAS system has implemented much of the relevant literature about prediction intervals on LMMs in the procedure MIXED (see, for example, Littell et al. (2006) and references therein).

In LMMs, responses are assumed to be normally distributed, however, the normality assumption is not always appropriate. When faced with such difficulties, the more applicable generalized linear mixed models (GLMMs) accommodate non-normally distributed responses. The topic of prediction intervals is less well developed for the important class of GLMMs. The SAS procedure PROC GLIMMIX computes prediction intervals for GLMMs using one of the following three methods: Pseudo-likelihood (PL), Laplace (L), and Quadrature (Q). In all three methods, an estimate of the predictor and its associated precision is used to construct a $100(1-\alpha)\%$ prediction interval using normal percentiles.

The PL method is based on Wolfinger and O'Connell (1993) who proposed an algorithm to calculate the parameter estimates and the values of fixed and random effects. The main idea of this algorithm is to approximate the GLMM as a LMM with use of a pseudo-variable obtained through a Taylor series expansion. The algorithm iterates between updates of the pseudo-variable and parameter estimator that result from the LMM computations.

The Laplace method is based on Booth and Hobert (1998) who utilized a Laplace approximation based on the work of de Bruijn (1981) to approximate the value of the BP. An iterative strategy is employed to obtain the eBP, first approximating it using current values of parameters and fixed effects, and then updating those values by approximating the likelihood using another Laplace approximation. This process continues until convergence is achieved. The precision of the eBP is evaluated using a Taylor series approximation to the conditional mean squared error (CMSE) derived in Booth and Hobert (1998). Zhao et al. (2006) and Skrondal and Rabe-Hesketh (2009) have also advocated CMSE as a suitable measure of precision.

The quadrature method also calculates the BP using a Laplace approximation, however the likelihood function is approximated via an adaptive quadrature approximation [see, for example, Golub and Welsch (1969), Abramowitz and Stegun (1972) and Pinheiro and Chao (2006)]. The advantage of the adaptive quadrature approximation is to improve the approximation of the likelihood function by centering and scaling the quadrature points. Again, the same iterative strategy as the Laplace method is employed until convergence criteria is met and CMSE is used to measure the precision of the eBP. It is worth noting that because the estimated CMSE is a function of parameter estimates, its value is not the same for the Laplace and quadrature methods since these methods calculate parameter estimates differently.

This dissertation proposes three new methods for constructing prediction intervals for GLMMs that have a single random factor that captures cluster effects in the. For example, the levels of the random factor could correspond to random block effects in an ANOVA design, random hospital effects in a clinical trial design, or random intercepts in a longitudinal data analysis design. Our GLMM context covers applications where the response variable is count data modeled by distributions such as Poisson, negative binomial or Bernoulli. Our aim is to propose a better prediction interval method for linear combinations of the underlying fixed and random effects.

Let y_{ij} denote the *j*-th sampling unit within the *i*-th cluster, for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. Let $s = (s_1, \ldots, s_m)'$ denote the unobservable random cluster effects and let μ_{ij} denote the conditional (given s_i) mean of the *j*-th observation from the *i*-th cluster. Our operating GLMM is defined as follows:

a. Conditional on s_i, the observations from the *i*-th cluster $\{y_{ij}\}_{j=1}^{n_i}$ are independently distributed from distributions whose probability functions are denoted by $f(\cdot \mid \mu_{ij}, \kappa)$

b. $g(\mu_{ij}) = x_{ij}'\beta + s_i$, where $x_{ij} = (1, x_{ij1}, \ldots, x_{ij,p-1})'$ is a vector of fixed covariates associated with *j*-th observation in the *i*-th cluster and $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ is a vector of unknown parameters

c. $\{s_i\}_{i=1}^{m}$ are independent and identically distributed from a $N(0, \sigma^2)$ distribution

Although other definitions of GLMMs can be found (see, for example, McCulloch et al. 2008) the framework we use is popular for reasons including the flexibility to extend beyond independent random effects and the availability of PROC GLIMMIX to fit these models. The covariate vector can be used to describe differences between the observations that are traced to identifiable fixed effects such as treatment effects. The parameter κ may or may not be needed, depending on the model specification. If, for example, f is a negative binomial distribution, κ represents the over-dispersion parameter, whereas if f is a Poisson or Bernoulli distribution, κ is not needed in the model specification. Our main purpose is on how to construct a prediction interval for a quantity such as $w = \lambda'\beta + \delta's$, where λ and δ are known $p \times 1$ and $m \times 1$ vectors of constants, respectively. Define $\theta = (\beta', \sigma^2, \kappa)'$, where it is understood that the κ parameter may not be needed. Let the observations from the i-th cluster be collectively referred to as $y_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$, and let all the observations from all of the clusters be collectively referred to as $y = (y_1', \ldots, y_m')'$. In our second proposed method, probability functions we will subsequently use are the conditional distribution of y_i, given s_i, $f(y_i \mid s_i; \theta) = \prod_{j=1}^{n_i} f(y_{ij} \mid \mu_{ij}, \kappa)$, the zero-mean Gaussian distribution for s_i,

$\varphi(s_i;\sigma^2)$, the joint distribution of y_i and s_i, $f(y_i, s_i;\theta) = f(y_i \mid s_i;\theta)\varphi(s_i \mid \sigma^2)$, the marginal distribution of y_i, $f(y_i;\theta) = \int_{-\infty}^{\infty} f(y_i, s_i;\theta)ds_i$, and the conditional distribution of s_i, given y_i, $f(s_i \mid y_i;\theta) = f(y_i, s_i;\theta) / f(y_i;\theta)$. The conditional distribution of s, given y, can be expressed as $f(s \mid y;\theta) = \prod_{i=1}^{m} f(s_i \mid y_i;\theta)$. The integrated likelihood function is $L(\theta \mid y) = \prod_{i=1}^{m} f(y_i;\theta)$ and the maximum likelihood estimator of θ is defined as $\hat{\theta} = \arg\max_{\theta} L(\theta \mid y)$.

This paper is organized into 6 sections. The first section is the introduction and motivation, which have already been discussed. The second section reviews the concepts of the PL method, the Laplace method and the quadrature method. In section 3, our proposed method, the BLP based method and the BP based method are derived. In section 4, we use three GLMMs with a single random factor to illustrate our proposed methods. In section 5, a simulation study is used to evaluate the alternative prediction interval methods. In section 6, we provide a summary and recommendation.

2. Related Work

Section 2.1 The PL method

Wolfinger and O'Connell (1993) proposed a PL approach to GLMMs based on the LMM methodologies. Let $\mu = [\mu_{11}, \cdots \mu_{mn_m}]'$ and define $\boldsymbol{R}(\mu, \kappa)$ as a diagonal matrix whose elements are the variance of y given s where κ can be neglected in special cases such as the Bernoulli and the Poisson GLMM, and let $\mathbf{1} = (1,\cdots,1)'$ denote a vector of 1's, $\boldsymbol{I}$ denote an identity matrix and Z denote a $n \times m$ matrix where $\sum_{i=1}^{m} n_i = n$. A GLMM is approximated as $y = \mu + e$, with $g(\mu) = X\beta + Zs$, $\text{Cov}(s) = \sigma^2 \boldsymbol{I}$, $E(e \mid \mu) = \mathbf{0}$ and $\text{Var}(e \mid \mu) = \boldsymbol{R}(\mu, \kappa)$. Let $\hat{\beta}$ and $\hat{s}$ are the current estimates of β and s. We define $\hat{\mu} = g^{-1}\left(X\hat{\beta} + Z\hat{s}\right)$ which is a vector consisting of evaluations of g^{-1} at each component of $X\hat{\beta} + Z\hat{s}$. The main idea of the PL approach is to create a pseudo variable, v, through the Taylor series expansions to $e = y - \mu$ expanding about $\hat{\beta}$ and $\hat{s}$ and iteratively apply the linear mixed model equations to estimate fixed effects and predict random effects. Define $v = g(\hat{\mu}) + g'(\hat{\mu})(y - \hat{\mu})$. Here, $g'(\hat{\mu})$ is a diagonal matrix with elements $g'(\hat{\mu}_{ij})$. Define $\boldsymbol{R}(\hat{\mu}, \kappa)$ as the approximation for $\boldsymbol{R}(\mu, \kappa)$. Wolfinger and O'Connell motivate the approximation $v \mid (\beta, s) \sim \text{MVN}\left(X\beta + Zs, g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\kappa)g'(\hat{\mu}) \right)$. Since $s \sim N(0, \sigma^2 \boldsymbol{I})$, we have:

$$v \sim \text{MVN}\left(X\beta, g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\kappa)g'(\hat{\mu}) + \sigma^2 ZZ' \right) \quad \text{Equation Section 2(2.1)}$$

Based on (2.1), the profile log likelihood can be derived and used to estimate σ^2 and κ as follows:

$$l(v; \sigma^2, \kappa) \propto -\frac{1}{2}\log|\boldsymbol{V}| - \frac{n}{2}\log r'\boldsymbol{V}^{-1}r \qquad (2.2)$$

where $\boldsymbol{V} = g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\kappa)g'(\hat{\mu}) + \sigma^2 ZZ'$ and $r = v - X\left(X'\boldsymbol{V}^{-1}X\right)^{-1}X'\boldsymbol{V}^{-1}v$. The estimates of σ^2 and κ, say $\hat{\sigma}^2$ and $\hat{\kappa}$, can then be used with mixed model equations to get the next iterate of $\hat{\beta}$ and $\hat{s}$ as follows:

$$\boldsymbol{H}_1\begin{bmatrix}\hat{\beta}\\ \hat{s}\end{bmatrix} = \begin{bmatrix}X'\left(g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\hat{\kappa})g'(\hat{\mu})\right)^{-1}v\\ Z'\left(g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\hat{\kappa})g'(\hat{\mu})\right)^{-1}v\end{bmatrix} \tag{2.3}$$

Where $\boldsymbol{H}_1 = \begin{bmatrix}X'\left(g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\hat{\kappa})g'(\hat{\mu})\right)^{-1}X & X'\left(g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\hat{\kappa})g'(\hat{\mu})\right)^{-1}Z\\ Z'\left(g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\hat{\kappa})g'(\hat{\mu})\right)^{-1}X & Z'\left(g'(\hat{\mu})\boldsymbol{R}(\hat{\mu},\hat{\kappa})g'(\hat{\mu})\right)^{-1}Z + \left(\hat{\sigma}^2\right)^{-1}\boldsymbol{I}\end{bmatrix}$.

Iterations that maximize (2.2) and then solve (2.3) can continue until the iterations converge.

We are interested in a prediction interval for w. An intuitive predictor for w is $\eta(y;\hat{\theta}) = \lambda'\hat{\beta} + \delta'\hat{s}$. A naïve estimate of the covariance matrix of $\left[\hat{\beta},\hat{s}'-s'\right]'$ is $\boldsymbol{H}_1^{-1}$. Thus, a naive estimate of the MSE of $\eta(y;\hat{\theta})$ is $\boldsymbol{L}_1^T\boldsymbol{H}_1^{-1}\boldsymbol{L}_1$ where $\boldsymbol{L}_1 = (\lambda',\delta')'$ and an approximate prediction interval for w is:

$$\eta(y;\hat{\theta}) - z_{\alpha/2}\sqrt{\boldsymbol{L}_1^T\boldsymbol{H}_1^{-1}\boldsymbol{L}_1} < w < \eta(y;\hat{\theta}) + z_{\alpha/2}\sqrt{\boldsymbol{L}_1^T\boldsymbol{H}_1^{-1}\boldsymbol{L}_1} \tag{2.4}$$

If we want to get a prediction interval for $h(w)$ from (2.4), we simply evaluate $h(\cdot)$ at the end points of the interval in (2.4).

Section 2.2 The Laplace method and Quadrature method

Booth and Hobert (1998) discuss using the BP, that is $E(s\,|\,y;\theta)$, as the predictor for s.

However, obtaining a closed form solution of the BP is often not possible as the high dimensional integration becomes mathematically intractable. Therefore, a Laplace approximation is applied to approximate the BP. The authors show $E(s \mid y;\theta) = s(y;\theta) \approx \tilde{s}(y;\theta)$ and $\tilde{s}(y;\theta)$ is solved by maximizing $\log f\left(y,s\right)$. Thus, the term $\tilde{s}(y;\theta)$ is an approximate value of the best predictor of s. In practice, the parameter θ is unknown. Therefore, an iterative strategy is employed by first finding $\tilde{s}(y;\theta)$ assuming a current value of θ, and then finding a new estimate of θ by approximating the integrated likelihood using another Laplace approximation. Iterations continue until convergence, and the resulting estimator is denoted as $\hat{\theta}$.

When θ is known, the corresponding BP for w is $\eta(y;\theta) = \lambda'\beta + \delta' E(s \mid y;\theta)$. When θ is unknown, the corresponding eBP for w is $\eta(y;\hat{\theta})$. Booth and Hobert assume $\eta(y;\hat{\theta}) - \eta(y;\theta)$ and $\eta(y;\theta) - w$ are independent and represent the conditional mean squared error (CMSE) of w as follows:

$$E\left[\left(\eta(y;\hat{\theta}) - w\right)^2 \middle| y\right] \approx \mathrm{Var}[w \mid y] + E\left\{\left(\eta(y;\hat{\theta}) - \eta(y;\theta)\right)^2 \middle| y\right\} = A(y;\theta) + B(y;\theta) \quad (2.5)$$

where $\mathrm{Var}[w \mid y] = A(y;\theta)$ and $E\left\{\left(\eta(y;\hat{\theta}) - \eta(y;\theta)\right)^2 \middle| y\right\} = B(y;\theta)$.

Define $l^{(2)}(\tilde{s}(y;\theta))$ as the second derivative of $\log f\left(y,s\right)$, evaluated at $\tilde{s}(y;\theta)$. Then the term, $A(y;\theta)$, is approximated using a Laplace approximation as follows:

$$A(y;\theta) \approx \delta'\left(-l^{(2)}(\tilde{s}(y;\theta))^{-1}\right)\delta \tag{2.6}$$

The term, $B(y;\theta)$, is approximated using the first order Taylor series expansion to $s(y;\theta)$ expanding about $\hat{\theta}$ and information matrix and as follows:

$$B(y;\theta) \approx C\left(y;\hat{\theta}\right)' I^{-1}(\theta) C\left(y;\hat{\theta}\right) \tag{2.7}$$

where $I(\theta)$ is the information matrix for θ based on the GLMMs and $C\left(y;\hat{\theta}\right)' = \left(\lambda' + \delta'\left(\frac{\partial s(y;\theta)}{\partial \beta}\right), \delta'\left(\frac{\partial s(y;\theta)}{\partial \sigma^2}\right), \delta'\left(\frac{\partial s(y;\theta)}{\partial \kappa}\right)\right)\Big|_{\theta=\hat{\theta}}$. From (2.6) and (2.7), the CMSE can be approximated as follows:

$$E\left[\left(\eta(y;\hat{\theta}) - w\right)^2 \middle| y\right] \approx \delta'\left(-l^{(2)}(\tilde{s}(y;\theta))^{-1}\right)\delta + C\left(y;\hat{\theta}\right)' I^{-1}(\theta) C\left(y;\hat{\theta}\right) \tag{2.8}$$

Thus, we make use of (2.8) to construct a prediction interval for w as follows:

$$\eta(y;\hat{\theta}) - z_{\alpha/2}\sqrt{A(y;\hat{\theta}) + B(y;\hat{\theta})} < w < \eta(y;\hat{\theta}) + z_{\alpha/2}\sqrt{A(y;\hat{\theta}) + B(y;\hat{\theta})} \tag{2.9}$$

Using the same procedure discussed at the end of section 2.1, a prediction interval for $h(w)$ can also be constructed. In terms of the Quadrature method, this method finds a new estimate of θ by approximating the integrated likelihood using the quadrature approximation instead of the Laplace approximation. The rest procedures are all as the same as the Laplace method. It is worth noting that the prediction interval expression for the quadrature method is as the same as (2.9). However, the values of $\hat{\theta}$ are not estimated identically using these two methods because of different approximations to the integrated likelihood. Thus, the values of the prediction interval from these two methods are not the same either.

3. Proposed Methods

Section 3.1 BLP Based Prediction Intervals

In this section, we derive a prediction interval for w by starting with the BLP of $g^{-1}(w)$. By doing so, we match the scale of y and $g^{-1}(w)$. We then use $g[\mathrm{BLP}(g^{-1}(w))]$ as a natural predictor for w and use it to derive the prediction interval (L,U) that we use for w. A prediction interval for an arbitrary function $h(w)$ is then $(h(L),h(U))$ and will automatically be contained in the proper domain. For example, $(g^{-1}(L),g^{-1}(U))$ will lie in the appropriate parameters space for μ, which might be $[0,1]$ or R^{+}.

The BLP($g^{-1}(w)$) can be represented as follows:

$$\mathrm{BLP}(g^{-1}(w)) = \mu_{g^{-1}(w)} + \mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}(y-\mu_y) \quad \text{Equation Section 3(3.1)}$$

Here, $\mu_{g^{-1}(w)} = E(g^{-1}(w))$, $\mathrm{V}_{g^{-1}(w),y} = Cov(g^{-1}(w),y)$, $\mathrm{V}_{y,y} = Cov(y,y)$, $\mu_y = E(y)$. Thus, $\mathrm{BLP}(g^{-1}(w))$ has four terms that need to be derived: $\mu_{g^{-1}(w)}$, $\mathrm{V}_{g^{-1}(w),y}$, $\mathrm{V}_{y,y}$ and μ_y. Let $\hat{g}[\mathrm{BLP}(g^{-1}(w))]$ be the value of $g[\mathrm{BLP}(g^{-1}(w))]$ after replacing θ in

$g[\text{BLP}(g^{-1}(w))]$ with $\hat{\theta}$ in $\hat{g}[\text{BLP}(g^{-1}(w))]$. In order to use $\hat{g}[\text{BLP}(g^{-1}(w))]$ to construct a prediction interval for w, we need to approximate the mean squared error of $\hat{g}[\text{BLP}(g^{-1}(w))]$. Let $\eta(y;\theta) = g[\text{BLP}(g^{-1}(w))]$ and $\eta(y;\hat{\theta}) = \hat{g}[\text{BLP}(g^{-1}(w))]$. Define

$$\begin{aligned} e &= w - \eta(y;\hat{\theta}) \\ &= [w - \eta(y;\theta)] + \left[\eta(y;\theta) - \eta(y;\hat{\theta})\right] \end{aligned} \tag{3.2}$$

and the exact MSE is $M(\theta) = E(e^2)$. In the case of LMMs, the two terms in (3.2) are uncorrelated (Kackar and Harville, 1984). Based on simulation results discussed in Section 5, we make the assumption that in our GLMM context they are at least approximately uncorrelated to obtain at the following approximation to $M(\theta)$

$$\begin{aligned} \dot{M}(\theta) &\approx E\left[\eta(y;\theta) - w\right]^2 + E\left[\eta(y;\theta) - \eta(y;\hat{\theta})\right]^2 \\ &= M_1(\theta) + M_2(\theta). \end{aligned} \tag{3.3}$$

where $M_1(\theta) = E\left[\eta(y;\theta) - w\right]^2$ and $M_2(\theta) = E\left[\eta(y;\theta) - \eta(y;\hat{\theta})\right]^2$. The first term in (3.3) is approximated using a Taylor expansion of $g(u)$ around $g^{-1}(w)$ yields: $\eta(y;\theta) - w \approx g'(g^{-1}(w))[\text{BLP}(g^{-1}(w)) - g^{-1}(w)]$. It follows that

$$\begin{aligned} M_1(\theta) \approx &\left\{\text{E}\left\{g'(g^{-1}(w))[\text{BLP}(g^{-1}(w)) - g^{-1}(w)]\right\}\right\}^2 \\ &+ \text{Var}\left\{g'(g^{-1}(w))[\text{BLP}(g^{-1}(w)) - g^{-1}(w)]\right\} \end{aligned} \tag{3.4}$$

In Appendix A.1, we show that $\left\{E\left\{g'(g^{-1}(w))[\text{BLP}(g^{-1}(w)) - g^{-1}(w)]\right\}\right\}^2 = \left\{\boldsymbol{L}_2' E\left(\boldsymbol{T}\right)\right\}^2$ where $\boldsymbol{L}_2 = \left(\left(\mu_{g^{-1}(w)} \text{-} \text{V}_{g^{-1}(w),y} \text{V}_{y,y}^{-1} \mu_y\right) \quad \left(\text{V}_{g^{-1}(w),y} \text{V}_{y,y}^{-1}\right) \quad -1\right)'$ and

$$\boldsymbol{T}=\left(g'(g^{-1}(w)) \quad g'(g^{-1}(w))\left[E\left(y|s\right)\right]' \quad g'(g^{-1}(w))g^{-1}(w)\right)'$$ and

$$E\left(y|s\right)=g^{-1}\left(X\beta+Zs\right)=\mu$$

And we also show:

$$\begin{aligned} &Var\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\right\} \\ &\quad =\boldsymbol{L}_2'\boldsymbol{H}_2\boldsymbol{L}_2+\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y|s\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)' \end{aligned} \quad (3.5)$$

Where $\boldsymbol{H}_2=Cov\left(\boldsymbol{T}\right)$. In (3.5), the term, $Var\left(y|s\right)$, is the covariance matrix of $y|s$ and off-diagonal entries are all zero and diagonal entries are the conditional variance of a response given its associated random effects. Thus, $M_1(\theta)$ in (3.4) can be approximated as follows:

$$M_1(\theta)\approx\left\{\boldsymbol{L}_2'E\left(\boldsymbol{T}\right)\right\}^2+\boldsymbol{L}_2'\boldsymbol{H}_2\boldsymbol{L}_2+\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)E\left\{\left(g'(g^{-1}(w))\right)^2\mathrm{Var}\left(y|s\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)'$$

(3.6)

Let $d(y;\theta)=\partial\eta(y;\theta)/\partial\theta$ and define $A(\theta)=E\left[d(y;\theta)d(y;\theta)'\right]$ and $B(\theta)=I^{-1}(\theta)$, where $I(\theta)$ is Fisher's information matrix whose *ij*-th element is given by $E\left[-\partial^2\log L(\theta\,|\,y)/\partial\theta_i\partial\theta_j\right]$. Using arguments that parallel Kackar and Harville’s (1984), a second order Taylor expansion of $\eta(y;\theta)-\eta(y;\hat{\theta})$ around $\hat{\theta}=\theta$ yields

$\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]^2 \approx \left[d(y;\theta)'(\hat{\theta}-\theta)\right]^2$ and then ultimately $M_2(\theta) \approx tr\left[A(\theta)\,B(\theta)\right]$ shown in Appendix A.2. An approximation to $\dot{M}(\theta)$ is therefore

$$\ddot{M}(\theta) = \left\{\boldsymbol{L}_2' \mathrm{E}\left(\boldsymbol{T}\right)\right\}^2 + \boldsymbol{L}_2' \boldsymbol{H}_2 \boldsymbol{L}_2 + \left(\mathrm{V}_{g^{-1}(w),y} \mathrm{V}_{y,y}^{-1}\right) E\left\{\left(g'(g^{-1}(w))\right)^2 \mathrm{Var}\left(y|s\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y} \mathrm{V}_{y,y}^{-1}\right)' + tr\left[A(\theta)\,B(\theta)\right]. \tag{3.7}$$

However, $\ddot{M}(\theta)$ still depends on unknown parameters. Thus, $\ddot{M}(\hat{\theta})$ is the estimate of $\ddot{M}(\theta)$, after substituting the parameter estimates. We propose the following $100\,(1-\alpha)\%$ prediction interval for w:

$$\eta(y;\hat{\theta}) - z_{\alpha/2}\sqrt{\ddot{M}(\hat{\theta})} < w < \eta(y;\hat{\theta}) + z_{\alpha/2}\sqrt{\ddot{M}(\hat{\theta})} \tag{3.8}$$

We are also interested in comparing the interval in (3.8) with the alternative interval

$$\eta(y;\hat{\theta}) - z_{\alpha/2}\sqrt{M_1(\hat{\theta})} < w < \eta(y;\hat{\theta}) + z_{\alpha/2}\sqrt{M_1(\hat{\theta})} \tag{3.9}$$

that is based on using a (naive) estimator of MSE that ignores the expected increase in MSE caused by having to use the eBLP instead of the BLP. Using the same procedure discussed at the end of section 2.1 or 2.2, a prediction interval for $h(w)$ can also be constructed.

Recognizing that neither $(\eta(y;\hat{\theta})-w)/\sqrt{\ddot{M}(\hat{\theta})}$ nor $(\eta(y;\hat{\theta})-w)/\sqrt{M_1(\hat{\theta})}$ may be adequately approximated by a $N(0,1)$ distribution, we also consider bootstrap percentile adjustments of (3.8) and (3.9). An Algorithm 1 below shows how to obtain the bootstrap percentile adjustments.

Algorithm 1

1. Use the data y to obtain parameter estimates $\hat{\theta}$ and eBLPs of the cluster effects $\hat{s} = (\hat{s}_1 , \ldots , \hat{s}_m)'$. The eBLPs are obtained as described in Section 3.1, choosing $\lambda = 0$ and $\delta = e_i$, where e_i is zero except for a one in the *i*-th position.
2. Compute $\hat{\mu}_{ij} = g^{-1}(x_{ij}'\hat{\beta} + \hat{s}_i)$, $i = 1,\ldots, m$, $j = 1, \ldots, n_i$
3. For $k = 1$ to B (we use $B = 1000$)
4. Simulate a conditional bootstrap data set, fixing $\hat{s}$, by generating the components of $y^{(k)}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$ independently from the distributions $f(\cdot \mid \hat{\mu}_{ij};\hat{\theta})$, and compute the bootstrap estimator $\hat{\theta}^{(k)}$
5. Compute $Z_k = \left[\eta(y^{(k)};\hat{\theta}^{(k)}) - \eta(y;\hat{\theta})\right] / \sqrt{\ddot{M}(\hat{\theta}^{(k)})}$ if using interval (3.8) or instead $Z_k = \left[\eta(y^{(k)};\hat{\theta}^{(k)}) - \eta(y;\hat{\theta})\right] / \sqrt{M_1(\hat{\theta}^{(k)})}$ if using interval (3.9)
6. Next k

7. Extract lower and upper $\alpha/2$ percentiles, $L_{\alpha/2}$ and $U_{\alpha/2}$ respectively, from the $\{Z_k\}_{k=1}^{K}$ quantities

8. Construct the bootstrap percentile interval

$$\eta(y;\hat{\theta})-U_{\alpha/2}\sqrt{\ddot{M}(\hat{\theta})}<w<\eta(y;\hat{\theta})-L_{\alpha/2}\sqrt{\ddot{M}(\hat{\theta})} \quad (3.10)$$

if using interval (3.8),

$$\eta(y;\hat{\theta})-U_{\alpha/2}\sqrt{M_1(\hat{\theta})}<w<\eta(y;\hat{\theta})-L_{\alpha/2}\sqrt{M_1(\hat{\theta})} \quad (3.11)$$

if using interval (3.9)

Again, using the same procedure discussed at the end of section 2.1 or 2.2, a prediction interval for $h(w)$ can also be constructed.

Section 3.2 BP Based Prediction Interval

There are four motivations to use the BP based method instead of the BLP based method. The first, in terms of the linear assumption, if we want to use the BLP, we have to consider the scale problem because the BLP is a linear function of y. Thus, we derive a prediction interval for $w=\lambda'\beta+\delta's$ by starting with the BLP of $g^{-1}(w)$. There is no scale problem for the BP because we know the BP is not always a linear function of y. The second, in terms of the prediction error, we can prove $w-\eta(y;\theta)$ and $\eta(y;\theta)-\eta(y;\hat{\theta})$ are uncorrelated when $\eta(y;\theta)$ and $\eta(y;\hat{\theta})$ are BP and eBP, respectively which is shown in Appendix A.3. However, this result does not hold when $\eta(y;\theta)$ and $\eta(y;\hat{\theta})$ are BLP and eBLP, respectively. The third, in terms of $M_1(\theta)$, if we use the BP, then $M_1(\theta) = \delta' E\left[Var(s \mid y;\theta)\right]\delta$ (shown in Appendix A.4) is estimated by using Monte Carlo simulation which we will discuss later. If we use the BLP, then $M_1(\theta)$ can only be approximated using a Taylor expansion in (3.6). The fourth, the Laplace method in PROC GLIMMIX also derives a prediction interval by starting with the BP, and so we can make comparisons with our method to the Laplace method.

In this section, we derive a prediction interval for w by starting with the BP of w. We then use $\eta(y;\theta) = \lambda'\beta + \delta' E(s \mid y;\theta)$ as a natural predictor for w and use it to derive the prediction interval (L,U) that we use for w. Again, a prediction interval for an arbitrary function $h(w)$ is then $(h(L),h(U))$ and will automatically be contained in the proper domain. In practice, θ is unknown and the predictor used is the so-called eBP, denoted as $\eta(y;\hat{\theta})$. The prediction error of the eBP is $e = w - \eta(y;\hat{\theta})$ and the exact MSE is $M(\theta) = E(e^2)$. As we have discussed and proved $w - \eta(y;\theta)$ and $\eta(y;\theta) - \eta(y;\hat{\theta})$ are uncorrelated, we know:

$$M(\theta) = M_1(\theta) + M_2(\theta) \tag{3.12}$$

Here, the first term in (3.12) is simply $M_1(\theta) = E\left[Var(w \mid y)\right] = \delta' E\left[Var(s \mid y;\theta)\right]\delta$ and

$M_2(\theta) \approx tr\left[A(\theta)\,B(\theta)\right]$. Therefore, an approximation to $M(\theta)$ is

$$\dot{M}(\theta) = \delta' E\left[Var(s \mid y;\theta)\right]\delta + tr\left[A(\theta)\,B(\theta)\right]. \tag{3.13}$$

Here, the Reader should note that the expressions needed to evaluate $A(\theta)$ are in (A.11). Consider estimating $\dot{M}(\theta)$ by using the plug-in estimator $\dot{M}(\hat{\theta})$. In this section, we discuss the details of computing this estimator. In so doing, we will make repeated use of Algorithm 2 (outlined on the next page) which is designed to receive $\hat{\theta}$ and an arbitrary function $q(y;\theta)$ (which for our use is a matrix-valued function) and return the plug-in estimator of $E\left[q(y;\theta)\right]$ through Monte Carlo evaluation of $\hat{E}\left[q(y;\hat{\theta})\right]$, where $\hat{E}(\cdot)$ denotes the estimated expectation with respect to the distribution $f(y;\hat{\theta})$. Algorithm 2 is specifically suited for situations where $E\left[q(y;\theta)\right]$ does not have a closed-form expression. We also note that $q(y;\theta)$ itself may not have a closed form expression.

The first term in (3.13) involves $E\left[Var(s \mid y)\right]$. The plug-in estimator of this quantity can be obtained from Algorithm 2 by choosing $q(y;\theta)$ equal to the matrix $Var(s \mid y)$. [Recall that $Var(s \mid y)$ is a diagonal matrix with elements $Var(s_i \mid y_i)$]. The second term in (3.13)separately involves both $A(\theta)$ and $B(\theta)$. Since $A(\theta) = E\left[d(y;\theta)d(y;\theta)'\right]$, Algorithm 2 will return $A(\hat{\theta})$ by choosing $q(y;\theta)$ equal to the matrix $d(y;\theta)d(y;\theta)'$. Since $B(\theta)=I^{-1}(\theta)$, the plug-in estimator $B(\hat{\theta})$ requires the plug-in estimator of the matrix $I(\theta)=E\left[-\partial^2 \log L(\theta \mid y)/\partial\theta_i\partial\theta_j\right]$. Algorithm 2 will return $I(\hat{\theta})$ by choosing $q(y;\theta)$ equal to the observed information matrix $I_o(\theta)=\left[-\partial^2 \log L(\theta \mid y)/\partial\theta_i\partial\theta_j\right]$.

Algorithm 2

1. For $k=1$ to K (we use $K=1000$)
2. Simulate $s_i^{(k)}$ independently from a $N(0,\hat{\sigma}^2)$ distribution, $i=1,\ldots,m$
3. Compute $\mu_{ij}^{(k)}=g^{-1}(x_{ij}'\hat{\beta}+s_i^{(k)})$, $i=1,\ldots,m$, $j=1,\ldots,n_i$
4. Simulate $y_i^{(k)}=(y_{i1}^{(k)},\ldots,y_{in_i}^{(k)})'$, $i=1,\ldots,m$ by generating the components independently from the distributions $f(\cdot \mid \mu_{ij}^{(k)};\hat{\theta})$ and let $y^{(k)}=(y_1'^{(k)},\ldots,y_m'^{(k)})'$
5. Compute $q(y^{(k)};\hat{\theta})$
6. Next k
7. Return $\sum_{k=1}^{K} q(y^{(k)};\hat{\theta})/K$

After obtaining $\dot{M}(\hat{\theta})$, we can propose the $100(1-\alpha)\%$ prediction interval for w as follows:

$$\eta(y;\hat{\theta}) - z_{\alpha/2}\sqrt{\dot{M}(\hat{\theta})} < w < \eta(y;\hat{\theta}) + z_{\alpha/2}\sqrt{\dot{M}(\hat{\theta})} \tag{3.14}$$

The Appendix A.5 summarizes all the calculations needed to compute (3.14). We are also interested in comparing the interval in (3.14) with the alternative interval

$$\eta(y;\hat{\theta}) - z_{\alpha/2}\sqrt{M_1(\hat{\theta})} < w < \eta(y;\hat{\theta}) + z_{\alpha/2}\sqrt{M_1(\hat{\theta})} \tag{3.15}$$

Again, we also proposed the bootstrap percentile adjustments to construct the bootstrap percentile interval as discussed in (3.10) and (3.11) because the Normal approximation might be not adequate.

4. Examples

Section 4.1 Negative Binomial GLMM

Section 4.1.1 BLP Based Prediction Interval

Here we have $f(y_{ij} \mid \mu_{ij}, \kappa) = \dfrac{\Gamma(y_{ij}+\kappa)}{\Gamma(y_{ij}+1)\Gamma(\kappa)}\left(\dfrac{\mu_{ij}}{\mu_{ij}+\kappa}\right)^{y_{ij}}\left(\dfrac{\kappa}{\mu_{ij}+\kappa}\right)^{\kappa}$ and the typical link function is $g(u) = \log u$. For our illustration, we assume $\log \mu_{ij} = \beta_0 + s_i$ and consider a prediction interval for μ_{ij}. In this case we have $\theta = (\beta_0, \sigma^2, \kappa)$ and

$$f(y_i, s_i; \theta) = \prod_{j=1}^{n_i}\left\{\frac{\Gamma(y_{ij}+\kappa)}{\Gamma(y_{ij}+1)\Gamma(\kappa)}\left(\frac{\mu_{ij}}{\mu_{ij}+\kappa}\right)^{y_{ij}}\left(\frac{\kappa}{\mu_{ij}+\kappa}\right)^{\kappa}\right\}\varphi(s_i;\sigma^2)$$

and we can express the likelihood function as follows:

$$L(\theta \mid y) \propto \prod_{i=1}^{m} \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} \left\{ \frac{\Gamma(y_{ij}+\kappa)}{\Gamma(\kappa)} \left(\frac{e^{\beta_0+s_i}}{e^{\beta_0+s_i}+\kappa} \right)^{y_{ij}} \left(\frac{\kappa}{e^{\beta_0+s_i}+\kappa} \right)^{\kappa} \right\} \varphi(s_i;\sigma^2)\, ds_i$$

and the MLE is obtained as $\hat{\theta} = \arg\max_{\theta} L(\theta \mid y)$ and can then be used to evaluate $\eta(y;\hat{\theta})$.

Define $A = e^{\beta_0+\sigma^2/2}$, $B = e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right)$ and $C = e^{2\beta_0+2\sigma^2}/\kappa$. In Appendix A.6, we can show:

$$\mathrm{BLP}(g^{-1}(w)) = A + \frac{n_i B}{A+C+n_i B}\left(\bar{y}_{i.} - A\right) \tag{4.1}$$

To evaluate $M_1(\hat{\theta})$, we use the formula in (3.6) with the required inputs shown in Appendix A.7 as follows:

$\boldsymbol{L}_2'\mathrm{E}\left(\boldsymbol{T}\right) = B(A+C)/(A^2(A+C+n_i B))$

$\boldsymbol{L}_2' H_2 \boldsymbol{L}_2 = (B(A+C)^2(A^2+B)^2)/(A^4(A+C+n_i B)^2)$

$\left(\mathrm{V}_{g^{-1}(w),y_i}\mathrm{V}_{y_i,y_i}^{-1}\right)\mathrm{E}\left\{\left(g'(g^{-1}(w))\right)^2 \mathrm{Var}\left(y_i \middle| s_i\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y_i}\mathrm{V}_{y_i,y_i}^{-1}\right)' = n_i(B/(A+C+n_i B))^2((A^2+B)/A^3+1/\kappa)$ T

o evaluate $M_2(\hat{\theta})$, we have to calculate $A(\theta)$ with the required inputs shown in Appendix A.8 as follows:

Let $\rho = n_i B/(A+C+n_i B)$, $d_1 = n_i AB/(A+C+n_i B)^2$, $d_2 = (1-\rho)A$,

$d_3 = \dfrac{3nAB+2nA^3+2nA^2C}{2\left(A+C+nB\right)^2}$, $d_4 = (1-\rho)A/2$, $d_5 = n_i BC/(\kappa(A+C+n_i B)^2)$, $d_6 = 0$

and we have:

$E\left(d(y;\theta)d'(y;\theta)\right)=[E(f_{ij}(\bar{y}_{i.}))]$ where $E(f_{ij}(\bar{y}_{i.}))\approx f_{ij}(A)+f''_{ij}(A)Var[\bar{y}_{i.}]/2$ and $Var\left[\bar{y}_{i.}\right]=B+(A+C)/n_i$ and computation of $f_{ij}(A)$ and $f''_{ij}(A)$ is enabled with the required pieces by

$$f_{11}(A)=d_2^2/A^2,\quad f''_{11}(A)=(-8A\rho d_1d_2+2A^2d_1^2+6\rho^2d_2^2)/A^4,$$

$$f_{22}(A)=d_4^2/A^2,\quad f''_{22}(A)=(-8A\rho d_3d_4+2A^2d_3^2+6\rho^2d_4^2)/A^4,$$

$$f_{33}(A)=d_6^2/A^2,\quad f''_{33}(A)=(-8A\rho d_5d_6+2A^2d_5^2+6\rho^2d_6^2)/A^4,$$

$$f_{12}(A)=\frac{d_2d_4}{A^2},\ f''_{12}(A)=\left[6\rho^2d_2d_4+2A(Ad_1d_3-2\rho d_1d_4-2\rho d_2d_3)\right]/A^4$$

$$f_{13}(A)=\frac{d_2d_6}{A^2},\quad f_{13}''(A)=\left[6\rho^2d_2d_6+2A(Ad_1d_5-2\rho d_1d_6-2\rho d_2d_5)\right]/A^4$$

$$f_{23}(A)=\frac{d_6d_4}{A^2},\ f''_{23}(A)=\left[6\rho^2d_4d_6+2A(Ad_3d_5-2\rho d_3d_6-2\rho d_4d_5)\right]/A^4$$

The proposed $100(1-\alpha)\%$ prediction interval for $w=\beta_0+s_i$ as discussed in section 3.1.

Section 4.1.2 BP Based Prediction Interval

Following the computational summary outlined in the Appendix A.5, $E(s_i\mid y_i;\theta)$ can be obtained from (A.9) and used to evaluate $\eta(y;\theta)=\beta_0+E(s_i\mid y_i;\theta)$. And the MLE can then be used to evaluate $\eta(y;\hat{\theta})$.

To evaluate $\dot{M}(\hat{\theta})$, we use Algorithm 2 as described in Section 3.2, using referenced formulas in the Appendix to derive the required inputs. $Var(s_i\mid y_i;\theta)$ can be computed directly using equation (A.10). Let $\psi(u)$ denote the digamma function. Computation of $\partial\eta(y;\theta)/\partial\beta_0$, $\partial\eta(y;\theta)/\partial\sigma^2$ and $\partial\eta(y;\theta)/\partial\kappa$ is enabled by (A.11) with the required pieces (A.12)-(A.14) and (A.15)-(A.17) being given by

$$\frac{\partial f(y_i,s_i;\theta)}{\partial\beta_0}=\frac{\kappa}{e^{\beta_0+s_i}+\kappa}(y_{i\cdot}-n_i e^{\beta_0+s_i})\,f(y_i,s_i;\theta)$$

$$\frac{\partial f(y_i,s_i;\theta)}{\partial\sigma^2}=\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)f(y_i,s_i;\theta)$$

$$\frac{\partial f(y_i,s_i;\theta)}{\partial\kappa}=\left\{\sum_{j=1}^{n_i}\psi(y_{ij}+\kappa)-n_i\psi(\kappa)-\frac{y_{i\cdot}-n_i e^{\beta_0+s_i}}{e^{\beta_0+s_i}+\kappa}+n_i\log\left(\frac{\kappa}{e^{\beta_0+s_i}+\kappa}\right)\right\}f(y_i,s_i;\theta)$$

$$\frac{\partial f(y_i;\theta)}{\partial\beta_0}=\int_{-\infty}^{\infty}\frac{\kappa}{e^{\beta_0+s_i}+\kappa}(y_{i\cdot}-n_i e^{\beta_0+s_i})\,f(y_i,s_i;\theta)\;ds_i$$

$$\frac{\partial f(y_i;\theta)}{\partial\sigma^2}=\int_{-\infty}^{\infty}\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)f(y_i,s_i;\theta)\;ds_i$$

$$\frac{\partial f(y_i;\theta)}{\partial\kappa}=\int_{-\infty}^{\infty}\left\{\sum_{j=1}^{n_i}\psi(y_{ij}+\kappa)-n_i\psi(\kappa)-\frac{y_{i\cdot}-n_i e^{\beta_0+s_i}}{e^{\beta_0+s_i}+\kappa}+n_i\log\left(\frac{\kappa}{e^{\beta_0+s_i}+\kappa}\right)\right\}f(y_i,s_i;\theta)\;ds_i$$

Finally, the observed information matrix quantities are calculated from (A.18)-(A.23). The quantities needed for these formulas are (A.24)-(A.29) and are respectively given by

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\beta_0}=\left\{\left[\frac{\kappa}{e^{\beta_0+s_i}+\kappa}(y_{i.}-n_ie^{\beta_0+s_i})\right]^2-\left[\frac{\kappa e^{\beta_0+s_i}}{(e^{\beta_0+s_i}+\kappa)^2}\left(y_{i.}-n_ie^{\beta_0+s_i}\right)+\frac{\kappa}{e^{\beta_0+s_i}+\kappa}n_ie^{\beta_0+s_i}\right]\right\}$$
$$\times f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\sigma^2\partial\beta_0}=\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)\frac{\kappa}{e^{\beta_0+s_i}+\kappa}(y_{i.}-n_ie^{\beta_0+s_i})\,f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\kappa\partial\beta_0}=\left\{\left[\sum_{j=1}^{n_i}\psi(y_{ij}+\kappa)-n_i\psi(\kappa)-\frac{y_{i.}-n_ie^{\beta_0+s_i}}{e^{\beta_0+s_i}+\kappa}+n_i\log\left(\frac{\kappa}{e^{\beta_0+s_i}+\kappa}\right)\right]\right.$$
$$\left.\times\left[\frac{\kappa}{e^{\beta_0+s_i}+\kappa}\left(y_{i.}-n_ie^{\beta_0+s_i}\right)\right]+\frac{e^{\beta_0+s_i}(y_{i.}-n_ie^{\beta_0+s_i})}{(e^{\beta_0+s_i}+\kappa)^2}\right\}f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\sigma^2}=\left[\frac{1}{4\sigma^4}\left(\frac{s_i^2}{\sigma^2}-1\right)^2-\frac{1}{\sigma^4}\left(\frac{s_i^2}{\sigma^2}-\frac{1}{2}\right)\right]f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\kappa\partial\sigma^2}=\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)\left\{\sum_{j=1}^{n_i}\psi(y_{ij}+\kappa)-n_i\psi(\kappa)-\frac{y_{i.}-n_ie^{\beta_0+s_i}}{e^{\beta_0+s_i}+\kappa}\right.$$
$$\left.+\;n_i\log\left(\frac{\kappa}{e^{\beta_0+s_i}+\kappa}\right)\right\}f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\kappa}=\left\{\left[\sum_{j=1}^{n_i}\psi(y_{ij}+\kappa)-n_i\psi(\kappa)-\frac{y_{i.}-n_ie^{\beta_0+s_i}}{e^{\beta_0+s_i}+\kappa}+n_i\log\left(\frac{\kappa}{e^{\beta_0+s_i}+\kappa}\right)\right]^2\right.$$
$$\left.+\;\sum_{j=1}^{n_i}\psi'(y_{ij}+\kappa)-n_i\psi'(\kappa)+\frac{\kappa y_{i.}+n_ie^{2(\beta_0+s_i)}}{\kappa(e^{\beta_0+s_i}+\kappa)^2}\right\}f(y_i,s_i;\theta)$$

Then, the proposed $100(1-\alpha)\%$ prediction interval for $w=\beta_0+s_i$ as obtained as discussed in section 3.2.

Section 4.2 The Poisson GLMM

Section 4.2.1 BLP Based Prediction Interval

Here we have $f(y_{ij} \mid \mu_{ij}) = \exp(-\mu_{ij})\mu_{ij}^{y_{ij}} / y_{ij}!$ and the typical link function is $g(u) = \log u$. We again assume $\log \mu_{ij} = \beta_0 + s_i$ and consider a prediction interval for $w = \beta_0 + s_i$. In this case we have $\theta = (\beta_0, \sigma^2)$ and

$$f(y_i, s_i; \theta) = \exp\left[-n_i e^{\beta_0 + s_i} + (\beta_0 + s_i) y_{i\cdot}\right] \varphi(s_i; \sigma^2) / \prod_{j=1}^{n_i} y_{ij}! \ .$$

The likelihood function is $L(\theta \mid y) \propto \prod_{i=1}^{m} \int_{-\infty}^{\infty} \exp\left[-n_i e^{\beta_0 + s_i} + (\beta_0 + s_i) y_{i\cdot}\right] \varphi(s_i; \sigma^2) ds_i$ and the MLE is obtained as $\hat{\theta} = \arg\max_{\theta} L(\theta \mid y)$ and can then be used to evaluate $\eta(y; \hat{\theta})$.

Define $\rho = nB / (A + nB)$ with $A = e^{\beta_0 + \sigma^2/2}$ and $B = e^{2\beta_0 + \sigma^2}\left(e^{\sigma^2} - 1\right)$. In Appendix A.9, we can show:

$$\mathrm{BLP}(g^{-1}(w)) = (1 - \rho)A + \rho \bar{y}_{i.} \tag{4.2}$$

To evaluate $M_1(\hat{\theta})$, we use the formula in (3.6) to calculate $M_1(\theta)$ in Appendix A.10 as follows:

$$M_1(\theta) \approx B\left(\frac{1}{A + nB}\right)^2 \left(1 + \frac{nB}{A} + \frac{3B}{A^2} + \frac{nB^2}{A^3} + \frac{B^2}{A^4}\right) \tag{4.3}$$

To evaluate $M_2(\hat{\theta})$, we have to calculate $M_2(\theta)$ by evaluating $A(\theta)$ in Appendix A.11 as follows:

$$M_2(\theta) \approx tr\left[(1-\rho)^2 \begin{bmatrix} 1 & \frac{1}{2} - \frac{\rho\left(\rho + nA(1-\rho)\right)}{nA(1-\rho)} \\ \frac{1}{2} - \frac{\rho\left(\rho + nA(1-\rho)\right)}{nA(1-\rho)} & \frac{1}{4} + \rho + nA(1-\rho) \end{bmatrix} E\left(\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)'\right)\right]$$

(4.4)

The proposed $100\,(1-\alpha)\%$ prediction interval for μ_{ij} as obtained by letting $h = g^{-1}$ as discussed in section 3.1.

Section 4.2.2 BP Prediction Interval

Because the negative binomial distribution becomes the Poisson distribution when $\kappa = \infty$, all of the results in Section 4.1.2 pertaining to evaluating $\dot{M}(\hat{\theta})$ apply to the Poisson GLMM case when the following two conventions are adopted: i) use the above formula for $f(y_i, s_i; \theta)$, ii) only use the derivative equations that are with respect to β_0 and/or σ^2, and use their limiting form as $\kappa \to \infty$.

Following the computational summary outlined in the Appendix A.5, $E(s_i \mid y_i; \theta)$ can be obtained from (A.9) and used to evaluate $\eta(y;\theta) = \beta_0 + E(s_i \mid y_i; \theta)$. And the MLE can then be used to evaluate $\eta(y;\hat{\theta})$. To evaluate $\dot{M}(\hat{\theta})$, we use Algorithm 2 as described in Section 2.2, using referenced formulas in the Appendix to derive the required inputs. $Var(s_i \mid y_i; \theta)$ can be computed directly using equation (A.10). Computation of $\partial\eta(y;\theta)/\partial\beta_0$ and $\partial\eta(y;\theta)/\partial\sigma^2$ is enabled by (A.11) with the required pieces (A.12)

-(A.14) and (A.15)-(A.17) being given by

$$\frac{\partial f(y_i, s_i; \theta)}{\partial \beta_0} = (y_{i\cdot} - n_i e^{\beta_0 + s_i}) f(y_i, s_i; \theta)$$

$$\frac{\partial f(y_i, s_i; \theta)}{\partial \sigma^2} = \frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2} - 1\right) f(y_i, s_i; \theta)$$

$$\frac{\partial f(y_i; \theta)}{\partial \beta_0} = \int_{-\infty}^{\infty} (y_{i\cdot} - n_i e^{\beta_0 + s_i}) f(y_i, s_i; \theta)\ ds_i$$

$$\frac{\partial f(y_i; \theta)}{\partial \sigma^2} = \int_{-\infty}^{\infty} \frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2} - 1\right) f(y_i, s_i; \theta)\ ds_i$$

Finally, the observed information matrix quantities are calculated from (A.18), (A.19) and (A.21). The quantities needed for these formulas are (A.24), (A.25) and (A.27)are respectively given by

$$\frac{\partial^2 f(y_i, s_i; \theta)}{\partial^2 \beta_0} = \left\{(y_{i\cdot} - n_i e^{\beta_0 + s_i})^2 - n_i e^{\beta_0 + s_i}\right\} f(y_i, s_i; \theta)$$

$$\frac{\partial^2 f(y_i, s_i; \theta)}{\partial \sigma^2 \partial \beta_0} = \frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2} - 1\right)(y_{i\cdot} - n_i e^{\beta_0 + s_i}) f(y_i, s_i; \theta)$$

$$\frac{\partial^2 f(y_i, s_i; \theta)}{\partial^2 \sigma^2} = \left[\frac{1}{4\sigma^4}\left(\frac{s_i^2}{\sigma^2} - 1\right)^2 - \frac{1}{\sigma^4}\left(\frac{s_i^2}{\sigma^2} - \frac{1}{2}\right)\right] f(y_i, s_i; \theta)$$

The proposed $100(1-\alpha)\%$ prediction interval for μ_{ij} as obtained by letting $h = g^{-1}$ as discussed in section 3.2.

Section 4.2.3 Bayes Interval

Define $w_i = e^{\beta_0 + s_i}$, $\boldsymbol{w} = \left(w_1, \cdots w_m\right)'$ and $y_i = \left(y_{i1}, \cdots y_{in}\right)'$. In Appendix A.12, we can show:

$$f(\boldsymbol{w} \mid y) \propto e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \int_0^{\infty} \int_{-\infty}^{\infty} \left(\sigma^2\right)^{-\frac{m}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(\log w_i - \beta_0\right)^2} f(\beta_0, \sigma^2) d\beta_0 d\sigma^2 \tag{4.5}$$

We assign the non-informative prior to $f(\beta_0)$ and the inverse gamma distribution to $f(\sigma^2)$ in (4.5). Define $\overline{\log w} = \sum_{i=1}^{m} \frac{\log w_i}{m}$ and $ms^2 = \sum_{i=1}^{m} \left(\log w_i - \overline{\log w}\right)^2$, and so we can show the following result in Appendix A.13:

$$f(\boldsymbol{w} \mid y) \propto e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \left(\frac{ms^2 + 2\beta}{2}\right)^{-\left(\alpha + \frac{m-1}{2}\right)} \tag{4.6}$$

Our strategy is to generate a sample from the multivariate distribution of $\boldsymbol{w} \mid y$ in (4.6) using the random walk Metropolis–Hastings algorithm as follows. This algorithm starts with a given $\boldsymbol{w}^{(0)} = \left(w_1^{(0)}, \cdots w_m^{(0)}\right)^T$ and a symmetric m-dimensional distribution h. For $t = 1, 2 \cdots$, let $\boldsymbol{w}^{(t)} = \boldsymbol{w}^{(t-1)} + \varepsilon^{(t)}$ where $\varepsilon^{(t)} \sim h(\cdot)$. Then compute

$$\alpha\left(\boldsymbol{w}^{(t)} \mid \boldsymbol{w}^{(t-1)}\right) = \min\left(1, \frac{f\left(\boldsymbol{w}^{(t)} \mid y\right)}{f\left(\boldsymbol{w}^{(t-1)} \mid y\right)}\right) \tag{4.7}$$

Here, $f\left(\boldsymbol{w} \mid y\right)$ is derived using (4.6). We accept $\boldsymbol{w}^{(t)}$ as the next draw from the posterior distribution with probability (4.7), otherwise we set $\boldsymbol{w}^{(t)} = \boldsymbol{w}^{(t-1)}$. After applying the Metropolis–Hastings algorithm, a Markov chain of $\left(\boldsymbol{w}^{(1)},\boldsymbol{w}^{(2)},\cdots\right)$ is used as a sample from the posterior distribution. When constructing a $100\left(1-\alpha\right)\%$ prediction interval for w_i, we consider a highest posterior density (HPD) interval using another Monte Carlo algorithm developed by Chen and Shao (1999). This Monte Carlo algorithm can be used to search for two percentiles which have the shortest interval length for a $100\left(1-\alpha\right)\%$ prediction interval, using the criteria of minimizing the interval length between two percentiles from a MCMC draws.

Section 4.3 The Bernoulli GLMM

Section 4.3.1 BLP Based Prediction Interval

Here we have $f(y_{ij}\mid\mu_{ij})=\mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}}$ and the typical link function is $g(u)=\log(u/(1-u))$. For our illustration, we assume $\log\left[\mu_{ij}/(1-\mu_{ij})\right]=\beta_0+s_i$ and consider a prediction interval for μ_{ij}. In this case we have $\theta=(\beta_0,\sigma^2)$ and

$$f(y_i,s_i;\theta)=\frac{e^{(\beta_0+s_i)y_{i\cdot}}}{(1+e^{\beta_0+s_i})^{n_i}}\varphi(s_i;\sigma^2)$$

and we can express the likelihood function as follows:

$$L(\theta\mid y)\propto\prod_{i=1}^{m}\int_{-\infty}^{\infty}\frac{e^{(\beta_0+s_i)y_{i\cdot}}}{(1+e^{\beta_0+s_i})^{n_i}}\varphi(s_i;\sigma^2)ds_i$$

and the MLE is obtained as $\hat{\theta}=\arg\max_{\theta}L(\theta\mid y)$ and can then be used to evaluate $\eta(y;\hat{\theta})$.

Define $\mathrm{E}\left[e^{\beta_0+s_i}\Big/\left(1+e^{\beta_0+s_i}\right)\right]$ and $\mathrm{E}\left[e^{\beta_0+s_i}\Big/\left(1+e^{\beta_0+s_i}\right)\right]^2$ are calculated numerically and denoted as T_1 and T_2 respectively. In Appendix A.14, we can show:

$$\mathrm{BLP}(g^{-1}(w)) = T_1 + \frac{\left(T_2 - T_1^2\right)}{\left(T_1 - T_2\right) + n_i\left(T_2 - T_1^2\right)}\left(y_{i.} - n_i T_1\right) \tag{4.8}$$

To evaluate $M_1(\hat{\theta})$, we use the formula in (3.6) with the required inputs shown in Appendix A.15 as follows:

Let $A_1 = T_1 - T_2$ and $B_1 = T_2 - T_1^2$ and denote $H_{2,ij}$ as the element in the i th row and j th column of $\boldsymbol{H}_2$:

$$\boldsymbol{L}_2 = \left(T_1 - nT_1 B_1 / (A_1 + nB_1) \quad B_1 / (A_1 + nB_1)\mathbf{1}' \quad -1\right)$$

$$E\left(\boldsymbol{T}\right) = \left(1/A + B/A^3 + 2 + A \quad (1+A)\mathbf{1}' \quad 1+A\right)'$$

$$H_{2,11} = 2B\left(1/A + B/A^3 - 1\right)\Big/A^2 + B\ , \quad H_{2,22} = B\boldsymbol{J}\ , \quad H_{2,33} = B, \quad H_{2,12} = \left(-B/A^2 + 2B\right)\mathbf{1},$$

$$H_{2,13} = -B/A^2 + 2B, \quad H_{2,23} = B\mathbf{1}$$

$$E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i \middle| s_i\right)\right\} = \left(\left(A^2 + B\right)\Big/A^3 + A + 2\right)\boldsymbol{I} \quad \text{and}$$

$$\mathrm{V}_{g^{-1}(w),y_i}\mathrm{V}_{y_i,y_i}^{-1} = [B_1 / (A_1 + n_i B_1)]\mathbf{1}'.$$

To evaluate $M_2(\hat{\theta})$, we have to calculate $A(\theta)$ with the required inputs shown in Appendix A.16 as follows:

Define

$$T_3 = \mathrm{E}\left[\left(e^{\beta_0+s_i}\right)^2 \Big/ \left(1+e^{\beta_0+s_i}\right)^3\right],$$

$$T_4 = \int \sigma^{-3} e^{-\frac{s_i^2}{2\sigma^2}+\beta_0+s_i}\left(-1+s_i^2\sigma^{-2}\right)/[2\sqrt{2\pi}(1+e^{\beta_0+s_i})]\ \mathrm{ds}_i$$

$$T_5 = \int \sigma^{-3} e^{-\frac{s_i^2}{2\sigma^2}+2(\beta_0+s_i)}\left(-1+s_i^2\sigma^{-2}\right)/[2\sqrt{2\pi}(1+e^{\beta_0+s_i})^2]\ \mathrm{ds}_i$$

and T_3, T_4 and T_5 are calculated numerically and $\rho = n_i B_1 / (A_1 + n_i B_1)$

$E\left(d(y;\theta)d'(y;\theta)\right) = [E(f_{ij}(\bar{y}_{i.}))]$ where $E(f_{ij}(\bar{y}_{i.})) \approx f_{ij}(T_1) + f''_{ij}(T_1)Var[\bar{y}_{i.}]/2$ and

$Var\left[\bar{y}_{i.}\right] = T_2 - T_1^2 + \dfrac{1}{n_i}\left(T_1 - T_2\right)$ and computation of $f_{ij}(T_1)$ and $f''_{ij}(T_1)$ is enabled with

the required pieces by

$$f_{11}(T_1) = \frac{c_2^2}{\left(T_1(1-T_1)\right)^2}$$

$$f_{11}''(T_1) = \frac{4c_2^2}{[T_1(1-T_1)]^2}\left[\frac{c_1}{c_2} - \frac{\rho}{T_1} + \frac{\rho}{1-T_1}\right]^2 - \frac{2c_2^2}{[T_1(1-T_1)]^2}\left[\frac{c_1^2}{c_2^2} - \frac{\rho^2}{T_1^2} - \frac{\rho^2}{(1-T_1)^2}\right]$$
$$= \frac{2c_2^2}{[T_1(1-T_1)]^2}\left[2\left[\frac{c_1}{c_2} - \frac{\rho}{T_1} + \frac{\rho}{1-T_1}\right]^2 - \left[\frac{c_1^2}{c_2^2} - \frac{\rho^2}{T_1^2} - \frac{\rho^2}{(1-T_1)^2}\right]\right]$$

$$f_{12}(T_1) = \frac{c_2 c_4}{\left(T_1(1-T_1)\right)^2}$$ and

$$f_{12}''(T_1) = \frac{c_2 c_4}{[T_1(1-T_1)]^2}\left[\left[\frac{c_1}{c_2} + \frac{c_3}{c_4} - \frac{2\rho}{T_1} + \frac{2\rho}{1-T_1}\right]^2 + \left[-\frac{c_1^2}{c_2^2} - \frac{c_3^2}{c_4^2} + \frac{2\rho^2}{T_1^2} + \frac{2\rho^2}{(1-T_1)^2}\right]\right]$$

$$f_{22}(T_1) = \frac{c_4^2}{\left(T_1(1-T_1)\right)^2}$$

$$f_{22}''(T_1) = \frac{2c_4^2}{[T_1(1-T_1)]^2}\left[2\left[\frac{c_3}{c_4} - \frac{\rho}{T_1} + \frac{\rho}{1-T_1}\right]^2 - \left[\frac{c_3^2}{c_4^2} - \frac{\rho^2}{T_1^2} - \frac{\rho^2}{(1-T_1)^2}\right]\right]$$

Where c_1, c_2, c_3, and c_4 are defined in Appendix A.16. The proposed $100(1-\alpha)\%$ prediction interval for μ_{ij} as obtained by letting $h = g^{-1}$ as discussed in section 3.1.

Section 4.3.2 BP Based Prediction Interval

Following the computational summary outlined in the Appendix A.5, $E(s_i \mid y_i;\theta)$ can be obtained from (A.9) and used to evaluate $\eta(y;\theta) = \beta_0 + E(s_i \mid y_i;\theta)$. And the MLE can then be used to evaluate $\eta(y;\hat{\theta})$. To evaluate $\dot{M}(\hat{\theta})$, we use Algorithm 2 as described in Section 2.2, using referenced formulas in the Appendix to derive the required inputs. $Var(s_i \mid y_i;\theta)$ can be computed directly using equation (A.10). Computation of $\partial\eta(y;\theta)/\partial\beta_0$ and $\partial\eta(y;\theta)/\partial\sigma^2$ is enabled by (A.11) with the required pieces (A.12)-(A.14) and (A.15)-(A.17) being given by

$$\frac{\partial f(y_i, s_i;\theta)}{\partial\beta_0} = \left(y_{i\cdot} - \frac{n_i e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}} \right) f(y_i, s_i;\theta)$$

$$\frac{\partial f(y_i, s_i;\theta)}{\partial\sigma^2} = \frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2} - 1 \right) f(y_i, s_i;\theta)$$

$$\frac{\partial f(y_i;\theta)}{\partial\beta_0} = \int_{-\infty}^{\infty} \left(y_{i\cdot} - \frac{n_i e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}} \right) f(y_i, s_i;\theta)\, ds_i$$

$$\frac{\partial f(y_i;\theta)}{\partial\sigma^2} = \int_{-\infty}^{\infty} \frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2} - 1 \right) f(y_i, s_i;\theta)\, ds_i \,.$$

Finally, the observed information matrix quantities are calculated from (A.18), (A.19) and (A.21). The quantities needed for these formulas are (A.24), (A.25) and (A.27)are respectively given by

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\beta_0}=\left[\left(y_{i\cdot}-\frac{n_i e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)^2-\frac{n_i e^{\beta_0+s_i}}{(1+e^{\beta_0+s_i})^2}\right]f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\sigma^2\partial\beta_0}=\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)\left(y_{i\cdot}-\frac{n_i e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)f(y_i,s_i;\theta)$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\sigma^2}=\left[\frac{1}{4\sigma^4}\left(\frac{s_i^2}{\sigma^2}-1\right)^2-\frac{1}{\sigma^4}\left(\frac{s_i^2}{\sigma^2}-\frac{1}{2}\right)\right]f(y_i,s_i;\theta)\quad .$$

The proposed $100(1-\alpha)\%$ prediction interval for μ_{ij} as obtained by letting $h=g^{-1}$ as discussed in section 3.2.

Section 4.3.3 Bayes Interval

Define $w_i=\dfrac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}$. In Appendix A.17, we can show:

$$f(\boldsymbol{w}\,|\,y)\propto\prod_{i=1}^{m}w_i^{y_{i\cdot}-1}\prod_{i=1}^{m}\left(1-w_i\right)^{n_i-y_{i\cdot}-1}\int_0^{\infty}\int_{-\infty}^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\beta_0\right)^2}f(\beta_0,\sigma^2)\,d\beta_0\,d\sigma^2 \quad (4.9)$$

We assign the non-informative prior to $f(\beta_0)$ and the inverse gamma distribution to $f(\sigma^2)$ in (4.9). Define $\overline{\log w}=\sum_{i=1}^{m}\dfrac{\log w_i}{m}$ and $ms^2=\sum_{i=1}^{m}\left(\log w_i-\overline{\log w}\right)^2$, and so we can show the following result in Appendix A.18:

$$f(\boldsymbol{w}\,|\,y)\propto\prod_{i=1}^{m}w_i^{y_{i\cdot}-1}\prod_{i=1}^{m}\left(1-w_i\right)^{n_i-y_{i\cdot}-1}\left(\frac{ms^2+2\beta}{2}\right)^{-\left(\alpha+\frac{m-1}{2}\right)} \qquad (4.10)$$

Again, we can construct a HPD interval as discussed in section 4.2.3 using (4.10).

5 Performance Evaluations

To compare the proposed prediction intervals with the intervals in SAS, we performed a simulation study to evaluate coverage probability and expected width of the alternative intervals. In terms of alternative GLMMs, we considered the three illustrative examples presented in Section 4. In terms of the proposed prediction intervals, we included both (3.8), (3.9), (3.14) and (3.15) in the simulation study – along with their bootstrap adjusted versions and a HPD interval. We compared those four intervals to the three intervals in SAS obtained by using the PL, L and Q methods.

For simulation parameters, we took $m \in \{10, 20\}$ and set $n_i \equiv n$ for $n \in \{5, 10\}$. When considering the negative binomial GLMM, we varied $\kappa \in \{1, 2, 5\}$. We also varied $\beta_0 \in \{1, 2\}$ and $\sigma \in \{.2, .4\}$. These combinations of parameter values correspond to the response variable having a mean that ranges between 2 and 8 and variance that ranges between 3 and 94. Because it is intractable to demonstrate a HPD interval under the negative binomial GLMM with $\kappa = \infty$, we illustrate a HPD interval with the Poisson GLMM, we varied $\beta_0 \in \{1, 1.5, 2, 2.5\}$ and $\sigma \in \{.2, .4\}$. These combinations of parameter values correspond to the response variable having a mean that ranges between 3 and 13 and variance that ranges between 3 and 43. For the Bernoulli GLMM, we varied $\beta_0 \in \{-2.5, -.5, .5, 2\}$ and $\sigma \in \{.2, 2\}$. These choices of parameter values corresponded to success probabilities that ranged from about 0.2 to 0.8. Using the Bernoulli GLMM, we can also derive a HPD interval. For each scenario of parameter settings and sample size values, we simulated 1000 data sets from the GLMM and then evaluated each of the alternative prediction intervals. The percentage of prediction intervals for each method that covered parameter was recorded.

Tables 1-4 and 6-9 show the coverage probabilities and Table 5 and 10 show expected widths of all the alternative prediction intervals for the scenarios covering the negative binomial and Poisson GLMMs, respectively. Table11-14 and Table 15 show corresponding results for the Bernoulli GLMMs.

We see that the three prediction interval methods implemented in SAS have coverage probabilities that are lower than expected, and in general, there is very little difference between these three methods. While intervals (3.8), (3.9), (3.14) and (3.15) have coverage probability closer to the nominal level, their bootstrapped versions offer the best solutions. Tables1-4, 6-9 and 11-14 show that bootstrap adjustments were less effective with the three SAS intervals and generally not adequate enough to make the coverage probabilities satisfactory. Using the BP base method, it is worth noting that we tried using a bias correction with this $2tr(A(\theta)B(\theta))$ (see, Harville and Jeske (1992)) and it resulted in only a modest 1% increase in the coverage probability. From Table 5,10 and 15 we can see increasing the number of clusters is more important than increasing the number of sampling units in terms of reducing the expected width. Namely, when doubling the number of clusters the expected width reduces by approximately 75% compared to approximately 50% when doubling the number of sampling units. We also see interval (3.8) and (3.14) have an appropriately slightly wider expected width compared to interval (3.9) and (3.15), respectively, due to the correction term in the MSE approximation.

6 Summary and Future Work

We have developed new prediction interval methodologies for a class of GLMMs that are suitable for analyzing clustered count data. In the BLP and BP based intervals, our approach was to derive an approximation to the MSE of the eBP or the eBLP using the technique applied by Kackar and Harville (1984) for LMMs. In a HPD interval, our approach is to use MCMC methods to sample from the conditional distribution of $\boldsymbol{w} \mid y$. We compared our proposed methods with three existing prediction interval methods that are implemented in the SAS procedure GLIMMIX, which are based upon pseudo-likelihood, Laplace, and quadrature approximations.

For three illustrated examples, our simulation study showed that the coverage probabilities for the intervals computed by SAS are too low. The coverage probabilities for our proposed interval (3.8) and (3.14) with bootstrap adjustments are quite close to the nominal value with an expected width that rapidly decreases as the number of clusters increases, and less rapidly as the number of sampling units within a cluster increases.

For the Poisson GLMM and the Bernoulli GLMM, the HPD interval performs as well as the BP based and the BLP based interval using the bootstrap percentile adjustment in terms of coverage probability. Thus, we construct the expected width from each simulation to compare these three methods in Table10-1 and 10-2 and Table15-1 and 15-2. It is clear to see the expected width from the HPD interval is shorter than the BLP based and the BP based interval using the bootstrap percentile adjustment, thus, the HPD interval outperforms the BLP based and the BP based. Thus, when the HPD interval is available, it has better performance, however, the BLP and the BP based interval are more generalized

When constructing a HPD interval, we must know the form of the density in (4.6) or (4.10), however, this density is usually mathematically intractable, so we then use the random walk Metropolis Hasting algorithm to sample from the target density in (4.6) or (4.10), to construct a HPD interval.

In addition, future work includes considering a standard GLMM having two or more random factors or a conjugate GLMM presented in Appendix A.19. When using this conjugate GLMM, there are two advantages and two disadvantages. The first advantage is that a closed form of the parameter estimate exists as shown in Appendix A.19 while the second advantage is that a closed form approximation of the MSE is derived in Appendix A.19. The first disadvantage is that it lacks flexibility to extend beyond independent random factors while the second disadvantage is that no available software can fit this conjugate GLMM. Due to these two disadvantages, the standard GLMM (our operating GLMM) was considered in this dissertation.

κ	β_0	σ	Prediction Interval									
			(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	1	.2	0.900	0.938	0.885	0.933	0.879	0.901	0.881	0.903	0.882	0.905
		.4	0.902	0.939	0.889	0.934	0.873	0.898	0.875	0.901	0.876	0.902
	2	.2	0.905	0.936	0.901	0.931	0.875	0.897	0.876	0.899	0.876	0.903
		.4	0.899	0.934	0.896	0.928	0.876	0.899	0.876	0.901	0.877	0.904
2	1	.2	0.908	0.938	0.905	0.935	0.869	0.896	0.871	0.898	0.873	0.901
		.4	0.911	0.940	0.907	0.936	0.871	0.899	0.872	0.900	0.872	0.902
	2	.2	0.905	0.941	0.899	0.939	0.868	0.894	0.870	0.896	0.871	0.899
		.4	0.902	0.939	0.895	0.936	0.873	0.897	0.874	0.899	0.875	0.900
5	1	.2	0.898	0.943	0.892	0.938	0.878	0.903	0.879	0.904	0.881	0.905
		.4	0.894	0.941	0.889	0.937	0.875	0.899	0.876	0.903	0.877	0.904
	2	.2	0.896	0.943	0.892	0.939	0.872	0.894	0.873	0.896	0.874	0.897
		.4	0.893	0.942	0.887	0.936	0.876	0.898	0.877	0.901	0.879	0.903

Table 1-1. Negative Binomial Coverage Probabilities for $(m, n) = (10, 5)$. Nominal Coverage is 0.95.

κ	β_0	σ	Correlation of two terms in (3.2)	Prediction Interval				
				(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	(3.14) w/2tr(AB)
1	1	.2	-0.002	0.903	0.937	0.886	0.930	0.912
		.4	0.009	0.904	0.938	0.901	0.936	0.908
	2	.2	0.012	0.902	0.935	0.898	0.929	0.913
		.4	-0.006	0.898	0.932	0.895	0.930	0.901
2	1	.2	0.003	0.906	0.941	0.903	0.934	0.917
		.4	-0.005	0.912	0.943	0.910	0.935	0.920
	2	.2	0.007	0.903	0.940	0.897	0.936	0.909
		.4	-0.001	0.903	0.938	0.899	0.935	0.911
5	1	.2	-0.006	0.900	0.940	0.891	0.937	0.904
		.4	0.014	0.895	0.939	0.890	0.935	0.901
	2	.2	-0.011	0.898	0.944	0.892	0.941	0.903
		.4	0.012	0.891	0.941	0.886	0.938	0.900

Table 1-2. Negative Binomial Coverage Probabilities for $(m, n) = (10, 5)$. Nominal Coverage is 0.95.

κ	β_0	σ	Prediction Interval									
			(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	1	.2	0.909	0.941	0.895	0.936	0.892	0.904	0.894	0.907	0.895	0.909
		.4	0.912	0.942	0.909	0.938	0.890	0.901	0.891	0.903	0.891	0.905
	2	.2	0.914	0.939	0.912	0.935	0.903	0.912	0.904	0.914	0.905	0.916
		.4	0.916	0.943	0.913	0.940	0.906	0.916	0.906	0.917	0.906	0.918
2	1	.2	0.911	0.939	0.908	0.936	0.908	0.915	0.910	0.916	0.911	0.917
		.4	0.915	0.942	0.912	0.940	0.904	0.912	0.906	0.913	0.906	0.914
	2	.2	0.920	0.940	0.916	0.938	0.898	0.908	0.901	0.910	0.902	0.912
		.4	0.917	0.939	0.915	0.935	0.896	0.903	0.897	0.904	0.899	0.906
5	1	.2	0.915	0.943	0.914	0.942	0.899	0.906	0.901	0.910	0.903	0.911
		.4	0.918	0.941	0.916	0.938	0.901	0.909	0.902	0.911	0.904	0.912
	2	.2	0.909	0.941	0.905	0.937	0.902	0.912	0.904	0.913	0.905	0.915
		.4	0.907	0.938	0.903	0.935	0.905	0.913	0.906	0.915	0.907	0.915

Table 2-1. Negative Binomial Coverage Probabilities for $(m, n) = (10, 10)$. Nominal Coverage is 0.95.

κ	β_0	σ	Correlation of two terms in (3.2)	Prediction Interval				
				(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	(3.14) w/2tr(AB)
1	1	.2	-0.010	0.907	0.938	0.901	0.933	0.915
		.4	0.003	0.914	0.943	0.906	0.935	0.918
	2	.2	0.012	0.912	0.941	0.910	0.935	0.919
		.4	0.006	0.915	0.946	0.913	0.943	0.921
2	1	.2	-0.001	0.910	0.937	0.906	0.934	0.918
		.4	-0.003	0.917	0.944	0.914	0.942	0.919
	2	.2	0.009	0.922	0.943	0.917	0.939	0.925
		.4	-0.010	0.915	0.939	0.911	0.932	0.926
5	1	.2	0.007	0.913	0.941	0.912	0.935	0.922
		.4	-0.010	0.920	0.943	0.916	0.937	0.925
	2	.2	-0.006	0.911	0.938	0.908	0.934	0.916
		.4	0.015	0.905	0.937	0.901	0.931	0.918

Table 2-2. Negative Binomial Coverage Probabilities for $(m, n) = (10, 10)$. Nominal Coverage is 0.95.

			Prediction Interval									
κ	β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	1	.2	0.916	0.936	0.912	0.933	0.912	0.918	0.914	0.919	0.915	0.920
		.4	0.917	0.938	0.914	0.935	0.918	0.923	0.920	0.923	0.921	0.924
	2	.2	0.921	0.940	0.919	0.938	0.909	0.915	0.911	0.916	0.912	0.918
		.4	0.923	0.941	0.922	0.940	0.913	0.920	0.913	0.922	0.915	0.924
2	1	.2	0.926	0.939	0.925	0.937	0.914	0.921	0.915	0.923	0.917	0.924
		.4	0.924	0.937	0.923	0.935	0.918	0.923	0.922	0.924	0.924	0.926
	2	.2	0.915	0.942	0.912	0.940	0.920	0.924	0.921	0.925	0.923	0.926
		.4	0.912	0.938	0.908	0.936	0.917	0.919	0.919	0.921	0.921	0.923
5	1	.2	0.925	0.941	0.922	0.939	0.915	0.923	0.917	0.925	0.919	0.926
		.4	0.928	0.944	0.925	0.941	0.911	0.916	0.911	0.916	0.913	0.918
	2	.2	0.919	0.942	0.917	0.940	0.921	0.926	0.922	0.927	0.924	0.928
		.4	0.917	0.940	0.915	0.939	0.918	0.921	0.919	0.923	0.919	0.925

Table 3-1. Negative Binomial Coverage Probabilities for $(m, n) = (20, 5)$. Nominal Coverage is 0.95.

κ	β_0	σ	Correlation of two terms in (3.2)	Prediction Interval (3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	(3.14) w/2tr(AB)
1	1	.2	0.005	0.918	0.933	0.910	0.929	0.922
		.4	0.007	0.915	0.936	0.913	0.933	0.924
	2	.2	-0.013	0.920	0.939	0.918	0.935	0.928
		.4	-0.006	0.924	0.942	0.921	0.937	0.929
2	1	.2	0.005	0.928	0.941	0.924	0.935	0.930
		.4	-0.006	0.925	0.938	0.921	0.933	0.928
	2	.2	0.001	0.913	0.936	0.910	0.934	0.920
		.4	0.006	0.910	0.937	0.908	0.934	0.918
5	1	.2	0.002	0.924	0.939	0.920	0.937	0.930
		.4	-0.016	0.930	0.945	0.927	0.940	0.932
	2	.2	0.012	0.917	0.941	0.913	0.938	0.924
		.4	-0.013	0.918	0.937	0.916	0.936	0.923

Table 3-2. Negative Binomial Coverage Probabilities for $(m, n) = (20, 5)$. Nominal Coverage is 0.95.

			Prediction Interval									
κ	β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	1	.2	0.919	0.944	0.915	0.942	0.921	0.930	0.922	0.931	0.924	0.932
		.4	0.923	0.948	0.918	0.946	0.925	0.931	0.926	0.932	0.926	0.933
	2	.2	0.925	0.943	0.924	0.942	0.918	0.925	0.919	0.926	0.921	0.926
		.4	0.929	0.947	0.927	0.946	0.923	0.928	0.923	0.928	0.924	0.929
2	1	.2	0.930	0.948	0.927	0.945	0.925	0.932	0.926	0.933	0.926	0.934
		.4	0.928	0.945	0.926	0.943	0.924	0.930	0.927	0.931	0.929	0.932
	2	.2	0.926	0.946	0.923	0.944	0.925	0.931	0.926	0.932	0.927	0.933
		.4	0.923	0.945	0.921	0.942	0.923	0.928	0.924	0.930	0.926	0.931
5	1	.2	0.926	0.943	0.925	0.942	0.929	0.933	0.931	0.934	0.932	0.935
		.4	0.929	0.947	0.928	0.946	0.926	0.929	0.927	0.931	0.927	0.931
	2	.2	0.925	0.942	0.924	0.941	0.926	0.930	0.927	0.931	0.929	0.932
		.4	0.927	0.944	0.926	0.943	0.921	0.928	0.923	0.928	0.925	0.929

Table 4-1. Negative Binomial Coverage Probabilities
for $(m, n) = (20, 10)$. Nominal Coverage is 0.95.

κ	β_0	σ	Correlation of two terms in (3.2)	Prediction Interval (3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	(3.14) w/2tr(AB)
1	1	.2	-0.004	0.917	0.944	0.913	0.939	0.924
		.4	-0.012	0.922	0.950	0.916	0.942	0.928
	2	.2	-0.011	0.927	0.948	0.922	0.941	0.929
		.4	-0.009	0.928	0.945	0.924	0.942	0.932
2	1	.2	0.005	0.932	0.952	0.923	0.947	0.935
		.4	-0.010	0.925	0.947	0.922	0.943	0.931
	2	.2	0.004	0.926	0.946	0.922	0.942	0.931
		.4	-0.012	0.924	0.943	0.919	0.942	0.928
5	1	.2	0.007	0.925	0.944	0.921	0.938	0.931
		.4	0.014	0.930	0.948	0.925	0.942	0.934
	2	.2	0.007	0.926	0.944	0.923	0.940	0.929
		.4	0.012	0.926	0.946	0.920	0.939	0.934

Table 4-2. Negative Binomial Coverage Probabilities
for $(m, n) = (20, 10)$. Nominal Coverage is 0.95.

			Expected Width for (m, n)							
			(10 , 5)				(10 , 10)			
κ	β_0	σ	(3.14)w/BS	(3.15)w/BS	(3.8) w/BS	(3.9)w/BS	(3.14) w/BS	(3.15) w/BS	(3.8) w/BS	(3.9)w/BS
1	1	.2	2.348	2.298	2.451	2.357	1.338	1.289	1.432	1.348
		.4	2.576	2.498	2.579	2.489	1.548	1.493	1.486	1.358
	2	.2	2.987	2.893	2.889	2.798	1.632	1.583	1.569	1.426
		.4	3.042	2.983	2.987	2.887	1.738	1.695	1.831	1.730
2	1	.2	2.563	2.478	2.487	2.376	1.284	1.238	1.312	1.231
		.4	2.681	2.512	2.689	2.563	1.292	1.258	1.322	1.245
	2	.2	3.034	2.983	2.987	2.872	1.318	1.263	1.431	1.331
		.4	3.142	3.015	3.241	3.126	1.322	1.301	1.298	1.265
5	1	.2	2.487	2.397	2.512	2.451	1.326	1.298	1.301	1.288
		.4	2.534	2.498	2.612	2.561	1.341	1.328	1.351	1.333
	2	.2	2.834	2.795	2.784	2.684	1.358	1.348	1.372	1.351
		.4	2.931	2.887	2.928	2.842	1.361	1.351	1.340	1.308

Table 5-1. Negative Binomial Expected Widths.
Nominal Coverage is 0.95.

κ	β_0	σ	Expected Width for (m, n)							
			(20 , 5)				(20 , 10)			
			(3.14)w/BS	(3.15)w/BS	(3.8) w/BS	(3.9)w/BS	(3.14) w/BS	(3.15) w/BS	(3.8) w/BS	(3.9)w/BS
1	1	.2	0.787	0.713	0.797	0.783	0.489	0.463	0.512	0.501
		.4	0.812	0.795	0.816	0.802	0.492	0.471	0.503	0.498
	2	.2	0.825	0.806	0.831	0.821	0.503	0.493	0.496	0.483
		.4	0.831	0.828	0.829	0.819	0.509	0.501	0.487	0.479
2	1	.2	0.756	0.748	0.776	0.764	0.499	0.488	0.505	0.493
		.4	0.779	0.763	0.785	0.769	0.512	0.503	0.498	0.482
	2	.2	0.781	0.772	0.790	0.781	0.516	0.510	0.501	0.491
		.4	0.792	0.781	0.798	0.784	0.522	0.518	0.526	0.512
5	1	.2	0.748	0.732	0.738	0.725	0.475	0.468	0.498	0.483
		.4	0.751	0.738	0.755	0.743	0.479	0.470	0.490	0.482
	2	.2	0.762	0.749	0.768	0.751	0.485	0.479	0.476	0.465
		.4	0.768	0.759	0.781	0.779	0.489	0.481	0.494	0.483

Table 5-2. Negative Binomial Expected Widths.
Nominal Coverage is 0.95.

		Prediction Interval									
β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	.2	0.903	0.939	0.898	0.937	0.884	0.902	0.884	0.903	0.885	0.904
	.4	0.901	0.940	0.896	0.938	0.881	0.905	0.883	0.906	0.886	0.907
1.5	.2	0.905	0.937	0.901	0.933	0.892	0.903	0.894	0.904	0.895	0.904
	.4	0.902	0.938	0.898	0.934	0.896	0.906	0.897	0.906	0.899	0.907
2	.2	0.894	0.940	0.889	0.938	0.875	0.894	0.876	0.896	0.878	0.897
	.4	0.896	0.942	0.892	0.939	0.879	0.893	0.881	0.895	0.882	0.895
2.5	.2	0.899	0.941	0.893	0.938	0.887	0.898	0.888	0.900	0.889	0.901
	.4	0.903	0.937	0.898	0.934	0.881	0.899	0.881	0.902	0.883	0.903

Table 6-1. Poisson GLMM Coverage Probabilities
for $(m, n) = (10, 5)$. Nominal Coverage is 0.95.

		Correlation of	Prediction Interval					
β_0	σ	two terms in (3.2)	(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
1	.2	0.004	0.903	0.938	0.895	0.936	0.937	0.911
	.4	0.006	0.901	0.939	0.897	0.935	0.934	0.906
1.5	.2	-0.012	0.905	0.937	0.902	0.935	0.938	0.912
	.4	0.008	0.908	0.939	0.905	0.938	0.940	0.908
2	.2	-0.009	0.895	0.941	0.891	0.937	0.941	0.897
	.4	0.005	0.896	0.943	0.890	0.938	0.939	0.900
2.5	.2	0.012	0.901	0.940	0.899	0.939	0.942	0.902
	.4	-0.008	0.894	0.938	0.891	0.936	0.939	0.908

Table 6-2. Poisson GLMM Coverage Probabilities
for $(m, n) = (10, 5)$. Nominal Coverage is 0.95.

		Prediction Interval									
β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	.2	0.916	0.940	0.916	0.936	0.899	0.911	0.901	0.912	0.903	0.913
	.4	0.920	0.941	0.917	0.940	0.894	0.908	0.898	0.910	0.899	0.912
1.5	.2	0.915	0.938	0.913	0.935	0.902	0.914	0.903	0.915	0.904	0.917
	.4	0.923	0.939	0.920	0.937	0.903	0.916	0.905	0.917	0.906	0.918
2	.2	0.922	0.939	0.918	0.936	0.890	0.906	0.892	0.909	0.892	0.911
	.4	0.918	0.936	0.915	0.932	0.904	0.915	0.905	0.917	0.905	0.918
2.5	.2	0.921	0.940	0.919	0.938	0.901	0.912	0.902	0.913	0.903	0.914
	.4	0.924	0.938	0.921	0.936	0.905	0.910	0.906	0.910	0.906	0.912

Table 7-1. Poisson GLMM Coverage Probabilities
for $(m, n) = (10, 10)$. Nominal Coverage is 0.95.

		Correlation of two terms in (3.2)	Prediction Interval					
β_0	σ		(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
1	.2	0.012	0.916	0.942	0.914	0.940	0.941	0.920
	.4	-0.014	0.921	0.940	0.916	0.938	0.937	0.928
1.5	.2	0.005	0.915	0.939	0.913	0.936	0.942	0.923
	.4	-0.009	0.924	0.941	0.920	0.938	0.940	0.931
2	.2	-0.013	0.922	0.942	0.916	0.938	0.941	0.926
	.4	-0.006	0.915	0.939	0.912	0.937	0.939	0.925
2.5	.2	0.007	0.918	0.944	0.915	0.941	0.942	0.928
	.4	0.015	0.920	0.943	0.918	0.939	0.936	0.930

Table 7-2. Poisson GLMM Coverage Probabilities
for $(m, n) = (10, 10)$. Nominal Coverage is 0.95.

		Prediction Interval									
β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	.2	0.921	0.941	0.918	0.939	0.915	0.927	0.917	0.929	0.918	0.930
	.4	0.925	0.936	0.923	0.935	0.918	0.923	0.919	0.925	0.919	0.927
1.5	.2	0.927	0.942	0.924	0.940	0.920	0.928	0.921	0.930	0.923	0.930
	.4	0.922	0.941	0.918	0.939	0.925	0.931	0.925	0.932	0.927	0.933
2	.2	0.920	0.943	0.919	0.941	0.911	0.922	0.912	0.924	0.912	0.924
	.4	0.924	0.944	0.921	0.941	0.921	0.924	0.923	0.925	0.924	0.926
2.5	.2	0.930	0.946	0.926	0.943	0.916	0.925	0.918	0.926	0.919	0.927
	.4	0.931	0.945	0.927	0.942	0.927	0.932	0.928	0.934	0.929	0.935

Table 8-1. Poisson GLMM Coverage Probabilities
for $(m, n) = (20, 5)$. Nominal Coverage is 0.95.

		Correlation of two terms in (3.2)	Prediction Interval					
β_0	σ		(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
1	.2	0.003	0.920	0.939	0.918	0.934	0.937	0.923
	.4	-0.002	0.924	0.946	0.921	0.941	0.942	0.929
1.5	.2	-0.005	0.921	0.945	0.918	0.943	0.944	0.935
	.4	-0.012	0.925	0.947	0.921	0.945	0.943	0.930
2	.2	0.016	0.923	0.944	0.920	0.940	0.941	0.931
	.4	-0.014	0.928	0.945	0.926	0.941	0.940	0.932
2.5	.2	0.005	0.924	0.942	0.922	0.940	0.942	0.936
	.4	-0.008	0.927	0.940	0.925	0.938	0.943	0.935

Table 8-2. Poisson GLMM Coverage Probabilities
for $(m, n) = (20, 5)$. Nominal Coverage is 0.95.

		Prediction Interval									
β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
1	.2	0.928	0.946	0.926	0.943	0.922	0.927	0.922	0.927	0.923	0.928
	.4	0.929	0.945	0.927	0.941	0.918	0.932	0.919	0.932	0.920	0.933
1.5	.2	0.925	0.946	0.921	0.942	0.925	0.930	0.926	0.931	0.926	0.931
	.4	0.921	0.943	0.918	0.939	0.927	0.932	0.928	0.933	0.929	0.935
2	.2	0.923	0.941	0.920	0.940	0.921	0.928	0.921	0.925	0.922	0.927
	.4	0.928	0.947	0.923	0.944	0.926	0.931	0.926	0.930	0.927	0.931
2.5	.2	0.930	0.948	0.926	0.945	0.929	0.934	0.930	0.935	0.931	0.936
	.4	0.929	0.945	0.925	0.941	0.931	0.936	0.931	0.936	0.932	0.936

Table 9-1. Poisson GLMM Coverage Probabilities
for $(m, n) = (20, 10)$. Nominal Coverage is 0.95.

		Correlation of	Prediction Interval					
β_0	σ	two terms in (3.2)	(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
1	.2	0.009	0.927	0.944	0.925	0.941	0.944	0.936
	.4	0.003	0.921	0.945	0.919	0.942	0.940	0.936
1.5	.2	-0.008	0.930	0.947	0.925	0.945	0.942	0.933
	.4	0.012	0.934	0.948	0.932	0.947	0.946	0.932
2	.2	-0.014	0.928	0.946	0.925	0.944	0.949	0.931
	.4	0.009	0.924	0.945	0.922	0.943	0.947	0.934
2.5	.2	-0.013	0.931	0.942	0.929	0.939	0.943	0.934
	.4	-0.005	0.929	0.948	0.926	0.945	0.947	0.935

Table 9-2. Poisson GLMM Coverage Probabilities
for $(m, n) = (20, 10)$. Nominal Coverage is 0.95.

β_0	σ	Expected Width for (m, n)									
		(10 , 5)					(10 , 10)				
		(3.14) w/BS	(3.15) w/BS	(3.8) w/BS	(3.9) w/BS	HPD	(3.14) w/BS	(3.15) w/BS	(3.8) w/BS	(3.9) w/BS	HPD
1	.2	5.124	4.987	5.234	5.019	4.897	2.348	2.257	2.579	2.462	2.198
	.4	5.579	5.321	5.679	5.458	5.238	2.357	2.145	2.487	2.378	2.013
1.5	.2	4.987	4.789	5.095	4.874	4.679	2.451	2.235	2.378	2.264	1.984
	.4	5.011	4.982	4.987	4.872	4.563	2.443	2.138	2.375	2263	1.982
2	.2	5.523	5.314	5.231	5.014	4.872	2.378	2.224	2.413	2.301	2.013
	.4	4.867	4.642	4.763	4.578	4.231	2.298	2.109	2.324	2.287	2.024
2.5	.2	5.021	4.809	4.984	4.783	4.349	2.543	2.238	2.489	2.384	2.231
	.4	5.145	4.908	5.012	4.983	4.592	2.621	2.348	2.587	2.485	2.140

Table 10-1. Poisson GLMM Expected Widths.
Nominal Coverage is 0.95.

β_0	σ	Expected Width for (m, n)									
		(20 , 5)					(20 , 10)				
		(3.14) w/BS	(3.15) w/BS	(3.8) w/BS	(3.9) w/BS	HPD	(3.14) w/BS	(3.15) w/BS	(3.8) w/BS	(3.9) w/BS	HPD
1	.2	1.267	1.187	1.256	1.187	1.004	0.684	0.662	0.691	0.673	0.631
	.4	1.253	1.165	1.248	1.128	1.104	0.692	0.671	0.683	0.663	0.623
1.5	.2	1.225	1.159	1.238	1.117	0.983	0.702	0.673	0.711	0.689	0.642
	.4	1.119	0.998	1.226	1.120	0.981	0.713	0.687	0.721	0.702	0.639
2	.2	1.112	0.992	1.229	1.112	0.984	0.698	0.663	0.683	0.669	0.644
	.4	1.243	1.082	1.183	0.988	0.975	0.732	0.701	0.714	0.693	0.648
2.5	.2	1.273	1.104	1.225	1.009	0.988	0.692	0.671	0.702	0.682	0.652
	.4	1.281	1.107	1.245	1.012	0.993	0.721	0.702	0.719	0.699	0.667

Table 10-2. Poisson GLMM Expected Widths.
Nominal Coverage is 0.95.

β_0	σ	Prediction Interval									
		(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
-2.5	0.2	0.895	0.935	0.892	0.934	0.869	0.878	0.870	0.879	0.871	0.879
	2	0.899	0.942	0.896	0.938	0.871	0.880	0.873	0.881	0.874	0.881
-0.5	0.2	0.908	0.941	0.905	0.939	0.875	0.883	0.876	0.884	0.877	0.886
	2	0.903	0.939	0.901	0.936	0.873	0.881	0.875	0.882	0.876	0.884
0.5	0.2	0.894	0.934	0.891	0.931	0.875	0.884	0.876	0.885	0.878	0.886
	2	0.897	0.935	0.893	0.932	0.878	0.889	0.878	0.891	0.879	0.892
2.0	0.2	0.905	0.940	0.902	0.938	0.874	0.885	0.876	0.885	0.877	0.886
	2	0.902	0.938	0.897	0.935	0.876	0.888	0.874	0.889	0.875	0.891

Table 11-1. Bernoulli GLMM Coverage Probabilities
for $(m, n) = (10, 5)$. Nominal Coverage is 0.95.

β_0	σ	Correlation of two terms in (3.2)	Prediction Interval					
			(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
-2.5	0.2	0.004	0.893	0.934	0.890	0.931	0.935	0.899
	2	0.008	0.896	0.940	0.894	0.936	0.937	0.902
-0.5	0.2	0.012	0.911	0.942	0.907	0.937	0.940	0.912
	2	-0.009	0.905	0.938	0.902	0.933	0.934	0.909
0.5	0.2	0.014	0.895	0.936	0.890	0.931	0.933	0.899
	2	0.006	0.897	0.934	0.893	0.929	0.934	0.902
2.0	0.2	-0.005	0.907	0.942	0.900	0.937	0.936	0.908
	2	0.013	0.903	0.940	0.898	0.937	0.937	0.906

Table 11-2. Bernoulli GLMM Coverage Probabilities
for $(m, n) = (10, 5)$. Nominal Coverage is 0.95.

		Prediction Interval									
β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
-2.5	0.2	0.911	0.942	0.907	0.938	0.903	0.908	0.906	0.910	0.907	0.911
	2	0.905	0.940	0.902	0.936	0.897	0.906	0.897	0.906	0.900	0.906
-0.5	0.2	0.915	0.944	0.911	0.942	0.899	0.906	0.901	0.907	0.903	0.908
	2	0.911	0.941	0.907	0.939	0.895	0.902	0.897	0.903	0.898	0.904
0.5	0.2	0.918	0.935	0.915	0.931	0.892	0.901	0.893	0.902	0.895	0.903
	2	0.920	0.938	0.916	0.935	0.896	0.904	0.896	0.905	0.898	0.906
2.0	0.2	0.917	0.943	0.914	0.938	0.901	0.906	0.904	0.907	0.906	0.908
	2	0.912	0.941	0.908	0.939	0.898	0.904	0.901	0.905	0.903	0.906

Table 12-1. Bernoulli GLMM Coverage Probabilities
for $(m, n) = (10, 10)$. Nominal Coverage is 0.95.

		Correlation of two terms in (3.2)	Prediction Interval					
β_0	σ		(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
-2.5	0.2	0.012	0.908	0.944	0.904	0.939	0.940	0.915
	2	-0.015	0.906	0.942	0.902	0.934	0.935	0.909
-0.5	0.2	0.007	0.913	0.945	0.909	0.939	0.943	0.918
	2	0.013	0.910	0.943	0.905	0.940	0.941	0.915
0.5	0.2	0.005	0.915	0.937	0.911	0.932	0.935	0.921
	2	-0.002	0.922	0.939	0.916	0.936	0.933	0.923
2.0	0.2	0.002	0.916	0.942	0.913	0.939	0.936	0.920
	2	0.008	0.913	0.939	0.910	0.938	0.938	0.918

Table 12-2. Bernoulli GLMM Coverage Probabilities
for $(m, n) = (10, 10)$. Nominal Coverage is 0.95.

β_0	σ	Prediction Interval									
		(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
-2.5	0.2	0.925	0.944	0.921	0.939	0.921	0.926	0.922	0.927	0.923	0.928
	2	0.921	0.941	0.916	0.939	0.919	0.924	0.920	0.925	0.921	0.926
-0.5	0.2	0.919	0.938	0.915	0.936	0.912	0.920	0.914	0.921	0.916	0.922
	2	0.923	0.941	0.918	0.940	0.914	0.922	0.915	0.923	0.918	0.925
0.5	0.2	0.924	0.942	0.919	0.940	0.922	0.928	0.922	0.930	0.924	0.931
	2	0.928	0.945	0.923	0.941	0.923	0.930	0.923	0.931	0.925	0.932
2.0	0.2	0.922	0.942	0.919	0.940	0.915	0.924	0.917	0.925	0.919	0.926
	2	0.918	0.944	0.914	0.941	0.911	0.917	0.914	0.918	0.916	0.920

Table 13-1. Bernoulli GLMM Coverage Probabilities for $(m, n) = (20, 5)$. Nominal Coverage is 0.95.

β_0	σ	Correlation of two terms in (3.2)	Prediction Interval					
			(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
-2.5	0.2	0.004	0.926	0.944	0.922	0.938	0.937	0.929
	2	0.007	0.922	0.942	0.917	0.936	0.940	0.926
-0.5	0.2	-0.013	0.917	0.936	0.912	0.933	0.939	0.925
	2	-0.008	0.921	0.943	0.918	0.939	0.941	0.928
0.5	0.2	0.007	0.926	0.943	0.921	0.940	0.942	0.929
	2	0.009	0.927	0.946	0.920	0.942	0.943	0.932
2.0	0.2	0.005	0.920	0.941	0.918	0.938	0.939	0.926
	2	-0.008	0.917	0.942	0.915	0.937	0.943	0.924

Table 13-2. Bernoulli GLMM Coverage Probabilities for $(m, n) = (20, 5)$. Nominal Coverage is 0.95.

		Prediction Interval									
β_0	σ	(3.14)	(3.14) w/BS	(3.15)	(3.15) w/BS	SAS/PL	SAS/PL w/BS	SAS/L	SAS/L w/BS	SAS/Q	SAS/Q w/BS
-2.5	0.2	0.929	0.947	0.928	0.946	0.925	0.931	0.926	0.931	0.927	0.931
	2	0.928	0.945	0.925	0.943	0.923	0.928	0.925	0.928	0.926	0.929
-0.5	0.2	0.933	0.946	0.929	0.945	0.919	0.925	0.921	0.926	0.922	0.927
	2	0.931	0.945	0.927	0.941	0.918	0.923	0.921	0.925	0.922	0.926
0.5	0.2	0.924	0.945	0.921	0.942	0.922	0.926	0.923	0.927	0.924	0.928
	2	0.929	0.948	0.926	0.944	0.924	0.930	0.924	0.931	0.925	0.932
2.0	0.2	0.928	0.946	0.925	0.943	0.925	0.932	0.926	0.933	0.927	0.934
	2	0.931	0.948	0.926	0.944	0.927	0.933	0.928	0.935	0.929	0.937

Table 14-1. Bernoulli GLMM Coverage Probabilities
for $(m, n) = (20, 10)$. Nominal Coverage is 0.95.

		Correlation of two terms in (3.2)	Prediction Interval					
β_0	σ		(3.8)	(3.8) w/BS	(3.9)	(3.9) w/BS	HPD	(3.14) w/2tr(AB)
-2.5	0.2	0.015	0.928	0.944	0.924	0.942	0.945	0.932
	2	0.006	0.926	0.945	0.922	0.942	0.944	0.931
-0.5	0.2	0.009	0.934	0.948	0.931	0.944	0.946	0.935
	2	0.010	0.929	0.943	0.925	0.939	0.943	0.936
0.5	0.2	-0.013	0.923	0.944	0.919	0.941	0.944	0.930
	2	0.011	0.931	0.949	0.927	0.946	0.946	0.935
2.0	0.2	-0.004	0.927	0.943	0.922	0.940	0.945	0.935
	2	0.003	0.932	0.946	0.925	0.941	0.942	0.938

Table 14-2. Bernoulli GLMM Coverage Probabilities
for $(m, n) = (20, 10)$. Nominal Coverage is 0.95.

		Expected Width for (m, n)									
		(10 , 5)					(10 , 10)				
		(3.14)	(3.15)	(3.8)	(3.9)	HPD	(3.14)	(3.15)	(3.8)	(3.9)	HPD
β_0	σ	w/BS	w/BS	w/BS	w/BS		w/BS	w/BS	w/BS	w/BS	
-2.5	0.2	0.883	0.879	0.886	0.882	0.843	0.534	0.528	0.527	0.522	0.498
	2	0.891	0.889	0.893	0.889	0.851	0.541	0.536	0.491	0.486	0.473
-0.5	0.2	0.873	0.869	0.869	0.865	0.846	0.487	0.479	0.502	0.496	0.468
	2	0.882	0.878	0.884	0.876	0.862	0.502	0.496	0.493	0.488	0.479
0.5	0.2	0.871	0.867	0.873	0.869	0.853	0.528	0.523	0.521	0.516	0.502
	2	0.874	0.871	0.876	0.871	0.861	0.538	0.534	0.518	0.512	0.504
2.0	0.2	0.869	0.864	0.872	0.868	0.849	0.545	0.537	0.526	0.522	0.523
	2	0.873	0.868	0.876	0.872	0.848	0.551	0.547	0.522	0.515	0.508

Table 15-1. Bernoulli GLMM Expected Widths.
Nominal Coverage is 0.95.

		Expected Width for (m, n)									
		(20 , 5)					(20 , 10)				
		(3.14)	(3.15)	(3.8)	(3.9)	HPD	(3.14)	(3.15)	(3.8)	(3.9)	HPD
β_0	σ	w/BS	w/BS	w/BS	w/BS		w/BS	w/BS	w/BS	w/BS	
-2.5	0.2	0.223	0.218	0.254	0.248	0.225	0.105	0.098	0.108	0.105	0.087
	2	0.231	0.228	0.248	0.241	0.218	0.125	0.115	0.126	0.124	0.107
-0.5	0.2	0.219	0.217	0.212	0.207	0.209	0.114	0.109	0.117	0.114	0.091
	2	0.228	0.224	0.218	0.214	0.212	0.118	0.112	0.107	0.103	0.089
0.5	0.2	0.226	0.223	0.231	0.226	0.218	0.104	0.099	0.109	0.105	0.092
	2	0.231	0.227	0.226	0.221	0.215	0.115	0.105	0.118	0.112	0.098
2.0	0.2	0.218	0.215	0.222	0.218	0.210	0.104	0.098	0.108	0.103	0.093
	2	0.231	0.226	0.216	0.208	0.198	0.129	0.117	0.123	0.118	0.105

Table 15-2. Bernoulli GLMM Expected Widths.
Nominal Coverage is 0.95.

Appendix

A.1

We derive the first term, $\left\{\mathrm{E}\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\right\}\right\}^2$, in (3.4) as follows:

$$\begin{aligned}
&\left\{E\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\right\}\right\}^2=\left\{E\left\{g'(g^{-1}(w))\left[\mu_{g^{-1}(w)}+\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}(y-\mu_y)-g^{-1}(w)\right]\right\}\right\}^2\\
&=\left\{E\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)g'(g^{-1}(w))+g'(g^{-1}(w))\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}y-g'(g^{-1}(w))g^{-1}(w)\right\}\right\}^2\\
&=\left\{E\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)g'(g^{-1}(w))\right\}+E\left\{g'(g^{-1}(w))\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}y\right\}-E\left\{g'(g^{-1}(w))g^{-1}(w)\right\}\right\}^2\\
&=\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)E\left[g'(g^{-1}(w))\right]+EE\left\{g'(g^{-1}(w))\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}y\middle|s\right\}-E\left\{g'(g^{-1}(w))g^{-1}(w)\right\}\right\}^2\\
&=\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)E\left[g'(g^{-1}(w))\right]+E\left\{g'(g^{-1}(w))\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\left[\mathrm{E}\left(y|s\right)\right]\right\}-E\left\{g'(g^{-1}(w))g^{-1}(w)\right\}\right\}^2\\
&=\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)E\left[g'(g^{-1}(w))\right]+\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}E\left\{g'(g^{-1}(w))\left[\mathrm{E}\left(y|s\right)\right]\right\}-E\left\{g'(g^{-1}(w))g^{-1}(w)\right\}\right\}^2\\
&=\left\{\boldsymbol{L}_2^T E\left(\boldsymbol{T}\right)\right\}^2
\end{aligned}$$

Where $\boldsymbol{L}_2=\left(\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)\quad\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)\quad -1\right)^T$ and

$\boldsymbol{T}=\left(g'(g^{-1}(w))\quad g'(g^{-1}(w))\left[E\left(y|s\right)\right]\quad g'(g^{-1}(w))g^{-1}(w)\right)'$

We can derive the second term, $\mathrm{Var}\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\right\}$, in (3.4) as follows:

$$\begin{aligned}
\mathrm{Var}\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\right\}&=\mathrm{Var}\left\{E\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\middle|s\right\}\right\}\\
&\quad+E\left\{\mathrm{Var}\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\middle|s\right\}\right\}
\end{aligned}\tag{A.1}$$

Here, we need to derive the first term, $\mathrm{Var}\left\{E\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\middle|s\right\}\right\}$, in (A.1) as follows:

$$
\begin{aligned}
&\mathrm{Var}\left\{E\left\{g'(g^{-1}(w))\left[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)\right]\middle|s\right\}\right\}\\
&=\mathrm{Var}\left\{E\left\{g'(g^{-1}(w))\left[\mu_{g^{-1}(w)}+\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}(y-\mu_y)-g^{-1}(w)\right]\middle|s\right\}\right\}\\
&=\mathrm{Var}\left\{E\left\{g'(g^{-1}(w))\left[\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)+\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}y-g^{-1}(w)\right]\middle|s\right\}\right\}\\
&=\mathrm{Var}\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)g'(g^{-1}(w))+\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}g'(g^{-1}(w))E\left\{y\mid s\right\}-g'(g^{-1}(w))g^{-1}(w)\right\}\\
&=\mathrm{Var}\left\{\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)g'(g^{-1}(w))\right\}+\mathrm{Var}\left\{\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}g'(g^{-1}(w))E\left(y\mid s\right)\right\}\\
&\quad+\mathrm{Var}\left\{g'(g^{-1}(w))g^{-1}(w)\right\}\\
&\quad+2\mathrm{Cov}\left(\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)g'(g^{-1}(w)),\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}g'(g^{-1}(w))E\left(y\mid s_i\right)\right)\\
&\quad-2\mathrm{Cov}\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y g'(g^{-1}(w)),g'(g^{-1}(w))g^{-1}(w)\right)\\
&\quad-2\mathrm{Cov}\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}g'(g^{-1}(w))E\left(y\mid s\right),g'(g^{-1}(w))g^{-1}(w)\right)\\
&=\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)^2\mathrm{Var}\left\{g'(g^{-1}(w))\right\}+\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)\mathrm{Var}\left\{g'(g^{-1}(w))E\left(y\mid s\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)'\\
&\quad+\mathrm{Var}\left\{g'(g^{-1}(w))g^{-1}(w)\right\}\\
&\quad+2\left(\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\right)\mathrm{Cov}\left(g'(g^{-1}(w)),g'(g^{-1}(w))E\left(y\mid s\right)\right)\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)'\\
&\quad-2\mu_{g^{-1}(w)}\text{-}\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mu_y\mathrm{Cov}\left(g'(g^{-1}(w)),g'(g^{-1}(w))g^{-1}(w)\right)\\
&\quad-2\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\mathrm{Cov}\left(g'(g^{-1}(w))E\left(y\mid s\right),g'(g^{-1}(w))g^{-1}(w)\right)\\
&=\boldsymbol{L}_2^T\boldsymbol{H}_2\boldsymbol{L}_2
\end{aligned}
$$

(A.2)

And we need to derive the second term, $E\left\{\mathrm{Var}\left\{g'(g^{-1}(w))[\mathrm{BLP}(g^{-1}(w))-g^{-1}(w)]\middle|s\right\}\right\}$,

in (A.1) as follows:

$$
\begin{aligned}
& E\left\{\mathrm{Var}\left\{g'(g^{-1}(w))\left[\mathrm{BLP}(g^{-1}(w)) - g^{-1}(w)\right]\middle| s\right\}\right\} \\
&= E\left\{\mathrm{Var}\left\{g'(g^{-1}(w))\left[\mu_{g^{-1}(w)} + \mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}(y-\mu_y) - g^{-1}(w)\right]\middle| s\right\}\right\} \\
&= E\left\{\mathrm{Var}\left\{g'(g^{-1}(w))\left[\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}y\right]\middle| s\right\}\right\} \\
&= E\left\{\mathrm{Var}\left\{g'(g^{-1}(w))\,\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\left(y\right)\middle| s\right\}\right\} \\
&= E\left\{\left(g'(g^{-1}(w))\right)^2\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)\mathrm{Var}\left(y\middle| s\right)\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)'\right\} \\
&= \left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)\mathrm{E}\left\{\left(g'(g^{-1}(w))\right)^2\mathrm{Var}\left(y\middle| s\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y}\mathrm{V}_{y,y}^{-1}\right)'
\end{aligned}
\qquad \text{(A.3).}
$$

A.2

Using arguments that parallel Kackar and Harville's (1984), the second term, $M_2(\theta)$, in (3.3) can be approximated by using the second order Taylor expansion as follows:

$$
\begin{aligned}
& \left\{\eta(y;\theta) - \eta(y;\hat{\theta})\right\}^2 \\
&\approx \left\{\eta(y;\theta) - \eta(y;\hat{\theta})\right\}^2\Big|_{\hat{\theta}=\theta} + \left.\frac{2\left\{\eta(y;\theta) - \eta(y;\hat{\theta})\right\}d(y;\theta)'}{1!}\right|_{\hat{\theta}=\theta}\left(\hat{\theta}-\theta\right) \\
&\quad + \frac{2}{2!}\left(\hat{\theta}-\theta\right)^T\left\{d(y;\theta)d(y;\theta)' + \left[\eta(y;\theta) - \eta(y;\hat{\theta})\right][\eta''(y;\theta)]\right\}\Big|_{\hat{\theta}=\theta}\left(\hat{\theta}-\theta\right) \qquad \text{(A.4)} \\
&= \left(\hat{\theta}-\theta\right)' d(y;\theta)d(y;\theta)'\left(\hat{\theta}-\theta\right) \\
&= d(y;\theta)'\left(\hat{\theta}-\theta\right)d(y;\theta)'\left(\hat{\theta}-\theta\right) = \left\{d(y;\theta)'\left(\hat{\theta}-\theta\right)\right\}^2
\end{aligned}
$$

where $\eta''(y;\theta) = [\partial^2\eta(y;\theta)/\partial\theta_i\partial\theta_j]$.

Thus,

$$
\begin{aligned}
M_2(\theta) &= E\left\{\eta(y;\theta)-\eta(y;\hat{\theta})\right\}^2 \\
&\approx E\left\{d(y;\theta)'\left(\hat{\theta}-\theta\right)\right\}^2 \\
&= E\left\{\left[d(y;\theta)'\left(\hat{\theta}-\theta\right)\right]\left[\left(\hat{\theta}-\theta\right)' d(y;\theta)\right]\right\} \\
&= E\left\{tr\left[d(y;\theta)'\left(\hat{\theta}-\theta\right)\right]\left[\left(\hat{\theta}-\theta\right)' d(y;\theta)\right]\right\} \qquad \text{(A.5)} \\
&= E\left\{tr\left\{\left[d(y;\theta)d(y;\theta)'\right]\left[\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)'\right]\right\}\right\} \\
&= tr\left\{E\left\{\left[d(y;\theta)d(y;\theta)'\right]\left[\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)'\right]\right\}\right\}
\end{aligned}
$$

Assume $\mathrm{Cov}\left[d_i(y;\theta)d_j(y;\theta),\left(\hat{\theta}_i-\theta_i\right)\left(\hat{\theta}_j-\theta_j\right)\right]=0$ where $d_i(y;\theta)$ is the i th element of $d(y;\theta)$ and $\hat{\theta}_i-\theta_i$ is the i th element of $\hat{\theta}-\theta$. Thus, $M_2(\theta)$ can be approximated as follows:

$$
\begin{aligned}
M_2(\theta) &= E\left\{\eta(y;\theta)-\eta(y;\hat{\theta})\right\}^2 \\
&\approx tr\left\{E\left\{\left[d(y;\theta)d(y;\theta)'\right]\left[\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)'\right]\right\}\right\} \qquad \text{(A.6)} \\
&\approx tr\left[E\left(d(y;\theta)d(y;\theta)'\right)\mathrm{E}\left(\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)'\right)\right]
\end{aligned}
$$

And we use the observed information matrix to approximate $\mathrm{E}\left(\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)'\right)$. Thus, $M_2(\theta)$ in (A.6) can be approximated by $tr\left[A(\theta)\,B(\theta)\right]$.

A.3

If $\eta(y;\theta)$ is the BP and $\eta(y;\hat{\theta})$ is the eBP, we can show

$E\left\{[w-\eta(y;\theta)]\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]\right\}=0$ as follows:

$$
\begin{aligned}
& E\left\{E\left\{[w-\eta(y;\theta)]\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right] \mid y\right\}\right\} \\
& =E\left\{\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]E\left\{[w-\eta(y;\theta)] \mid y\right\}\right\} \\
& =E\left\{\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]E\left\{[w-E(w \mid y) \mid y\right\}\right\} \\
& =E\left\{\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right][E(w \mid y)-E(w \mid y)]\right\} \\
& =0
\end{aligned}
$$

Then we can prove $w-\eta(y;\theta)$ and $\eta(y;\theta)-\eta(y;\hat{\theta})$ are uncorrelated. Thus, the BP has one good property as follows:

$$
\begin{aligned}
E(e^2) &= E\left[\eta(y;\theta)-w\right]^2 + E\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]^2 \\
&= M_1(\theta) + M_2(\theta)
\end{aligned}
$$

where $M_1(\theta)=E\left[\eta(y;\theta)-w\right]^2$ and $M_2(\theta)=E\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]^2$.

If $\eta(y;\theta)$ is the BLP and $\eta(y;\hat{\theta})$ is the eBLP, we can't show

$E\left\{[w-\eta(y;\theta)]\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]\right\}=0$ as follows:

$$
\begin{aligned}
& E\left\{E\left\{[w-\eta(y;\theta)]\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right] \mid y\right\}\right\} \\
& =E\left\{\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]E\left\{[w-\eta(y;\theta)] \mid y\right\}\right\} \\
& =E\left\{\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]E\left\{[w-[\mu_y+V_{wy}V_{yy}^{-1}(y-\mu_y)]] \mid y\right\}\right\} \qquad \text{(A.7)} \\
& =E\left\{\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right][E(w \mid y)-[\mu_y+V_{wy}V_{yy}^{-1}(y-\mu_y)]]\right\} \\
& =\text{Unknown}
\end{aligned}
$$

In (A.7), in the case we have the normal theorem, then

$E\left\{[w-\eta(y;\theta)]\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]\right\}=0$. The normal theorem means the joint distribution of (w,y) is normally distributed.

Because we can't prove $w-\eta(y;\theta)$ and $\eta(y;\theta)-\eta(y;\hat{\theta})$ are uncorrelated, we can only approximate $E(e^2)$ by assuming $w-\eta(y;\theta)$ and $\eta(y;\theta)-\eta(y;\hat{\theta})$ are uncorrelated. Thus, we have the following result:

$$\begin{aligned} E(e^2) &\approx E\left[\eta(y;\theta)-w\right]^2 + E\left[\eta(y;\theta)-\eta(y;\hat{\theta})\right]^2 \\ &= M_1(\theta) + M_2(\theta) \end{aligned}$$

A.4

$$\begin{aligned} M_1(\theta) &= E(w-\eta(y;\theta))^2 \\ &= E[E(w-\eta(y;\theta))^2 \mid y] \\ &= E[\mathrm{Var}(w \mid y)] \\ &= \delta' E\left[Var(s \mid y;\theta)\right]\delta \end{aligned} \tag{A.8}$$

A.5

Here we summarize all the formulas needed to compute the prediction interval in (4), including those formulas that are needed in conjunction with use of Algorithm 1. All of the integrals required for these computations are one-dimensional integrals that can be easily evaluated using Gaussian quadrature. The best predictor $\eta(y;\theta)$ requires the quantities

$$E(s_i \mid y_i;\theta) = \int_{-\infty}^{\infty} s_i f(s_i \mid y_i\,;\theta) ds_i \tag{A.9}$$

A standard optimization routine can be used to maximize the integrated likelihood $L(\theta \mid y)$ to find $\hat{\theta}$, which in turn can be used to find $\eta(y;\hat{\theta})$.

Algorithm 1 can be used three times, as detailed below, to obtain $\dot{M}(\hat{\theta})$. For the first use of Algorithm 1, send it the matrix $Var(s \mid y) = Diag\left[Var(s_i \mid y_i;\theta) \right]_{i=1}^{m}$, using

$$Var(s_i \mid y_i;\theta) = \int_{-\infty}^{\infty} s_i^2 f(s_i \mid y_i;\theta) ds_i - \left(\int_{-\infty}^{\infty} s_i f(s_i \mid y_i;\theta) ds_i \right)^2 \qquad \text{(A.10)}$$

For the second use, send it the matrix $d(y;\theta)d(y;\theta)' = \left[\frac{\partial \eta(y;\theta)}{\partial \theta} \right] \left[\frac{\partial \eta(y;\theta)}{\partial \theta} \right]'$, where $\partial \eta(y;\theta)/\partial \theta_l = \lambda' \left(\partial \beta / \partial \theta_l \right) + \delta' \left(\partial E(s \mid y;\theta)/\partial \theta_l \right)$. Evaluations of $\partial \beta / \partial \theta_l$ are either zero or one and

$$\begin{aligned} \frac{\partial E(s_i \mid y_i;\theta)}{\partial \theta_l} &= \int_{-\infty}^{\infty} s_i \left(\partial f(s_i \mid y_i;\theta)/\partial \theta_l \right) ds_i \\ &= \frac{ f(y_i;\theta) \int_{-\infty}^{\infty} s_i \frac{\partial f(y_i, s_i;\theta)}{\partial \theta_l} ds_i - \frac{\partial f(y_i;\theta)}{\partial \theta_l} \int_{-\infty}^{\infty} s_i f(y_i, s_i;\theta) ds_i }{ f^2(y_i;\theta) } \end{aligned} \qquad \text{(A.11)}$$

Expressions needed to evaluate (A.11) are

$$\partial f(y_i, s_i;\theta)/\partial \beta_l = \left(\sum_{j=1}^{n_i} x_{ijl} \frac{\partial \log f(y_{ij} \mid \mu_{ij}, \kappa)/\partial \mu_{ij}}{g'(\mu_{ij})} \right) f(y_i, s_i;\theta) \qquad \text{(A.12)}$$

$$\partial f(y_i, s_i;\theta)/\partial \sigma^2 = \frac{1}{2\sigma^2} \left(\frac{s_i^2}{\sigma^2} - 1 \right) f(y_i, s_i;\theta) \qquad \text{(A.13)}$$

$$\partial f(y_i, s_i;\theta)/\partial \kappa = \left(\sum_{j=1}^{n_i} \partial \log f(y_{ij} \mid \mu_{ij}, \kappa)/\partial \kappa \right) f(y_i, s_i;\theta) \qquad \text{(A.14)}$$

and the integrated forms of (A.12)-(A.14)

$$\partial f(y_i;\theta)/\partial \beta_l = \int_{-\infty}^{\infty} \left(\partial f(y_i, s_i;\theta)/\partial \beta_l \right) ds_i \qquad \text{(A.15)}$$

$$\partial f(y_i;\theta)/\partial \sigma^2 = \int_{-\infty}^{\infty} \left(\partial f(y_i, s_i;\theta)/\partial \sigma^2 \right) ds_i \qquad \text{(A.16)}$$

$$\partial f(y_i;\theta)/\partial\kappa = \int_{-\infty}^{\infty}\left(\partial f(y_i,s_i;\theta)/\partial\kappa\right)ds_i \tag{A.17}$$

Finally, for the third use of Algorithm 2, send it the observed information matrix $I_o(\theta)$. Since $\log L(\theta\,|\,y) = \sum_{i=1}^{m}\log f(y_i;\theta)$, and

$$\frac{\partial^2 \log f(y_i;\theta)}{\partial\theta_j\partial\theta_k} = \frac{f(y_i;\theta)\partial^2 f(y_i;\theta)/\partial\theta_j\partial\theta_k - \left(\partial f(y_i;\theta)/\partial\theta_j\right)\left(\partial f(y_i;\theta)/\partial\theta_k\right)}{f^2(y_i;\theta)} \quad ,$$

it suffices to combine the expressions (A.15)-(A.17) with expressions for the Hessian matrix of $f(y_i;\theta)$. Starting with (A.15)-(A.17), we find

$$\frac{\partial^2 f(y_i;\theta)}{\partial\beta_k\partial\beta_l} = \int_{-\infty}^{\infty}\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\beta_k\partial\beta_l}ds_i \tag{A.18}$$

$$\frac{\partial^2 f(y_i;\theta)}{\partial\sigma^2\partial\beta_l} = \int_{-\infty}^{\infty}\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\sigma^2\partial\beta_l}ds_i \tag{A.19}$$

$$\frac{\partial^2 f(y_i;\theta)}{\partial\kappa\,\partial\beta_l} = \int_{-\infty}^{\infty}\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\kappa\,\partial\beta_l}ds_i \tag{A.20}$$

$$\frac{\partial^2 f(y_i;\theta)}{\partial^2\sigma^2} = \int_{-\infty}^{\infty}\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\sigma^2}ds_i \tag{A.21}$$

$$\frac{\partial^2 f(y_i;\theta)}{\partial\kappa\,\partial\sigma^2} = \int_{-\infty}^{\infty}\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\kappa\,\partial\sigma^2}ds_i \tag{A.22}$$

$$\frac{\partial^2 f(y_i;\theta)}{\partial^2\kappa} = \int_{-\infty}^{\infty}\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\kappa}ds_i \tag{A.23}$$

We can see from (A.18)-(A.23) that what we ultimately need is the Hessian matrix of $f(y_i,s_i;\theta)$, which can be shown, starting with (A.15)-(A.17), to be the following:

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\beta_k\partial\beta_l}=\Bigg\{\left(\sum_{j=1}^{n_i}x_{ijl}\,\frac{\partial\log f(y_{ij}\mid\mu_{ij},\kappa)/\partial\mu_{ij}}{g'(\mu_{ij})}\right)\left(\sum_{j=1}^{n_i}x_{ijk}\,\frac{\partial\log f(y_{ij}\mid\mu_{ij},\kappa)/\partial\mu_{ij}}{g'(\mu_{ij})}\right)\;+$$

$$\sum_{j=1}^{n_i}x_{ijk}x_{ijl}\left[\frac{1}{g'(\mu_{ij})^2}\frac{\partial^2\log f(y_{ij}\mid\mu_{ij},\kappa)}{\partial^2\mu_{ij}}-\frac{g''(\mu_{ij})}{g'(\mu_{ij})^3}\frac{\partial\log f(y_{ij}\mid\mu_{ij},\kappa)}{\partial\mu_{ij}}\right]\Bigg\}f(y_i,s_i;\theta)$$

(A.24)

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\sigma^2\partial\beta_l}=\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)\frac{\partial f(y_i,s_i;\theta)}{\partial\beta_l} \tag{A.25}$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\kappa\,\partial\beta_l}=\left(\sum_{j=1}^{n_i}\frac{\partial\log f(y_{ij}\mid\mu_{ij},\kappa)}{\partial\kappa}\right)\frac{\partial f(y_i,s_i;\theta)}{\partial\beta_l}\;+$$

$$\left(\sum_{j=1}^{n_i}x_{ijl}\,\frac{1}{g'(\mu_{ij})}\frac{\partial^2\log f(y_{ij}\mid\mu_{ij},\kappa)}{\partial\mu_{ij}\partial\kappa}\right)f(y_i,s_i;\theta) \tag{A.26}$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\sigma^2}=\left[\frac{1}{4\sigma^4}\left(\frac{s_i^2}{\sigma^2}-1\right)^2-\frac{1}{\sigma^4}\left(\frac{s_i^2}{\sigma^2}-\frac{1}{2}\right)\right]f(y_i,s_i;\theta) \tag{A.27}$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial\kappa\,\partial\sigma^2}=\frac{1}{2\sigma^2}\left(\frac{s_i^2}{\sigma^2}-1\right)\frac{\partial f(y_i,s_i;\theta)}{\partial\kappa} \tag{A.28}$$

$$\frac{\partial^2 f(y_i,s_i;\theta)}{\partial^2\kappa}=\left[\left(\sum_{j=1}^{n_i}\frac{\partial\log f(y_{ij}\mid\mu_{ij},\kappa)}{\partial\kappa}\right)^2+\sum_{j=1}^{n_i}\frac{\partial^2\log f(y_{ij}\mid\mu_{ij},\kappa)}{\partial^2\kappa}\right]f(y_i,s_i;\theta)$$

(A.29)

A.6

According to (3.1), the four elements, $\mu_{g^{-1}(w)}$, $\mathrm{V}_{g^{-1}(w),y_i}$, V_{y_i,y_i}, and μ_{y_i} in

$\mathrm{BLP}(g^{-1}(w))$, are derived as follows:

$$\mu_{g^{-1}(w)} = E(g^{-1}(w)) = E(e^{\beta_0+s_i}) = e^{\beta_0+\frac{\sigma^2}{2}} = A \tag{A.30}$$

$$\begin{aligned} V_{g^{-1}(w),y_i} &= Cov(g^{-1}(w), y_{i1}\cdots y_{in_i}) \\ &= Cov(g^{-1}(w), y_{i1})\mathbf{1'} \\ &= \left(EE(g^{-1}(w)y_{i1} \mid s_i) - E(g^{-1}(w))\mathrm{EE}(y_{i1} \mid s_i)\right)\mathbf{1'} \\ &= \left(e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right)\right)\mathbf{1'} = B\mathbf{1'} \end{aligned} \tag{A.31}$$

For the diagonal elements in V_{y_i,y_i} , we have:

$$\begin{aligned} V(y_{ij}) &= EVar(y_{ij} \mid s_i) + VarE(y_{ij} \mid s_i) \\ &= E(e^{\beta_0+s_i} + \frac{e^{2\beta_0+2s_i}}{k}) + Var(e^{\beta_0+s_i}) \\ &= A + B + C \end{aligned}$$

For the off diagonal elements in V_{y_i,y_i} , we have:

$$\begin{aligned} &Cov(y_{ij}, y_{ij'}) = EE(y_{ij}y_{ij'} \mid s_i) - EE(y_{ij} \mid s_i)EE(y_{ij'} \mid s_i) \\ &= E\left[E(y_{ij} \mid s_i)E(y_{ij'} \mid s_i)\right] - \left(e^{\beta_0+\frac{\sigma^2}{2}}\right)\left(e^{\beta_0+\frac{\sigma^2}{2}}\right) \\ &= E\left(e^{2\beta_0+2s_i}\right) - e^{2\beta_0+\sigma^2} \\ &= B \end{aligned}$$

Therefore, denote $\boldsymbol{I}$ as an identity matrix and $\boldsymbol{J} = \left[a_{ij}\right]$ where $a_{ij} = 1$

$$V_{y_i,y_i} = (A+C)\boldsymbol{I} + B\boldsymbol{J}$$

Thus, the inverse matrix of V_{y_i,y_i} can be expressed as follows:

$$V^{-1}_{y_i,y_i} = \frac{1}{A+C}\left(\boldsymbol{I} - \frac{B}{A+C+n_iB}\boldsymbol{J}\right) \tag{A.32}$$

Now,

$$\mu_{y_i} = \left[EE\left(y_{ij} \mid s_i\right) \right] \mathbf{1}$$
$$= e^{\beta_0 + \sigma^2/2} \mathbf{1} = A\mathbf{1} \tag{A.33}$$

Assemble pieces from (A.30) to (A.33), it is clear to show (4.1).

A.7

To evaluate $M_1(\theta)$, we need to calculate elements from (A.34) to (A.45) as follows:

$$\begin{aligned}\left(\mu_{g^{-1}(w)} \text{-} \mathrm{V}_{g^{-1}(w),y_i} \mathrm{V}_{y_i,y_i}^{-1} \mu_{y_i}\right) &= A - \frac{B\mathbf{1}'}{A+C}\left(\boldsymbol{I} - \frac{B}{A+C+n_iB}\boldsymbol{J}\right)A\mathbf{1} \\ &= A - \frac{B}{A+C}\left(\mathbf{1}' - \frac{n_iB}{A+C+n_iB}\mathbf{1}'\right)A\mathbf{1} \\ &= A - \frac{B}{A+C}\left(\frac{A+C}{A+C+n_iB}\mathbf{1}'\right)A\mathbf{1} \\ &= A - \frac{n_iAB}{A+C+n_iB}\end{aligned} \tag{A.34}$$

$$\begin{aligned}\left(\mathrm{V}_{g^{-1}(w),y} \mathrm{V}_{y,y}^{-1}\right) &= \frac{B\mathbf{1}'}{A+C}\left(\boldsymbol{I} - \frac{B}{A+C+n_iB}\boldsymbol{J}\right) \\ &= \frac{B}{A+C}\left(\mathbf{1}' - \frac{n_iB}{A+C+n_iB}\mathbf{1}'\right) \\ &= \frac{B}{(A+C+n_iB)}\mathbf{1}'\end{aligned} \tag{A.35}$$

Thus, $\boldsymbol{L}_2 = \left(A - n_iAB/(A+C+n_iB) \quad B/(A+C+n_iB)\mathbf{1}' \quad -1\right)'$ and derive

$E(\boldsymbol{T}) = \mathrm{E}\left(g'(g^{-1}(w)) \quad g'(g^{-1}(w))\left[E\left(y_i \middle| s_i\right)\right]' \quad g'(g^{-1}(w))g^{-1}(w) \right)'$ as follows:

$$\mathrm{E}\left[g'(g^{-1}(w))\right] = \mathrm{E}\left[\frac{1}{e^{\beta_0+s_i}}\right] = e^{-\beta_0+\frac{\sigma^2}{2}} = \frac{1}{A}\left(1+\frac{B}{A^2}\right) \tag{A.36}$$

$$E\left\{g'(g^{-1}(w))\left[E\left(y_i|s_i\right)\right]\right\} = E\left\{\frac{1}{e^{\beta_0+s_i}}\left[e^{\beta_0+s_i}\mathbf{1}\right]\right\} = E\left\{\mathbf{1}\right\} = \mathbf{1} \tag{A.37}$$

$$E\left\{g'(g^{-1}(w))g^{-1}(w)\right\} = E\left\{\frac{1}{e^{\beta_0+s_i}}e^{\beta_0+s_i}\right\} = 1 \tag{A.38}$$

Thus, $E(\boldsymbol{T}) = \left(\frac{1}{A}\left(1+\frac{B}{A^2}\right) \quad \mathbf{1} \quad 1\right)'$

And derive $\boldsymbol{H}_2$ as follows:

$$Var\left\{g'(g^{-1}(w))\right\} = \frac{e^{\sigma^2}}{e^{2\beta_0}}\left(e^{\sigma^2}-1\right) = \frac{B}{A^4}\left(1+\frac{B}{A^2}\right)^2 \tag{A.39}$$

$$Var\left\{g'(g^{-1}(w))E\left(y_i|s_i\right)\right\} = \mathbf{0} \tag{A.40}$$

$$Var\left\{g'(g^{-1}(w))g^{-1}(w)\right\} = 0 \tag{A.41}$$

$$Cov\left(g'(g^{-1}(w)), g'(g^{-1}(w))E\left(y_i|s_i\right)\right) = \mathbf{0} \tag{A.42}$$

$$Cov\left(g'(g^{-1}(w)), g'(g^{-1}(w))g^{-1}(w)\right) = 0 \tag{A.43}$$

$$Cov\left(g'(g^{-1}(w))E\left(y_i|s_i\right), g'(g^{-1}(w))g^{-1}(w)\right) = \mathbf{0} \tag{A.44}$$

And we can derive $E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i|s_i\right)\right\}$ as follows:

$$E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i|s_i\right)\right\} = E\left\{\left(\frac{1}{e^{\beta_0+s_i}}\right)^2\begin{pmatrix} e^{\beta_0+s_i}+\frac{e^{2\beta_0+2s_i}}{\kappa} & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 0 \\ & & & e^{\beta_0+s_i}+\frac{e^{2\beta_0+2s_i}}{\kappa} \end{pmatrix}\right\}$$

$$= \left(e^{-\beta_0+\frac{\sigma^2}{2}} + \frac{1}{\kappa}\right)\boldsymbol{I} = \left(\frac{1}{A}\left(1+\frac{B}{A^2}\right)+\frac{1}{\kappa}\right)\boldsymbol{I}$$

(A.45)

Thus, we can evaluate $M_1(\theta)$ as follows:

$$M_1(\theta) \approx \left\{ \boldsymbol{L}_2' E\left(\boldsymbol{T}\right)\right\}^2 + \boldsymbol{L}_2' H_2 \boldsymbol{L}_2 + \left(\mathrm{V}_{g^{-1}(w),y_i} \mathrm{V}_{y_i,y_i}^{-1}\right) E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i \middle| s_i\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y_i} \mathrm{V}_{y_i,y_i}^{-1}\right)'$$

Where

$$\boldsymbol{L}_2' E\left(\boldsymbol{T}\right) = \begin{pmatrix} A - \dfrac{n_i AB}{A+C+n_iB} & \dfrac{B}{A+C+n_iB}\mathbf{1}' & -1 \end{pmatrix} \begin{pmatrix} \dfrac{1}{A}\left(1+\dfrac{B}{A^2}\right) & \mathbf{1} & 1 \end{pmatrix}'$$

$$= \left(A - \frac{n_i AB}{A+C+n_iB}\right)\frac{1}{A}\left(1+\frac{B}{A^2}\right) + \frac{n_iB}{A+C+n_iB} - 1$$

$$= \left(\frac{A^2 + AC + n_iAB - n_iAB}{A+C+n_iB}\right)\frac{1}{A}\left(1+\frac{B}{A^2}\right) + \frac{n_iB}{A+C+n_iB} - 1$$

$$= A\left(\frac{A+C}{A+C+n_iB}\right)\frac{1}{A}\left(1+\frac{B}{A^2}\right) + \frac{n_iB}{A+C+n_iB} - 1$$

$$= \left(\frac{A+C}{A+C+n_iB}\right)\left(\frac{A^2+B}{A^2}\right) + \frac{n_iB}{A+C+n_iB} - 1$$

$$= \left(\frac{(A+C)(A^2+B) + n_iA^2B}{A^2\left(A+C+n_iB\right)}\right) - 1$$

$$= \frac{B(A+C)}{A^2\left(A+C+n_iB\right)}$$

and

$$\boldsymbol{L}_2' H_2 \boldsymbol{L}_2 = \left(A - \frac{n_iAB}{A+C+n_iB}\right)\frac{B}{A^4}\left(1+\frac{B}{A^2}\right)^2\left(A - \frac{n_iAB}{A+C+n_iB}\right)$$

$$= A\left(\frac{A+C}{A+C+n_iB}\right)\frac{B}{A^4}\left(1+\frac{B}{A^2}\right)^2 A\left(\frac{A+C}{A+C+n_iB}\right)$$

$$= \frac{B}{A^2}\left(\frac{A+C}{A+C+n_iB}\right)^2\left(1+\frac{B}{A^2}\right)^2$$

$$= \frac{B}{A^4}\left(\frac{A+C}{A+C+n_iB}\right)^2\left(A^2+B\right)^2$$

And

$$\left(\mathrm{V}_{g^{-1}(w),y_i}\mathrm{V}^{-1}_{y_i,y_i}\right)E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i \middle| s_i\right)\right\}\left(\mathrm{V}_{g^{-1}(w),y_i}\mathrm{V}^{-1}_{y_i,y_i}\right)'$$

$$=\left(\frac{B}{A+C+n_iB}\mathbf{1}'\right)\left(\frac{1}{A}\left(1+\frac{B}{A^2}\right)+\frac{1}{\kappa}\right)\boldsymbol{I}\left(\frac{B}{A+C+n_iB}\mathbf{1}\right)$$

$$=n_i\left(\frac{B}{A+C+n_iB}\right)^2\left(\frac{1}{A}\left(1+\frac{B}{A^2}\right)+\frac{1}{\kappa}\right)$$

$$=n_i\left(\frac{B}{A+C+n_iB}\right)^2\left(\frac{A^2+B}{A^3}+\frac{1}{\kappa}\right)$$

A.8

To evaluate $A(\theta)$, we know:

$$\mathrm{BLP}(g^{-1}(w)) = A+\frac{n_iB}{A+C+n_iB}\left(\bar{y}_{i.}-A\right) \tag{A.46}$$

We can parameterize (A.46) as follows:

$$\begin{aligned}\mathrm{BLP}(g^{-1}(w)) &= A+\frac{n_iB}{A+C+n_iB}\left(\bar{y}_{i.}-A\right)\\ &=(1-\rho)A+\rho\bar{y}_{i.}\end{aligned} \tag{A.47}$$

Thus $d(y;\theta)=\left(\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\beta_0},\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\sigma^2},\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\kappa}\right)'$

is derived by using $\frac{\partial A}{\partial\beta_0}=A$, $\frac{\partial B}{\partial\beta_0}=2B$, $\frac{\partial C}{\partial\beta_0}=2C$, $\frac{\partial\rho}{\partial\beta_0}=\frac{n_iAB}{\left(A+C+n_iB\right)^2}$,

$\frac{\partial A}{\partial\sigma^2}=\frac{1}{2}A$, $\frac{\partial B}{\partial\sigma^2}=2B+A^2$, $\frac{\partial C}{\partial\sigma^2}=2C$, $\frac{\partial\rho}{\partial\sigma^2}=\frac{3n_iAB+2n_iA^3+2n_iA^2C}{2\left(A+C+n_iB\right)^2}$, $\frac{\partial A}{\partial\kappa}=0$,

$\frac{\partial B}{\partial\kappa}=0$, $\frac{\partial C}{\partial\kappa}=-C/\kappa$ and $\frac{\partial\rho}{\partial\kappa}=n_iBC/(\kappa(A+C+n_iB)^2)$ as follows:

$$\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\beta_0}=\frac{\partial\log\left[(1-\rho)A+\rho\bar{y}_{i.}\right]}{\partial\beta_0}$$

$$=\frac{\frac{\partial(1-\rho)}{\partial\beta_0}A+(1-\rho)\frac{\partial A}{\partial\beta_0}+\frac{\partial\rho}{\partial\beta_0}\bar{y}_{i.}}{(1-\rho)A+\rho\bar{y}_{i.}} \qquad \text{(A.48)}$$

$$=\frac{d_1(\bar{y}_{i.}-A)+d_2}{(1-\rho)A+\rho\bar{y}_{i.}}$$

Where $d_1=\frac{\partial\rho}{\partial\beta_0}$ and $d_2=(1-\rho)\frac{\partial A}{\partial\beta_0}$.

And

$$\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\sigma^2}=\frac{\partial\log\left[(1-\rho)A+\rho\bar{y}_{i.}\right]}{\partial\sigma^2}$$

$$=\frac{\frac{\partial(1-\rho)}{\partial\sigma^2}A+(1-\rho)\frac{\partial A}{\partial\sigma^2}+\frac{\partial\rho}{\partial\sigma^2}\bar{y}_{i.}}{(1-\rho)A+\rho\bar{y}_{i.}} \qquad \text{(A.49)}$$

$$=\frac{d_3(\bar{y}_{i.}-A)+d_4}{(1-\rho)A+\rho\bar{y}_{i.}}$$

Where $d_3=\frac{\partial\rho}{\partial\sigma^2}$ and $d_4=(1-\rho)\frac{\partial A}{\partial\sigma^2}$.

And

$$\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\kappa}=\frac{\partial\log\left[(1-\rho)A+\rho\bar{y}_{i.}\right]}{\partial\kappa}$$

$$=\frac{\frac{\partial(1-\rho)}{\partial\kappa}A+(1-\rho)\frac{\partial A}{\partial\kappa}+\frac{\partial\rho}{\partial\kappa}\bar{y}_{i.}}{(1-\rho)A+\rho\bar{y}_{i.}} \qquad \text{(A.50)}$$

$$=\frac{d_5(\bar{y}_{i.}-A)+d_6}{(1-\rho)A+\rho\bar{y}_{i.}}$$

Where $d_5=\frac{\partial\rho}{\partial\kappa}$ and $d_6=(1-\rho)\frac{\partial A}{\partial\kappa}$.

Based on (A.48), (A.49) and (A.50), the term, $E\left(d(y;\theta)d'(y;\theta)\right)$, can be derived as follows:

$$E\left(d(y;\theta)d'(y;\theta)\right) = [E(f_{ij}(\bar{y}_{i.}))] \qquad \text{(A.51)}$$

where $[E(f_{ij}(\bar{y}_{i.}))]$ is a 3 by 3 matrix and given by the following elements

$f_{11}(\bar{y}_{i.}) = \dfrac{(d_1(\bar{y}_{i.} - A) + d_2)^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}$, $f_{22}(\bar{y}_{i.}) = \dfrac{(d_3(\bar{y}_{i.} - A) + d_4)^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}$, $f_{33}(\bar{y}_{i.}) = \dfrac{(d_5(\bar{y}_{i.} - A) + d_6)^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}$,

$f_{12}(\bar{y}_{i.}) = f_{21}(\bar{y}_{i.}) = \dfrac{(d_1(\bar{y}_{i.} - A) + d_2)(d_3(\bar{y}_{i.} - A) + d_4)}{((1-\rho)A + \rho\bar{y}_{i.})^2}$,

$f_{13}(\bar{y}_{i.}) = f_{31}(\bar{y}_{i.}) = \dfrac{(d_1(\bar{y}_{i.} - A) + d_2)(d_5(\bar{y}_{i.} - A) + d_6)}{((1-\rho)A + \rho\bar{y}_{i.})^2}$ and

$f_{23}(\bar{y}_{i.}) = f_{32}(\bar{y}_{i.}) = \dfrac{(d_3(\bar{y}_{i.} - A) + d_4)(d_5(\bar{y}_{i.} - A) + d_6)}{((1-\rho)A + \rho\bar{y}_{i.})^2}$. In (A.51), we approximate $E(f_{ij}(\bar{y}_{i.}))$ by considering a second Taylor expansion of $f_{ij}(\bar{y}_{i.})$ around $\mathrm{E}(\bar{y}_{i.}) = A$.

Then we have: $E(f_{ij}(\bar{y}_{i.})) \approx f_{ij}(A) + f''_{ij}(A)Var[\bar{\mathrm{y}}_{i.}]/2$ and $Var\left[\bar{y}_{i.}\right] = B + (A+C)/n_i$ and the calculation of $f_{ij}(A)$ and $f''_{ij}(A)$ are as follows:

For the diagonal elements in $[E(f_{ij}(\bar{y}_{i.}))]$:

$$f_{11}(A) = \frac{(d_1(A-A) + d_2)^2}{((1-\rho)A + \rho A)^2} = \frac{d_2^2}{A^2},$$

And we know:

$$f'_{11}(\bar{y}_{i.}) = \frac{(d_1(\bar{y}_{i.} - A) + d_2)^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}\left[\frac{2d_1}{d_1(\bar{y}_{i.} - A) + d_2} - \frac{2\rho}{((1-\rho)A + \rho\bar{y}_{i.})}\right]$$

$$f''_{11}(\bar{y}_{i.}) = \frac{(d_1(\bar{y}_{i.} - A) + d_2)^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}\left[\frac{2d_1}{d_1(\bar{y}_{i.} - A) + d_2} - \frac{2\rho}{((1-\rho)A + \rho\bar{y}_{i.})}\right]^2 + \frac{(d_1(\bar{y}_{i.} - A) + d_2)^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}\left[\frac{-2d_1^2}{(d_1(\bar{y}_{i.} - A) + d_2)^2} + \frac{2\rho^2}{((1-\rho)A + \rho\bar{y}_{i.})^2}\right]$$

Thus,

$$f_{11}''(A)=\frac{(d_1(A-A)+d_2)^2}{((1-\rho)A+\rho A)^2}\left[\frac{2d_1}{d_1(A-A)+d_2}-\frac{2\rho}{((1-\rho)A+\rho A)}\right]^2$$

$$+\frac{(d_1(A-A)+d_2)^2}{((1-\rho)A+\rho A)^2}\left[\frac{-2d_1^2}{(d_1(A-A)+d_2)^2}+\frac{2\rho^2}{((1-\rho)A+\rho A)^2}\right]$$

$$=\frac{d_2^2}{A^2}\left\{\left[\frac{2d_1}{d_2}-\frac{2\rho}{A}\right]^2-\frac{2d_1^2}{d_2^2}+\frac{2\rho^2}{A^2}\right\}$$

$$=\frac{d_2^2}{A^2}\left\{\left[\frac{4d_1^2}{d_2^2}+\frac{4\rho^2}{A^2}-\frac{8\rho d_1}{Ad_2}\right]-\frac{2d_1^2}{d_2^2}+\frac{2\rho^2}{A^2}\right\}$$

$$=\frac{d_2^2}{A^2}\left[\frac{2d_1^2}{d_2^2}+\frac{6\rho^2}{A^2}-\frac{8\rho d_1}{Ad_2}\right]$$

$$=\frac{d_2^2}{A^2}\left[\frac{2A^2d_1^2+6\rho^2d_2^2-8\rho Ad_1d_2}{A^2d_2^2}\right]$$

$$=\frac{2A^2d_1^2+6\rho^2d_2^2-8\rho Ad_1d_2}{A^4}$$

Similarly,

$$f_{22}(A)=\frac{(d_3(A-A)+d_4)^2}{((1-\rho)A+\rho A)^2}=\frac{d_4^2}{A^2},\quad f_{22}''(A)=(-8A\rho d_3d_4+2A^2d_3^2+6\rho^2d_4^2)/A^4,$$

$$f_{33}(A)=\frac{(d_5(A-A)+d_6)^2}{((1-\rho)A+\rho A)^2}=\frac{d_6^2}{A^2},\quad f_{33}''(A)=(-8A\rho d_5d_6+2A^2d_5^2+6\rho^2d_6^2)/A^4,$$

For the off diagonal elements in $[E(f_{ij}(\bar{y}_{i.}))]$:

$$f_{12}(A)=\frac{(d_1(A-A)+d_2)(d_3(A-A)+d_4)}{((1-\rho)A+\rho A)^2}=\frac{d_2d_4}{A^2},$$

And we know:

$$f_{12}'(A)=\frac{(d_1(\bar{y}_{i.}-A)+d_2)(d_3(\bar{y}_{i.}-A)+d_4)}{((1-\rho)A+\rho A)^2}\left[\frac{d_1}{(d_1(\bar{y}_{i.}-A)+d_2)}+\frac{d_3}{(d_3(\bar{y}_{i.}-A)+d_4)}-\frac{2\rho}{((1-\rho)A+\rho A)}\right]$$

$$f_{12}''(A)=\frac{(d_1(\bar{y}_{i.}-A)+d_2)(d_3(\bar{y}_{i.}-A)+d_4)}{((1-\rho)A+\rho A)^2}\left[\frac{d_1}{(d_1(\bar{y}_{i.}-A)+d_2)}+\frac{d_3}{(d_3(\bar{y}_{i.}-A)+d_4)}-\frac{2\rho}{((1-\rho)A+\rho A)}\right]^2$$
$$+\frac{(d_1(\bar{y}_{i.}-A)+d_2)(d_3(\bar{y}_{i.}-A)+d_4)}{((1-\rho)A+\rho A)^2}\left[\frac{-d_1^2}{(d_1(\bar{y}_{i.}-A)+d_2)^2}+\frac{-d_3^2}{(d_3(\bar{y}_{i.}-A)+d_4)^2}+\frac{2\rho^2}{((1-\rho)A+\rho A)^2}\right]$$

Thus,

$$f_{12}''(A)=\frac{(d_1(A-A)+d_2)(d_3(A-A)+d_4)}{((1-\rho)A+\rho A)^2}\left[\frac{d_1}{(d_1(A-A)+d_2)}+\frac{d_3}{(d_3(A-A)+d_4)}-\frac{2\rho}{((1-\rho)A+\rho A)}\right]^2$$
$$+\frac{(d_1(A-A)+d_2)(d_3(A-A)+d_4)}{((1-\rho)A+\rho A)^2}\left[\frac{-d_1^2}{(d_1(A-A)+d_2)^2}+\frac{-d_3^2}{(d_3(A-A)+d_4)^2}+\frac{2\rho^2}{((1-\rho)A+\rho A)^2}\right]$$
$$=\frac{d_2d_4}{A^2}\left[\frac{d_1}{d_2}+\frac{d_3}{d_4}-\frac{2\rho}{A}\right]^2+\frac{d_2d_4}{A^2}\left[\frac{-d_1^2}{d_2^2}+\frac{-d_3^2}{d_4^2}+\frac{2\rho^2}{A^2}\right]$$
$$=\frac{d_2d_4}{A^2}\left[\frac{d_1^2}{d_2^2}+\frac{d_3^2}{d_4^2}+\frac{4\rho^2}{A^2}+2\left(\frac{d_1d_3}{d_2d_4}-\frac{d_1}{d_2}\frac{2\rho}{A}-\frac{d_3}{d_4}\frac{2\rho}{A}\right)+\frac{-d_1^2}{d_2^2}+\frac{-d_3^2}{d_4^2}+\frac{2\rho^2}{A^2}\right]$$
$$=\frac{d_2d_4}{A^2}\left[\frac{6\rho^2}{A^2}+2\left(\frac{d_1d_3}{d_2d_4}-\frac{d_1}{d_2}\frac{2\rho}{A}-\frac{d_3}{d_4}\frac{2\rho}{A}\right)\right]$$
$$=\frac{d_2d_4}{A^2}\left[\frac{6\rho^2}{A^2}+2A\left(\frac{Ad_1d_3-2\rho d_1d_4-2\rho d_2d_3}{d_2d_4A^2}\right)\right]$$
$$=\frac{1}{A^4}\left[6\rho^2d_2d_4+2A(Ad_1d_3-2\rho d_1d_4-2\rho d_2d_3)\right]$$

Similarly,

$$f_{13}(A)=\frac{(d_1(A-A)+d_2)(d_5(A-A)+d_6)}{((1-\rho)A+\rho A)^2}=\frac{d_2d_6}{A^2},$$

$$f_{13}''(A)=\frac{1}{A^4}\left[6\rho^2d_2d_6+2A(Ad_1d_5-2\rho d_1d_6-2\rho d_2d_5)\right]$$

$$f_{23}(A)=\frac{(d_3(A-A)+d_4)(d_5(A-A)+d_6)}{((1-\rho)A+\rho A)^2}=\frac{d_6d_4}{A^2},$$

$$f_{23}''(A)=\frac{1}{A^4}\left[6\rho^2d_4d_6+2A(Ad_3d_5-2\rho d_3d_6-2\rho d_4d_5)\right]$$

Thus, we can approximate $\mathrm{E}\left(d(y;\theta)d'(y;\theta)\right)$ using the result of approximation of

$[E(f_{ij}(\bar{y}_{i.}))]$.

A.9

According to (3.1), the four elements, $\mu_{g^{-1}(w)}$, $\mathrm{V}_{g^{-1}(w),y_i}$, V_{y_i,y_i}, and μ_{y_i} in

$\mathrm{BLP}(g^{-1}(w))$, are derived as follows:

$$\mu_{g^{-1}(w)} = \mathrm{E}(g^{-1}(w)) = \mathrm{E}(e^{\beta_0+s_i}) = e^{\beta_0+\frac{\sigma^2}{2}} = A \tag{A.52}$$

$$\begin{aligned}\mathrm{V}_{g^{-1}(w),y_i} &= Cov(g^{-1}(w), y_{i1}\cdots y_{in_i}) \\ &= Cov(g^{-1}(w), y_{i1})\mathbf{1}' \\ &= \left(EE(g^{-1}(w)y_{i1} \mid s_i) - E(g^{-1}(w))EE(y_{i1} \mid s_i)\right)\mathbf{1}' \\ &= \left(e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right)\right)\mathbf{1}' = B\mathbf{1}'\end{aligned} \tag{A.53}$$

For the diagonal elements in V_{y_i,y_i}, we have:

$$\begin{aligned}Var(y_{ij}) &= EVar(y_{ij} \mid s_i) + VarE(y_{ij} \mid s_i) \\ &= A + B\end{aligned}$$

For the off diagonal elements in V_{y_i,y_i}, we have:

$$\begin{aligned}&Cov(y_{ij}, y_{ij'}) = EE(y_{ij}y_{ij'} \mid s_i) - EE(y_{ij} \mid s_i)EE(y_{ij'} \mid s_i) \\ &= E\left[E(y_{ij} \mid s_i)E(y_{ij'} \mid s_i)\right] - \left(e^{\beta_0+\frac{\sigma^2}{2}}\right)\left(e^{\beta_0+\frac{\sigma^2}{2}}\right) \\ &= E\left(e^{2\beta_0+2s_i}\right) - e^{2\beta_0+\sigma^2} \\ &= B\end{aligned}$$

Thus, the inverse matrix of V_{y_i,y_i} can be expressed as follows:

$$\mathrm{V}_{y_i,y_i}^{-1} = \frac{1}{A}\left(\boldsymbol{I} - \frac{B}{A+nB}\boldsymbol{J}\right) \tag{A.54}$$

Now,

$$\begin{aligned}\mu_{y_i} &= \left[EE\left(y_{ij} \mid s_i\right)\right]\mathbf{1} \\ &= e^{\beta_0+\sigma^2/2}\mathbf{1} = A\mathbf{1}\end{aligned} \tag{A.55}$$

Assemble pieces from (A.52) to (A.55), it is clear to show (4.2).

A.10

Based on A.9, we have: $\boldsymbol{L}_2 = \left(\left(\dfrac{A^2}{A+n_i B} \right) \quad \left(\dfrac{B}{A+n_i B} \right) \mathbf{1}' \quad -1 \right)'$ and derive

$E(\boldsymbol{T}) = E\left(g'(g^{-1}(w)) \quad g'(g^{-1}(w)) \left[E\left(y_i | s_i \right) \right]^T \quad g'(g^{-1}(w)) g^{-1}(w) \right)'$ as follows:

$$E\left[g'(g^{-1}(w)) \right] = E\left[\frac{1}{e^{\beta_0 + s_i}} \right] = e^{-\beta_0 + \frac{\sigma^2}{2}} \tag{A.56}$$

$$E\left\{ g'(g^{-1}(w)) \left[E\left(y_i | s_i \right) \right] \right\} = E\left\{ \frac{1}{e^{\beta_0 + s_i}} \left[e^{\beta_0 + s_i} \mathbf{1} \right] \right\} = \mathrm{E}\left\{ \mathbf{1} \right\} = \mathbf{1} \tag{A.57}$$

$$\mathrm{E}\left\{ g'(g^{-1}(w)) g^{-1}(w) \right\} = E\left\{ \frac{1}{e^{\beta_0 + s_i}} e^{\beta_0 + s_i} \right\} = 1 \tag{A.58}$$

And derive $\boldsymbol{H}_2$ as follows:

$$Var\left\{ g'(g^{-1}(w)) \right\} = \frac{e^{\sigma^2}}{e^{2\beta_0}} \left(e^{\sigma^2} - 1 \right) \tag{A.59}$$

$$Var\left\{ g'(g^{-1}(w)) E\left(y_i | s_i \right) \right\} = \mathbf{0} \tag{A.60}$$

$$Var\left\{ g'(g^{-1}(w)) g^{-1}(w) \right\} = 0 \tag{A.61}$$

$$Cov\left(g'(g^{-1}(w)), g'(g^{-1}(w)) E\left(y_i | s_i \right) \right) = \mathbf{0} \tag{A.62}$$

$$Cov\left(g'(g^{-1}(w)), g'(g^{-1}(w)) g^{-1}(w) \right) = 0 \tag{A.63}$$

$$Cov\left(g'(g^{-1}(w)) E\left(y_i | s_i \right), g'(g^{-1}(w)) g^{-1}(w) \right) = \mathbf{0} \tag{A.64}$$

From (A.56) to (A.64), we have the following result:

$$\boldsymbol{H}_2 = \begin{pmatrix} \frac{e^{\sigma^2}}{e^{2\beta_0}}\left(e^{\sigma^2}-1\right) & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & 0 \end{pmatrix} \tag{A.65}$$

And we can derive $E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i \middle| s_i\right)\right\}$ as follows:

$$E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i \middle| s_i\right)\right\} = E\left\{\left(\frac{1}{e^{\beta_0+s_i}}\right)^2 \begin{pmatrix} e^{\beta_0+s_i} & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 0 \\ & & & e^{\beta_0+s_i} \end{pmatrix}\right\} = e^{-\beta_0+\frac{\sigma^2}{2}}\boldsymbol{I}$$

(A.66)

Thus, we have the following result:

$$\begin{aligned} M_1(\theta) &\approx \left\{\frac{A^2}{A+n_iB}\left(e^{-\beta_0+\frac{\sigma^2}{2}}\right) - \frac{A}{A+n_iB}\right\}^2 + \left(\frac{A^2}{A+n_iB}\right)^2 \frac{e^{\sigma^2}}{e^{2\beta_0}}\left(e^{\sigma^2}-1\right) + n_i\left(\frac{B}{A+n_iB}\right)^2\left(e^{-\beta_0+\frac{\sigma^2}{2}}\right) \\ &= \left\{\frac{A}{A+n_iB}\left(1+\frac{B}{A^2}\right) - \frac{A}{A+n_iB}\right\}^2 + \left(\frac{1}{A+n_iB}\right)^2\left(B\left(1+\frac{B}{A^2}\right)^2\right) + n_i\left(\frac{B}{A+n_iB}\right)^2\left(\frac{1}{A}\left(1+\frac{B}{A^2}\right)\right) \\ &= B\left(\frac{1}{A+n_iB}\right)^2\left(1+\frac{n_iB}{A}+\frac{3B}{A^2}+\frac{n_iB^2}{A^3}+\frac{B^2}{A^4}\right) \end{aligned}$$

A.11

From **(4.2)**, we know:

$$g[\text{BLP}(g^{-1}(\beta_0+s_i))] = \log\left(A + \frac{B}{A+n_iB}\left(y_{i.} - n_iA\right)\right) \tag{A.67}$$

We can parameterize (A.67) as follows:

$$g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]=\log\left(A+\frac{B}{A+n_iB}\left(y_{i.}-n_iA\right)\right)$$
$$=\log\left(A+\frac{n_iB}{A+n_iB}\left(\bar{y}_{i.}-A\right)\right) \quad \text{(A.68)}$$
$$=\log\left((1-\rho)A+\rho\bar{y}_{i.}\right)$$

Thus $d(y;\theta)=\left(\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\beta_0},\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\sigma^2}\right)'$ is derived by using

$\frac{\partial A}{\partial\beta_0}=A$, $\frac{\partial B}{\partial\beta_0}=2B$, $\frac{\partial\rho}{\partial\beta_0}=\frac{n_iAB}{\left(A+n_iB\right)^2}$, $\frac{\partial A}{\partial\sigma^2}=\frac{1}{2}A$, $\frac{\partial B}{\partial\sigma^2}=2B+A^2$ and

$\frac{\partial\rho}{\partial\sigma^2}=\frac{3n_iAB+2n_iA^3}{2\left(A+n_iB\right)^2}$ as follows:

$$\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\beta_0}=\frac{\partial\log\left((1-\rho)A+\rho\bar{y}_{i.}\right)}{\partial\beta_0}$$
$$=\frac{\frac{-n_iAB}{\left(A+n_iB\right)^2}A+(1-\rho)A+\frac{n_iAB}{\left(A+n_iB\right)^2}\bar{y}_{i.}}{(1-\rho)A+\rho\bar{y}_{i.}}$$
$$=\frac{\frac{-A^2}{A+n_iB}\rho+(1-\rho)A+\frac{A}{A+n_iB}\rho\bar{y}_{i.}}{(1-\rho)A+\rho\bar{y}_{i.}}$$
$$=\frac{\rho\left(\frac{-A^2}{A+n_iB}+\frac{A}{A+n_iB}\bar{y}_{i.}\right)+(1-\rho)A}{(1-\rho)A+\rho\bar{y}_{i.}}$$
$$=\frac{\rho\left(-A\left(1-\rho\right)+\left(1-\rho\right)\bar{y}_{i.}\right)+(1-\rho)A}{(1-\rho)A+\rho\bar{y}_{i.}}$$
$$=\frac{\rho\left(\left(1-\rho\right)\left(\bar{y}_{i.}-A\right)\right)+(1-\rho)A}{(1-\rho)A+\rho\bar{y}_{i.}}$$
$$=\frac{\left(1-\rho\right)\left(\rho\left(\bar{y}_{i.}-A\right)+A\right)}{(1-\rho)A+\rho\bar{y}_{i.}}=\left(1-\rho\right) \quad \text{(A.69)}$$

and

$$
\begin{aligned}
\frac{\partial g[\text{BLP}(g^{-1}(\beta_0+s_i))]}{\partial\sigma^2} &= \frac{\partial\log\left((1-\rho)A+\rho\bar{y}_{i.}\right)}{\partial\sigma^2} \\
&= \frac{-\left(\dfrac{3n_iAB+2n_iA^3}{2\left(A+n_iB\right)^2}\right)A+(1-\rho)\dfrac{A}{2}+\left(\dfrac{3n_iAB+2n_iA^3}{2\left(A+n_iB\right)^2}\right)\bar{y}_{i.}}{(1-\rho)A+\rho\bar{y}_{i.}} \\
&= \frac{\dfrac{3n_iAB+2n_iA^3}{2\left(A+n_iB\right)^2}\left(\bar{y}_{i.}-A\right)+\dfrac{1}{2}\left(1-\rho\right)A}{(1-\rho)A+\rho\bar{y}_{i.}} \\
&= \frac{\left(\dfrac{3}{2}\times\dfrac{n_iAB}{\left(A+n_iB\right)^2}+\dfrac{n_iA^3}{\left(A+n_iB\right)^2}\right)\left(\bar{y}_{i.}-A\right)+\dfrac{1}{2}\left(1-\rho\right)A}{(1-\rho)A+\rho\bar{y}_{i.}} \\
&= \frac{\left(\dfrac{3}{2}\rho\left(1-\rho\right)+n_iA\left(1-\rho\right)^2\right)\left(\bar{y}_{i.}-A\right)+\dfrac{1}{2}\left(1-\rho\right)A}{(1-\rho)A+\rho\bar{y}_{i.}} \\
&= \frac{\left(1-\rho\right)\left[\left(\dfrac{3}{2}\rho+n_iA\left(1-\rho\right)\right)\left(\bar{y}_{i.}-A\right)+\dfrac{1}{2}A\right]}{(1-\rho)A+\rho\bar{y}_{i.}} \\
&= \frac{\left(1-\rho\right)\left[\left(3\rho+2n_iA\left(1-\rho\right)\right)\left(\bar{y}_{i.}-A\right)+A\right]}{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]} \\
&= \frac{\left(1-\rho\right)\left[\left(3\rho+2n_iA\left(1-\rho\right)\right)\bar{y}_{i.}+A\left(1-2n_iA(1-\rho)-3\rho\right)\right]}{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]} \\
&= \left(1-\rho\right)f\left(\bar{y}_{i.}\right)
\end{aligned}
$$

(A.70)

Where $f\left(\bar{y}_{i.}\right)=\dfrac{\left[\left(3\rho+2n_iA\left(1-\rho\right)\right)\bar{y}_{i.}+A\left(1-2n_iA(1-\rho)-3\rho\right)\right]}{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]}$

Based on (A.69) and (A.70), the term, $E\left(d(y;\theta)d'(y;\theta)\right)$, can be derived as follows:

$$E\left(d(y;\theta)d'(y;\theta)\right)=E\left[\begin{pmatrix}1-\rho\\(1-\rho)f(\bar{y}_{i.})\end{pmatrix}\begin{pmatrix}1-\rho & (1-\rho)f(\bar{y}_{i.})\end{pmatrix}\right]$$
$$=E\begin{bmatrix}(1-\rho)^2 & (1-\rho)^2 f(\bar{y}_{i.})\\(1-\rho)^2 f(\bar{y}_{i.}) & (1-\rho)^2 f^2(\bar{y}_{i.})\end{bmatrix}=(1-\rho)^2 E\begin{bmatrix}1 & f(\bar{y}_{i.})\\ f(\bar{y}_{i.}) & f^2(\bar{y}_{i.})\end{bmatrix} \quad \text{(A.71)}$$

In (A.71), we approximate $E\left(f(\bar{y}_{i.})\right)$ and $E\left(f^2(\bar{y}_{i.})\right)$ by considering the second order Taylor expansion of $f(\bar{y}_{i.})$ and $f^2(\bar{y}_{i.})$ around $E(\bar{y}_{i.})$ as follows:

$$f(\bar{y}_{i.})\approx f\left(E(\bar{y}_{i.})\right)+f'\left(E(\bar{y}_{i.})\right)\left[\bar{y}_{i.}-E(\bar{y}_{i.})\right]+\frac{1}{2}f''\left(E(\bar{y}_{i.})\right)\left[\bar{y}_{i.}-E(\bar{y}_{i.})\right]^2$$

(A.72)

Thus,

$$E\left[f(\bar{y}_{i.})\right]\approx E\left[f\left(E(\bar{y}_{i.})\right)\right]+\frac{1}{2}f''\left(E(\bar{y}_{i.})\right)E\left[\bar{y}_{i.}-E(\bar{y}_{i.})\right]^2$$
$$=f\left(E(\bar{y}_{i.})\right)+\frac{1}{2}f''\left(E(\bar{y}_{i.})\right)Var\left[\bar{y}_{i.}\right]$$

(A.73)

where $f'\left(E(\bar{y}_{i.})\right)$, $f''\left(E(\bar{y}_{i.})\right)$ and $Var\left[\bar{y}_{i.}\right]$ are derived as follows:

$$f'(\bar{y}_{i.})=\frac{2\left(3\rho+2n_iA(1-\rho)\right)\left[(1-\rho)A+\rho\bar{y}_{i.}\right]-2\rho\left\{\left[3\rho+2n_iA(1-\rho)\right]\bar{y}_{i.}+A\left(1-3\rho-2n_iA(1-\rho)\right)\right\}}{\left\{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]\right\}^2}$$
$$=\frac{\left(3\rho+2n_iA(1-\rho)\right)\left\{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]-2\rho\bar{y}_{i.}+2A\rho\right\}-2A\rho}{\left\{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]\right\}^2}$$
$$=\frac{\left(3\rho+2n_iA(1-\rho)\right)\left\{2\left[(1-\rho)A\right]+2A\rho\right\}-2A\rho}{\left\{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]\right\}^2}$$
$$=\frac{2A\left(2\rho+2n_iA(1-\rho)\right)}{\left\{2\left[(1-\rho)A+\rho\bar{y}_{i.}\right]\right\}^2}$$
$$=\frac{A\left(\rho+n_iA(1-\rho)\right)}{\left[(1-\rho)A+\rho\bar{y}_{i.}\right]^2}$$

(A.74)

And

$$f''(\bar{y}_{i.}) = \left(\frac{A\left(\rho + n_i A(1-\rho)\right)}{\left[(1-\rho)A + \rho\bar{y}_{i.}\right]^2} \right)' \\ = \frac{-2A\rho\left(\rho + n_i A(1-\rho)\right)}{\left[(1-\rho)A + \rho\bar{y}_{i.}\right]^3} \tag{A.75}$$

Using (A.74) and (A.75), we have the following results:

$$f'\left(\mathrm{E}\left(\bar{y}_{i.}\right)\right) = \frac{A\left(\rho + n_i A(1-\rho)\right)}{\left[(1-\rho)A + \rho\mathrm{E}\left(\bar{y}_{i.}\right)\right]^2} \\ = \frac{A\left(\rho + n_i A(1-\rho)\right)}{\left[(1-\rho)A + \rho A\right]^2} \\ = \frac{A\left(\rho + n_i A(1-\rho)\right)}{A^2} \\ = \frac{\left(\rho + n_i A(1-\rho)\right)}{A} \tag{A.76}$$

And

$$f''\left(\mathrm{E}\left(\bar{y}_{i.}\right)\right) = \frac{-2A\rho\left(\rho + n_i A(1-\rho)\right)}{\left[(1-\rho)A + \rho A\right]^3} \\ = \frac{-2A\rho\left(\rho + n_i A(1-\rho)\right)}{A^3} \\ = \frac{-2\rho\left(\rho + n_i A(1-\rho)\right)}{A^2} \tag{A.77}$$

And

$$
\begin{aligned}
Var\left[\bar{y}_{i.}\right] &= VarE\left[\bar{y}_{i.} \mid s_i\right] + EVar\left[\bar{y}_{i.} \mid s_i\right] \\
&= Var\left(e^{\beta_0 + s_i}\right) + E\left(\frac{e^{\beta_0 + s_i}}{n_i}\right) \qquad \text{(A.78)} \\
&= B + \frac{A}{n_i} = \frac{A + nB}{n_i} = \frac{B}{\rho}
\end{aligned}
$$

Now, using (A.76), (A.77) and (A.78), we have the following results:

$$
\begin{aligned}
\mathrm{E}\left[f\left(\bar{y}_{i.}\right)\right] &\approx E\left[f\left(E\left(\bar{y}_{i.}\right)\right)\right] + \frac{1}{2} f''\left(E\left(\bar{y}_{i.}\right)\right) E\left[\bar{y}_{i.} - E\left(\bar{y}_{i.}\right)\right]^2 \\
&= f\left(E\left(\bar{y}_{i.}\right)\right) + \frac{1}{2} f''\left(E\left(\bar{y}_{i.}\right)\right) Var\left[\bar{y}_{i.}\right] \\
&= \frac{\left[\left(3\rho + 2n_i A(1-\rho)\right) A + A\left(1 - 2n_i A(1-\rho) - 3\rho\right)\right]}{2\left[(1-\rho)A + \rho A\right]} + \frac{1}{2} \times \frac{-2\rho\left(\rho + n_i A(1-\rho)\right)}{A^2} \times \frac{B}{\rho} \\
&= \frac{3\rho A + 2n_i A^2 (1-\rho) + A - 2n_i A^2 (1-\rho) - 3\rho A}{2\left[(1-\rho)A + \rho A\right]} + \frac{-B\left(\rho + n_i A(1-\rho)\right)}{A^2} \\
&= \frac{A}{2\left[(1-\rho)A + \rho A\right]} + \frac{-\rho\left(\rho + n_i A(1-\rho)\right)}{n_i A(1-\rho)} = \frac{1}{2} - \frac{\rho\left(\rho + n_i A(1-\rho)\right)}{n_i A(1-\rho)}
\end{aligned}
$$

(A.79)

Now, a second order Taylor expansion of $f^2\left(\bar{y}_{i.}\right)$ around $E\left(\bar{y}_{i.}\right)$ yields:

$$
\begin{aligned}
f^2\left(\bar{y}_{i.}\right) \approx\ & f^2\left(E\left(\bar{y}_{i.}\right)\right) + 2f\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right)\left[\bar{y}_{i.} - E\left(\bar{y}_{i.}\right)\right] \\
&+ \left[f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right)\right]\left[\bar{y}_{i.} - E\left(\bar{y}_{i.}\right)\right]^2 \qquad \text{(A.80)}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E\left[f^2\left(\bar{y}_{i.}\right)\right] &\approx E\left[f^2\left(E\left(\bar{y}_{i.}\right)\right)\right] + \left[f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right)\right] E\left[\bar{y}_{i.} - E\left(\bar{y}_{i.}\right)\right]^2 \\
&= f^2\left(E\left(\bar{y}_{i.}\right)\right) + \left[f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right)\right] E\left[\bar{y}_{i.} - E\left(\bar{y}_{i.}\right)\right]^2 \\
&= f^2\left(E\left(\bar{y}_{i.}\right)\right) + \left[f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right)\right] Var\left[\bar{y}_{i.}\right]
\end{aligned}
$$

(A.81)

Using (A.76) and (A.77), we can derive $f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right)$ as follows:

$$\begin{aligned} f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right) &= \left[\frac{\left(\rho + n_i A(1-\rho)\right)}{A}\right]^2 + \frac{1}{2} \times \frac{-2\rho\left(\rho + n_i A(1-\rho)\right)}{A^2} \\ &= \frac{\rho + n_i A(1-\rho)}{A^2}\left[\rho + n_i A(1-\rho) + \frac{1}{2}(-2\rho)\right] \\ &= \frac{n_i A(1-\rho)\left(\rho + n_i A(1-\rho)\right)}{A^2} = \frac{n_i(1-\rho)\left(\rho + n_i A(1-\rho)\right)}{A} \end{aligned}$$

(A.82)

We know $f\left(E\left(\bar{y}_{i.}\right)\right) = \frac{1}{2}$ and $f^2\left(E\left(\bar{y}_{i.}\right)\right) = \frac{1}{4}$ and after replacing (A.81) with (A.78) and (A.82). We have the following result:

$$\begin{aligned} E\left[f^2\left(\bar{y}_{i.}\right)\right] &\approx f^2\left(E\left(\bar{y}_{i.}\right)\right) + \left[f'\left(E\left(\bar{y}_{i.}\right)\right) f'\left(E\left(\bar{y}_{i.}\right)\right) + f\left(E\left(\bar{y}_{i.}\right)\right) f''\left(E\left(\bar{y}_{i.}\right)\right)\right] Var\left[\bar{y}_{i.}\right] \\ &= \frac{1}{4} + \left[\frac{n_i(1-\rho)\left(\rho + n_i A(1-\rho)\right)}{A}\right]\frac{B}{\rho} \\ &= \frac{1}{4} + \left[\frac{n_i\left(\rho + n_i A(1-\rho)\right)B}{A}\right]\frac{1-\rho}{\rho} \\ &= \frac{1}{4} + \frac{n_i\left(\rho + n_i A(1-\rho)\right)B}{A} \times \frac{A}{n_i B} = \frac{1}{4} + \rho + n_i A(1-\rho) \end{aligned}$$

(A.83)

Thus, using (A.79) and (A.83), we can approximate $E\left(d(y;\theta)d'(y;\theta)\right)$ as follows:

$$E\left(d(y;\theta)d'(y;\theta)\right) = (1-\rho)^2 E\begin{bmatrix} 1 & f\left(\bar{y}_{i.}\right) \\ f\left(\bar{y}_{i.}\right) & f^2\left(\bar{y}_{i.}\right) \end{bmatrix} \approx (1-\rho)^2 \begin{bmatrix} 1 & \frac{1}{2} - \frac{\rho\left(\rho + n_i A(1-\rho)\right)}{n_i A(1-\rho)} \\ \frac{1}{2} - \frac{\rho\left(\rho + n_i A(1-\rho)\right)}{n_i A(1-\rho)} & \frac{1}{4} + \rho + n_i A(1-\rho) \end{bmatrix}$$

(A.84)

A.12

We derive $f(\boldsymbol{w}\mid y)$ as follows:

$$\begin{aligned}
f(\boldsymbol{w}\mid y) &\propto \int_0^{\infty}\int_{-\infty}^{\infty}\prod_{i=1}^{m}\left[f(y_i\mid w_i,\beta_0,\sigma^2)f(w_i\mid\beta_0,\sigma^2)\right]f(\beta_0,\sigma^2)d\beta_0 d\sigma^2 \\
&=\int_0^{\infty}\int_{-\infty}^{\infty}\prod_{i=1}^{m}\left[e^{-n_i w_i}w_i^{y_{i.}}\times\frac{1}{w_i\sqrt{2\pi}\sigma}e^{-\frac{(\log w_i-\beta_0)^2}{2\sigma^2}}\right]f(\beta_0,\sigma^2)d\beta_0 d\sigma^2 \\
&\propto \int_0^{\infty}\int_{-\infty}^{\infty}e^{-n_i\sum_{i=1}^{m}w_i}\prod_{i=1}^{m}w_i^{y_{i.}-1}\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\log w_i-\beta_0)^2}f(\beta_0,\sigma^2)d\beta_0 d\sigma^2 \\
&=e^{-n_i\sum_{i=1}^{m}w_i}\prod_{i=1}^{m}w_i^{y_{i.}-1}\int_0^{\infty}\int_{-\infty}^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\log w_i-\beta_0)^2}f(\beta_0,\sigma^2)d\beta_0 d\sigma^2
\end{aligned}$$

A.13

We derive $f(\boldsymbol{w}\mid y)$ by assigning the prior distributions as follows:

$$\begin{aligned}
f(\boldsymbol{w}\mid y) &\propto e^{-n_i\sum_{i=1}^{m}w_i}\prod_{i=1}^{m}w_i^{y_{i.}-1}\int_0^{\infty}\int_{-\infty}^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\log w_i-\beta_0)^2}\left(\sigma^2\right)^{-\alpha-1}e^{-\frac{\beta}{\sigma^2}}d\beta_0 d\sigma^2 \\
&=e^{-n_i\sum_{i=1}^{m}w_i}\prod_{i=1}^{m}w_i^{y_{i.}-1}\int_0^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}-\alpha-1}e^{-\frac{\beta}{\sigma^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\log w_i-\beta_0)^2}d\beta_0 d\sigma^2 \qquad \text{(A.85)} \\
&=e^{-n_i\sum_{i=1}^{m}w_i}\prod_{i=1}^{m}w_i^{y_{i.}-1}\int_0^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}-\alpha-1}e^{-\frac{\beta}{\sigma^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\log w_i-\overline{\log w}+\overline{\log w}-\beta_0)^2}d\beta_0 d\sigma^2
\end{aligned}$$

Note that:

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\log w_i - \overline{\log w} + \overline{\log w} - \beta_0\right)^2} d\beta_0 = \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2 + \sum_{i=1}^{m}\left(\overline{\log w} - \beta_0\right)^2\right]} d\beta_0$$

$$= \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2 + m\left(\overline{\log w} - \beta_0\right)^2\right]} d\beta_0 \tag{A.86}$$

$$= e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2\right]} \times \sqrt{\frac{2\pi\sigma^2}{m}}$$

Thus, define $ms^2 = \sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2$ and we have the following result:

$$f(\boldsymbol{w} \mid y) \propto e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \int_0^{\infty} \left(\sigma^2\right)^{-\frac{m}{2}-\alpha-1} e^{-\frac{\beta}{\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2\right]} \times \sqrt{\frac{2\pi\sigma^2}{m}} d\sigma^2$$

$$\propto e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \int_0^{\infty} \left(\sigma^2\right)^{-\frac{m}{2}-\alpha-1} e^{-\frac{\beta}{\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2\right]} \times \left(\sigma^2\right)^{\frac{1}{2}} d\sigma^2$$

$$= e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \int_0^{\infty} \left(\sigma^2\right)^{-\frac{m}{2}-\alpha-\frac{1}{2}} e^{-\frac{\beta}{\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log w_i - \overline{\log w}\right)^2\right]} d\sigma^2$$

$$= e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \int_0^{\infty} \left(\sigma^2\right)^{-\frac{m}{2}-\alpha-\frac{1}{2}} e^{-\frac{2\beta+ms^2}{2\sigma^2}} d\sigma^2$$

$$= e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \times \frac{\Gamma\left(\alpha + \frac{m-1}{2}\right)}{\left(\frac{ms^2+2\beta}{2}\right)^{\alpha+\frac{m-1}{2}}}$$

$$\propto e^{-n_i \sum_{i=1}^{m} w_i} \prod_{i=1}^{m} w_i^{y_{i.}-1} \left(\frac{ms^2+2\beta}{2}\right)^{-\left(\alpha+\frac{m-1}{2}\right)}$$

(A.87)

A.14

Let $w=\beta_0+s_i$. According to (3.1), the four elements, $\mu_{g^{-1}(w)}$, $\mathrm{V}_{g^{-1}(w),y_i}$, V_{y_i,y_i}, and μ_{y_i} in $\mathrm{BLP}(g^{-1}(w))$, are derived as follows:

$$\mu_{g^{-1}(w)}=E(g^{-1}(w))=E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)=T_1 \tag{A.88}$$

$$\begin{aligned}\mathrm{V}_{g^{-1}(w),y_i}&=Cov(g^{-1}(w),y_{i1}\cdots y_{in_i})\\&=Cov(g^{-1}(w),y_{i1})\mathbf{1}\\&=\left(EE(g^{-1}(w)y_{i1}\mid s_i)-E(g^{-1}(w))EE(y_{i1}\mid s_i)\right)\mathbf{1}'\\&=\left(E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)^2-\left(E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)\right)^2\right)\mathbf{1}'=\left(T_2-T_1^2\right)\mathbf{1}'\end{aligned} \tag{A.89}$$

.

Where $E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)$ and $E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)^2$ are calculated numerically and denoted as T_1 and T_2 respectively.

For the diagonal elements in V_{y_i,y_i}, we have:

$$\begin{aligned}Var(y_{ij})&=VarE(y_{ij}\mid s_i)+EVar(y_{ij}\mid s_i)\\&=Var\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)+E\left[\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)\left(1-\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)\right]\\&=E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)^2-\left(E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)\right)^2+E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)-E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)^2\\&=E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)-\left(E\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)\right)^2\\&=T_1-T_1^2=T_1\left(1-T_1\right)\end{aligned}$$

For the off diagonal elements in V_{y_i,y_i}, we have:

$$
\begin{aligned}
Cov(y_{ij}, y_{ij'}) &= EE(y_{ij} y_{ij'} \mid s_i) - EE(y_{ij} \mid s_i) EE(y_{ij'} \mid s_i) \\
&= E\left[E(y_{ij} \mid s_i) E(y_{ij'} \mid s_i) \right] - \left(E\left(\frac{e^{\beta_0 + s_i}}{1 + e^{\beta_0 + s_i}} \right) \right)^2 \\
&= E\left(\frac{e^{\beta_0 + s_i}}{1 + e^{\beta_0 + s_i}} \right)^2 - \left(E\left(\frac{e^{\beta_0 + s_i}}{1 + e^{\beta_0 + s_i}} \right) \right)^2 \\
&= T_2 - T_1^2
\end{aligned}
$$

Therefore, define $A_1 = T_1 - T_2$ and $B_1 = T_2 - T_1^2$, we have:

$$
\mathrm{V}_{y_i, y_i} = A_1 \boldsymbol{I} + B_1 \boldsymbol{J}
$$

Thus, the inverse matrix of $\mathrm{V}_{\boldsymbol{y}_i, \boldsymbol{y}_i}$ can be expressed as follows:

$$
\mathrm{V}_{y_i, y_i}^{-1} = \frac{1}{A_1} \left(\boldsymbol{I} - \frac{B_1}{A_1 + n_i B_1} \boldsymbol{J} \right) \tag{A.90}
$$

Now,

$$
\begin{aligned}
\mu_{\boldsymbol{y}_i} &= \left[EE\left(y_{ij} \mid s_i \right) \right] \mathbf{1} \\
&= E\left[\frac{e^{\beta_0 + s_i}}{1 + e^{\beta_0 + s_i}} \right] \mathbf{1} = T_1 \mathbf{1}
\end{aligned} \tag{A.91}
$$

Thus, using (A.88) through (A.91), $\mathrm{BLP}(g^{-1}(w))$ has been derived as follows:

$$
\mathrm{BLP}(g^{-1}(w)) = T_1 + \frac{\left(T_2 - T_1^2 \right)}{\left(T_1 - T_2 \right) + n_i \left(T_2 - T_1^2 \right)} \left(y_{i.} - n_i T_1 \right) \tag{A.92}
$$

So, $\mathrm{g}\left(\mathrm{BLP}(g^{-1}(v)) \right)$ is derived by letting g equal to the Logit function.

A.15

Based on A.14, we have: $\boldsymbol{L}_2 = \left(T_1 - nT_1 B_1 / (A_1 + nB_1) \quad B_1 / (A_1 + nB_1) \mathbf{1}' \quad -1 \right)$ and derive

$E(\boldsymbol{T}) = E\left(g'(g^{-1}(w)) \quad g'(g^{-1}(w))\left[E\left(y_i \middle| s_i\right)\right]^T \quad g'(g^{-1}(w))g^{-1}(w) \right)'$ as follows:

The next step is to evaluate $M_1(\theta)$ in (3.4) as follows:

Note that $A = e^{\beta_0 + \frac{\sigma^2}{2}}$ and $B = e^{2\beta_0 + \sigma^2}(e^{\sigma^2} - 1)$.

Then, we derive the following three elements as follows:

$$E\left[g'(g^{-1}(w))\right] = E\left[\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}\right] = E\left[e^{-\beta_0-s_i} + 2 + e^{\beta_0+s_i}\right] = 1/A + B/A^3 + 2 + A$$

(A.93)

$$E\left\{g'(g^{-1}(w))\left[E\left(y_i \middle| s_i\right)\right]\right\} = E\left\{\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}\left[\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\mathbf{1}\right]\right\} = E\left\{\left(1+e^{\beta_0+s_i}\right)\mathbf{1}\right\} = \left(1+e^{\beta_0+\frac{\sigma^2}{2}}\right)\mathbf{1} = (1+A)\mathbf{1}$$

(A.94)

$$E\left\{g'(g^{-1}(w))g^{-1}(w)\right\} = E\left\{\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right\} = E\left\{1+e^{\beta_0+s_i}\right\} = 1+e^{\beta_0+\frac{\sigma^2}{2}} = 1+A$$

(A.95)

From (A.2), we derive the following six elements as follows:

$$Var\left\{g'(g^{-1}(w))\right\} = Var\left\{\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}\right\} = \left(e^{-2\beta_0+\sigma^2} + e^{2\beta_0+\sigma^2} - 2\right)\left(e^{\sigma^2} - 1\right) = 2B\left(1/A + B/A^3 - 1\right)/A^2 + B$$

(A.96)

$$Var\left\{g'(g^{-1}(w))E\left(y_i \middle| s_i\right)\right\} = Var\left\{1+e^{\beta_0+s_i}\right\}\boldsymbol{J} = Var\left\{1+e^{\beta_0+s_i}\right\}\boldsymbol{J} = e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right)\boldsymbol{J} = B\boldsymbol{J}$$

(A.97)

$$Var\left\{g'(g^{-1}(w))g^{-1}(w)\right\} = Var\left\{1+e^{\beta_0+s_i}\right\} = Var\left\{e^{\beta_0+s_i}\right\} = e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right) = B$$

(A.98)

$$Cov\left(g'(g^{-1}(w)), g'(g^{-1}(w))E\left(y_i|s_i\right)\right) = Cov\left(\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}, \left(1+e^{\beta_0+s_i}\right)\mathbf{1}\right)$$
$$= \left(1-e^{\sigma^2}+e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right)\right)\mathbf{1} = \left(-B/A^2+B\right)\mathbf{1}$$

(A.99)

$$Cov\left(g'(g^{-1}(w)), g'(g^{-1}(w))g^{-1}(w)\right) = Cov\left(\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}, 1+e^{\beta_0+s_i}\right) \qquad \text{(A.100)}$$
$$= -B/A^2+B$$

$$\mathrm{Cov}\left(g'(g^{-1}(w))E\left(y_i|s_i\right), g'(g^{-1}(w))g^{-1}(w)\right) = Cov\left(\left(1+e^{\beta_0+s_i}\right)\mathbf{1}, 1+e^{\beta_0+s_i}\right) = e^{2\beta_0+\sigma^2}\left(e^{\sigma^2}-1\right)\mathbf{1} = B\mathbf{1}$$

(A.101)

In (A.3), we derive the following element:

$$E\left\{\left(g'(g^{-1}(w))\right)^2 Var\left(y_i|s_i\right)\right\} = E\left\{\left(\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}\right)^2 \begin{pmatrix} \frac{e^{\beta_0+s_i}}{\left(1+e^{\beta_0+s_i}\right)^2} & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 0 \\ & & & \frac{e^{\beta_0+s_i}}{\left(1+e^{\beta_0+s_i}\right)^2} \end{pmatrix}\right\} = E\left(\frac{\left(1+e^{\beta_0+s_i}\right)^2}{e^{\beta_0+s_i}}\right)\boldsymbol{I}$$
$$= \left(e^{-\beta_0+\frac{\sigma^2}{2}}+e^{\beta_0+\frac{\sigma^2}{2}}+2\right)\boldsymbol{I} = \left(\left(A^2+B\right)/A^3+A+2\right)\boldsymbol{I}$$

(A.102)

Thus, we can evaluate the value of $M_1(\theta)$ using (A.93) to (A.102).

A.16

Then, we can derive $M_2(\theta)$ as follows:

From (A.92), we know:

$$\text{BLP}(g^{-1}(w)) = T_1 + \frac{\left(T_2 - T_1^2\right)}{\left(T_1 - T_2\right) + n_i\left(T_2 - T_1^2\right)}\left(y_{i.} - n_i T_1\right) \tag{A.103}$$

We can parameterize (A.103) as follows:

$$\begin{aligned}\text{BLP}(g^{-1}(w)) &= T_1 + \frac{\left(T_2 - T_1^2\right)}{\left(T_1 - T_2\right) + n_i\left(T_2 - T_1^2\right)}\left(y_{i.} - n_i T_1\right) \\ &= \left(1-\rho\right)T_1 + \rho\bar{y}_{i.}\end{aligned} \tag{A.104}$$

Thus $d(y;\theta) = \left(\frac{\partial g[\text{BLP}(g^{-1}(\beta_0 + s_i))]}{\partial \beta_0}, \frac{\partial g[\text{BLP}(g^{-1}(\beta_0 + s_i))]}{\partial \sigma^2}\right)'$ is derived by using

$\frac{\partial A_1}{\partial \beta_0} = T_1 - T_2 - 2T_3$, $\frac{\partial B_1}{\partial \beta_0} = 2T_3 - 2T_1\left(T_1 - T_2\right)$, $\frac{\partial T_1}{\partial \beta_0} = T_1 - T_2$, $\frac{\partial T_2}{\partial \beta_0} = 2T_3$,

$\frac{\partial \rho}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\frac{B_1}{A_1 + n_i B_1} = \frac{\left(\frac{\partial}{\partial \beta_0}B_1\right)(A_1 + n_i B_1) - B_1\left(\frac{\partial}{\partial \beta_0}A_1 + n_i\frac{\partial}{\partial \beta_0}B_1\right)}{(A_1 + n_i B_1)^2}$, $\frac{\partial A_1}{\partial \sigma^2} = T_4 - T_5$,

$\frac{\partial B_1}{\partial \sigma^2} = T_5 - 2T_1T_4$, $\frac{\partial T_1}{\partial \sigma^2} = T_4$, $\frac{\partial T_2}{\partial \sigma^2} = T_5$ and

$\frac{\partial \rho}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\frac{B_1}{A_1 + n_i B_1} = \frac{\left(\frac{\partial}{\partial \sigma^2}B_1\right)(A_1 + n_i B_1) - B_1\left(\frac{\partial}{\partial \sigma^2}A_1 + n_i\frac{\partial}{\partial \sigma^2}B_1\right)}{(A_1 + n_i B_1)^2}$ as follows:

$$\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial \beta_0}=\frac{\partial\left\{\log\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]-\log\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]\right\}}{\partial \beta_0}$$

$$=\frac{\frac{\partial(1-\rho)}{\partial\beta_0}T_1+(1-\rho)\frac{\partial T_1}{\partial\beta_0}+\frac{\partial\rho}{\partial\beta_0}\bar{y}_{i.}}{\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]}$$

$$=\frac{c_1(\bar{y}_{i.}-T_1)+c_2}{\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]}$$

(A.105)

Where $c_1=\frac{\partial\rho}{\partial\beta_0}$ and $c_2=(1-\rho)\frac{\partial T_1}{\partial\beta_0}$.

And

$$\frac{\partial g[\mathrm{BLP}(g^{-1}(\beta_0+s_i))]}{\partial \sigma^2}=\frac{\partial\left\{\log\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]-\log\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]\right\}}{\partial \sigma^2}$$

$$=\frac{\frac{\partial(1-\rho)}{\partial\sigma^2}T_1+(1-\rho)\frac{\partial T_1}{\partial\sigma^2}+\frac{\partial\rho}{\partial\sigma^2}\bar{y}_{i.}}{\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]}$$

$$=\frac{c_3(\bar{y}_{i.}-T_1)+c_4}{\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]}$$

(A.106)

Where $c_3=\frac{\partial\rho}{\partial\sigma^2}$ and $c_4=(1-\rho)\frac{\partial T_1}{\partial\sigma^2}$.

Based on (A.105) and (A.106), the term, $E\left(d(y;\theta)d'(y;\theta)\right)$, can be derived as follows:

$$E\left(d(y;\theta)d'(y;\theta)\right)=E\begin{bmatrix} f_{11}(\bar{y}_{i.}) & f_{12}(\bar{y}_{i.}) \\ f_{21}(\bar{y}_{i.}) & f_{22}(\bar{y}_{i.}) \end{bmatrix} \quad \text{(A.107)}$$

Where

$$f_{11}(\bar{y}_{i.})=\frac{(c_1(\bar{y}_{i.}-T_1)+c_2)^2}{\left\{\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[(1-\rho)T_1+\rho\bar{y}_{i.}\right]\right]\right\}^2},$$

$$f_{12}\left(\bar{y}_{i.}\right)=\frac{(c_1(\bar{y}_{i.}-T_1)+c_2)(c_3(\bar{y}_{i.}-T_1)+c_4)}{\left\{\left[\left(1-\rho\right)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[\left(1-\rho\right)T_1+\rho\bar{y}_{i.}\right]\right]\right\}^2}\quad\text{and}$$

$$f_{22}\left(\bar{y}_{i.}\right)=\frac{(c_3(\bar{y}_{i.}-T_1)+c_4)^2}{\left\{\left[\left(1-\rho\right)T_1+\rho\bar{y}_{i.}\right]\left[1-\left[\left(1-\rho\right)T_1+\rho\bar{y}_{i.}\right]\right]\right\}^2}\quad\text{and}$$

In (A.107), we approximate $E(f_{11}(\bar{y}_{i.}))$, $E(f_{12}(\bar{y}_{i.}))$ and $E(f_{22}(\bar{y}_{i.}))$ by considering a second Taylor expansion of $f_{11}(\bar{y}_{i.})$, $f_{12}(\bar{y}_{i.})$ and $f_{22}(\bar{y}_{i.})$ around $E(\bar{y}_{i.})=T_1$. Then we have: $E(f_{11}(\bar{y}_{i.}))\approx f_{11}(T_1)+f_{11}''(T_1)Var[\bar{y}_{i.}]/2$ where $f_{11}(T_1)=\dfrac{c_2^2}{\left(T_1(1-T_1)\right)^2}$ and

$$\begin{aligned}f_{11}''(T_1)&=\frac{4c_2^2}{[T_1(1-T_1)]^2}\left[\frac{c_1}{c_2}-\frac{\rho}{T_1}+\frac{\rho}{1-T_1}\right]^2-\frac{2c_2^2}{[T_1(1-T_1)]^2}\left[\frac{c_1^2}{c_2^2}-\frac{\rho^2}{T_1^2}-\frac{\rho^2}{(1-T_1)^2}\right]\\&=\frac{2c_2^2}{[T_1(1-T_1)]^2}\left[2\left[\frac{c_1}{c_2}-\frac{\rho}{T_1}+\frac{\rho}{1-T_1}\right]^2-\left[\frac{c_1^2}{c_2^2}-\frac{\rho^2}{T_1^2}-\frac{\rho^2}{(1-T_1)^2}\right]\right]\end{aligned}$$

and

$$\begin{aligned}Var\left[\bar{y}_{i.}\right]&=VarE\left[\bar{y}_{i.}\mid s_i\right]+EVar\left[\bar{y}_{i.}\mid s_i\right]\\&=Var\left(\frac{e^{\beta_0+s_i}}{1+e^{\beta_0+s_i}}\right)+\frac{1}{n_i}E\left(\frac{e^{\beta_0+s_i}}{(1+e^{\beta_0+s_i})^2}\right)\\&=T_2-T_1^2+\frac{1}{n_i}\left(T_1-T_2\right)\end{aligned}\qquad\text{(A.108)}$$

And $E(f_{12}(\bar{y}_{i.}))\approx f_{12}(T_1)+f_{12}''(T_1)Var[\bar{y}_{i.}]/2$ where $f_{12}(T_1)=\dfrac{c_2c_4}{\left(T_1(1-T_1)\right)^2}$ and

$$f_{12}''(T_1)=\frac{c_2c_4}{[T_1(1-T_1)]^2}\left[\left[\frac{c_1}{c_2}+\frac{c_3}{c_4}-\frac{2\rho}{T_1}+\frac{2\rho}{1-T_1}\right]^2+\left[-\frac{c_1^2}{c_2^2}-\frac{c_3^2}{c_4^2}+\frac{2\rho^2}{T_1^2}+\frac{2\rho^2}{(1-T_1)^2}\right]\right]$$

and $E(f_{22}(\bar{y}_{i.}))\approx f_{22}(T_1)+f_{22}''(T_1)Var[\bar{y}_{i.}]/2$ where

$$f_{22}(T_1)=\frac{c_4^2}{\left(T_1(1-T_1)\right)^2}$$

$$f_{22}''(T_1)=\frac{2c_4^2}{[T_1(1-T_1)]^2}\left[2\left[\frac{c_3}{c_4}-\frac{\rho}{T_1}+\frac{\rho}{1-T_1}\right]^2-\left[\frac{c_3^2}{c_4^2}-\frac{\rho^2}{T_1^2}-\frac{\rho^2}{(1-T_1)^2}\right]\right]$$

Thus, we can approximate $E\left(d(y;\theta)d'(y;\theta)\right)$ using the result of approximation of

$E(f_{11}(\bar{y}_{i.}))$, $E(f_{12}(\bar{y}_{i.}))$ and $E(f_{22}(\bar{y}_{i.}))$.

A17

We derive $f(\boldsymbol{w}\,|\,y)$ as follows:

$$f(\boldsymbol{w}\,|\,y)\propto\int_0^\infty\int_{-\infty}^\infty\prod_{i=1}^m\left[f(y_i\,|\,w_i,\beta_0,\sigma^2)f(w_i\,|\,\beta_0,\sigma^2)\right]f(\beta_0,\sigma^2)d\beta_0 d\sigma^2$$

$$=\int_0^\infty\int_{-\infty}^\infty\prod_{i=1}^m\left[w_i^{y_{i.}}\left(1-w_i\right)^{n_i-y_{i.}}\times\frac{1}{w_i(1-w_i)\sqrt{2\pi}\sigma}e^{-\frac{\left(\log\left(\frac{w_i}{1-w_i}\right)-\beta_0\right)^2}{2\sigma^2}}\right]f(\beta_0,\sigma^2)d\beta_0 d\sigma^2$$

$$\propto\int_0^\infty\int_{-\infty}^\infty\prod_{i=1}^m w_i^{y_{i.}-1}\prod_{i=1}^m\left(1-w_i\right)^{n_i-y_{i.}-1}\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^m\left(\log\left(\frac{w_i}{1-w_i}\right)-\beta_0\right)^2}f(\beta_0,\sigma^2)d\beta_0 d\sigma^2$$

$$=\prod_{i=1}^m w_i^{y_{i.}-1}\prod_{i=1}^m\left(1-w_i\right)^{n_i-y_{i.}-1}\int_0^\infty\int_{-\infty}^\infty\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^m\left(\log\left(\frac{w_i}{1-w_i}\right)-\beta_0\right)^2}f(\beta_0,\sigma^2)d\beta_0 d\sigma^2$$

A.18

We derive $f(\boldsymbol{w}\,|\,y)$ by assigning the prior distributions as follows:

We assign the non-informative prior to $f(\beta_0)$ and the inverse gamma distribution to

$f(\sigma^2)$ in (4.10). Thus, define $\overline{\log\left(\frac{w}{1-w}\right)}=\sum_{i=1}^m\frac{\log\left(\frac{w_i}{1-w_i}\right)}{m}$ and we have the following

results:

$$f(\boldsymbol{w}\mid y)\propto\prod_{i=1}^{m}w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n-y_{i.}-1}\int_0^{\infty}\int_{-\infty}^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\beta_0\right)^2}\left(\sigma^2\right)^{-\alpha-1}e^{-\frac{\beta}{\sigma^2}}d\beta_0 d\sigma^2$$

$$=\prod_{i=1}^{m}w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\int_0^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}-\alpha-1}e^{-\frac{\beta}{\sigma^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\beta_0\right)^2}d\beta_0 d\sigma^2$$

$$=\prod_{i=1}^{m}w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\int_0^{\infty}\left(\sigma^2\right)^{-\frac{m}{2}-\alpha-1}e^{-\frac{\beta}{\sigma^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}+\overline{\log\left(\frac{w}{1-w}\right)}-\beta_0\right)^2}d\beta_0 d\sigma^2$$

(A.109)

Note that:

$$\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}+\overline{\log\left(\frac{w}{1-w}\right)}-\beta_0\right)^2}d\beta_0=\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2+\sum_{i=1}^{m}\left(\overline{\log\left(\frac{w}{1-w}\right)}-\beta_0\right)^2\right]}d\beta_0$$

$$=\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2+m\left(\overline{\log\left(\frac{w}{1-w}\right)}-\beta_0\right)^2\right]}d\beta_0$$

$$=e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2\right]}\times\sqrt{\frac{2\pi\sigma^2}{m}}$$

(A.110)

Thus, define $ms^2=\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2$ and we have the following result:

$$
\begin{aligned}
f(\boldsymbol{w}\mid y) &\propto \prod_{i=1}^{m} w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\int_0^{\infty}(\sigma^2)^{-\frac{m}{2}-\alpha-1} e^{-\frac{\beta}{\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2\right]} \times\sqrt{\frac{2\pi\sigma^2}{m}}d\sigma^2 \\
&\propto \prod_{i=1}^{m} w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\int_0^{\infty}(\sigma^2)^{-\frac{m}{2}-\alpha-1} e^{-\frac{\beta}{\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2\right]} \times(\sigma^2)^{\frac{1}{2}}d\sigma^2 \\
&= \prod_{i=1}^{m} w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\int_0^{\infty}(\sigma^2)^{-\frac{m}{2}-\alpha-\frac{1}{2}} e^{-\frac{\beta}{\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}\left(\log\left(\frac{w_i}{1-w_i}\right)-\overline{\log\left(\frac{w}{1-w}\right)}\right)^2\right]} d\sigma^2 \\
&= \prod_{i=1}^{m} w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\int_0^{\infty}(\sigma^2)^{-\frac{m}{2}-\alpha-\frac{1}{2}} e^{-\frac{2\beta+ms^2}{2\sigma^2}} d\sigma^2 \\
&= \prod_{i=1}^{m} w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\times\frac{\Gamma\left(\alpha+\frac{m-1}{2}\right)}{\left(\frac{ms^2+2\beta}{2}\right)^{\alpha+\frac{m-1}{2}}} \\
&\propto \prod_{i=1}^{m} w_i^{y_{i.}-1}\prod_{i=1}^{m}(1-w_i)^{n_i-y_{i.}-1}\left(\frac{ms^2+2\beta}{2}\right)^{-\left(\alpha+\frac{m-1}{2}\right)}
\end{aligned}
$$

(A.111)

A.19

A conjugate GLMM is presented as follows:

$$
\begin{aligned}
& y_{ij}\mid\mu_{ij}\sim \text{Poi}(\mu_{ij}) \qquad i=1,\cdots,m\,,\ \ j=1,\cdots,n \\
& \mu_{ij}=s_i \\
& s_i\sim\exp(\lambda) \\
& \log\lambda = x_{ij}'\beta
\end{aligned}
$$

For our example, we take $\log\lambda=\beta_0$. Therefore, we know $s_i\sim\exp(e^{\beta_0})$ and $y_{i.}\mid s_i\sim\text{Poi}(ns_i)$. Thus we have the following result:

$$
\begin{aligned}
P(y_{i.}=y) &= \int_0^\infty \frac{e^{-ns_i}(ns_i)^y}{y!}\, e^{\beta_0} e^{-e^{\beta_0}s_i}\, ds_i \\
&= \frac{e^{\beta_0}n^y}{y!}\int_0^\infty e^{-(n+e^{\beta_0})s_i} s_i^{\,y}\, ds_i \\
&= \frac{e^{\beta_0}n^y}{y!}\cdot\frac{\Gamma(y+1)}{(n+e^{\beta_0})^{y+1}} \\
&= \left(\frac{n}{n+e^{\beta_0}}\right)^y \left(\frac{e^{\beta_0}}{n+e^{\beta_0}}\right) \\
&= \left(\frac{\frac{n}{e^{\beta_0}}}{\frac{n}{e^{\beta_0}}+1}\right)^y \left(\frac{1}{\frac{n}{e^{\beta_0}}+1}\right) \\
&\equiv \text{NegBin}\left[\frac{n}{e^{\beta_0}},1\right]
\end{aligned}
$$

Also,

$$
\begin{aligned}
f(s_i \mid y_{i.}) &\propto f(y_{i.} \mid s_i)\, f(s_i) \\
&= \frac{e^{-ns_i}(ns_i)^{y_{i.}}}{y_{i.}!}\, e^{\beta_0} e^{-e^{\beta_0}s_i} \\
&\propto (s_i)^{y_{i.}} e^{-(n+e^{\beta_0})s_i} \\
&\equiv \text{Gamma}(n+e^{\beta_0},\, y_{i.}+1)
\end{aligned}
$$

So for $w = s_i$, we denote the BP as $\eta(y;\beta_0) = \dfrac{y_{i.}+1}{n+e^{\beta_0}}$ and derive the likelihood function and the parameter estimator as follows:

$$L(\beta_0) \propto \left(\frac{n}{n+e^{\beta_0}}\right)^{y_{..}} \left(\frac{e^{\beta_0}}{n+e^{\beta_0}}\right)^{m}$$

$$\log L \propto -y_{..} \log(n+e^{\beta_0}) + m\beta_0 - m\log(n+e^{\beta_0})$$

$$\frac{\partial \log L}{\partial \beta_0} = -\frac{(y_{..}+m)e^{\beta_0}}{n+e^{\beta_0}} + m = 0$$

$$\Rightarrow \frac{(y_{..}+m)e^{\hat{\beta}_0}}{n+e^{\hat{\beta}_0}} = m$$

$$\Rightarrow \frac{e^{\hat{\beta}_0}}{n+e^{\hat{\beta}_0}} = \frac{m}{y_{..}+m} \Rightarrow y_{..}e^{\hat{\beta}_0} + me^{\hat{\beta}_0} = mn + me^{\hat{\beta}_0}$$

$$\Rightarrow e^{\hat{\beta}_0} = \frac{mn}{y_{..}} = \frac{1}{\bar{y}}$$

We can also derive the Fisher Information matrix as follows:

$$\frac{\partial^2 \log L(\beta_0)}{\partial^2 \beta_0} = -(y_{..}+m)\left\{\frac{(n+e^{\beta_0})e^{\beta_0} - e^{\beta_0}e^{\beta_0}}{(n+e^{\beta_0})^2}\right\}$$

$$= -(y_{..}+m)\,\frac{ne^{\beta_0} + e^{2\beta_0} - e^{2\beta_0}}{(n+e^{\beta_0})^2}$$

$$= -(y_{..}+m)\,\frac{ne^{\beta_0}}{(n+e^{\beta_0})^2}$$

$$E\left\{-\frac{\partial^2 \log L(\beta_0)}{\partial^2 \beta_0}\right\} = \frac{ne^{\beta_0}}{(n+e^{\beta_0})^2} E(y_{..}+m)$$

$$= \frac{ne^{\beta_0}}{(n+e^{\beta_0})^2}\left(\frac{mn}{e^{\beta_0}} + m\right)$$

$$= \frac{mne^{\beta_0}}{(n+e^{\beta_0})^2}\left(\frac{n}{e^{\beta_0}} + 1\right)$$

$$= \frac{mn}{n+e^{\beta_0}}$$

We denote the Fisher information matrix as $I(\beta_0) = \frac{mn}{n+e^{\beta_0}}$. Since we know $e^{\hat{\beta}_0} = \frac{1}{\bar{y}}$, we can replace β_0 with $\hat{\beta}_0$ in the eBP as follows:

$$\eta(y;\hat{\beta}_0) = \frac{y_{i.}+1}{n+e^{\hat{\beta}_0}} = \frac{y_{i.}+1}{n+\frac{1}{\bar{y}_{..}}}$$

Following Kackar and Harville (1984), we can approximate the MSE of $\eta(y;\hat{\beta}_0)$ as follows:

$$d(y;\beta_0) = \frac{\partial\eta(y;\beta_0)}{\partial\beta_0} = \frac{-(y_{i.}+1)e^{\beta_0}}{(n+e^{\beta_0})^2}$$

and

$$\begin{aligned}
A(\beta_0) &= E\left(\frac{(y_{i.}+1)^2 e^{2\beta_0}}{(n+e^{\beta_0})^4}\right) \\
&= \frac{e^{2\beta_0}}{(n+e^{\beta_0})^4} E(y_{i.}+1)^2 \\
&= \frac{e^{2\beta_0}}{(n+e^{\beta_0})^4}\left[E(y_{i.}^2+2y_{i.}+1)\right] \\
&= \frac{e^{2\beta_0}}{(n+e^{\beta_0})^4}\left(\frac{n}{e^{\beta_0}}+\frac{n^2}{e^{2\beta_0}}+\frac{n^2}{e^{2\beta_0}}+\frac{2n}{e^{\beta_0}}+1\right) \\
&= \frac{e^{2\beta_0}}{(n+e^{\beta_0})^4}\left(\frac{3ne^{\beta_0}+2n^2+e^{2\beta_0}}{e^{2\beta_0}}\right)
\end{aligned}$$

and

$$B(\beta_0) = \mathrm{I}^{-1}(\beta_0) = \frac{n+e^{\beta_0}}{mn}$$

Thus,

$$
\begin{aligned}
\mathrm{M}_1(\beta_0) &= E\left[s_i - E\left(s_i \mid y_{i.}\right)\right]^2 \\
&= E\left\{Var\left(s_i \mid y_{i.}\right)\right\} \\
&= E\left\{\frac{y_{i.}+1}{(n+e^{\beta_0})^2}\right\} \\
&= \frac{\frac{n}{e^{\beta_0}}+1}{(n+e^{\beta_0})^2} = \frac{n+e^{\beta_0}}{e^{\beta_0}(n+e^{\beta_0})^2} = \frac{1}{e^{\beta_0}(n+e^{\beta_0})}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{M}_2(\beta_0) &\approx \mathrm{tr}(A(\beta_0)B(\beta_0)) \\
&= \frac{e^{2\beta_0}}{(n+e^{\beta_0})^4}\left(\frac{3ne^{\beta_0}+2n^2+e^{2\beta_0}}{e^{2\beta_0}}\right)\frac{n+e^{\beta_0}}{mn} \\
&= \frac{3ne^{\beta_0}+2n^2+e^{2\beta_0}}{mn(n+e^{\beta_0})^3}
\end{aligned}
$$

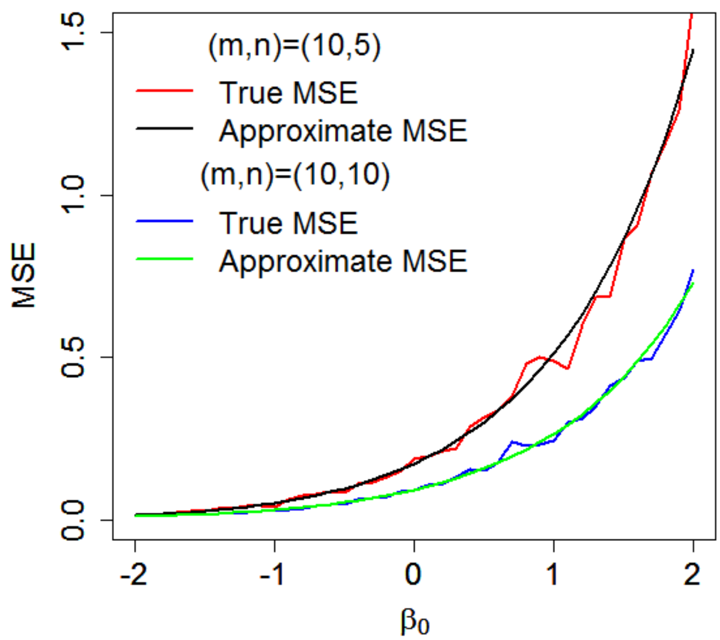
$$
\begin{aligned}
\dot{M}(\beta_0) &= \frac{1}{e^{\beta_0}(n+e^{\beta_0})} + \frac{3ne^{\beta_0}+2n^2+e^{2\beta_0}}{mn(n+e^{\beta_0})^3} \\
&= \frac{1}{(n+e^{\beta_0})}\left(\frac{1}{e^{\beta_0}} + \frac{3ne^{\beta_0}+2n^2+e^{2\beta_0}}{mn(n+e^{\beta_0})^2}\right)
\end{aligned}
$$

It is worth noting that $E\left(d(y;\beta_0)\right) \neq 0$, which is shown as follows:

$$
\begin{aligned}
E\left(d(y;\beta_0)\right) &= E\left(\frac{-(y_{i.}+1)e^{\beta_0}}{(n+e^{\beta_0})^2}\right) \\
&= -\frac{e^{\beta_0}}{(n+e^{\beta_0})^2}E(y_{i.}+1) \\
&= -\frac{e^{\beta_0}}{(n+e^{\beta_0})^2}\left(\frac{n}{e^{\beta_0}}+1\right) \\
&= \frac{-1}{n+e^{\beta_0}} \neq 0
\end{aligned}
$$

Thus, we can conclude that $E\left(d^2(y;\beta_0)\right) \neq Var\left(d(y;\beta_0)\right)$.

To check the performance of $\dot{M}(\beta_0)$ when approximating $M(\beta_0)$, we estimate the true MSE by Monte Carlo simulation. The following figure shows the results of the performance:



From the above figure, the performance of the MSE approximation, $\dot{M}(\beta_0)$ is very good since it is very close to the true MSE, $M(\beta_0)$ in both cases.

References

Abramowitz, M., Stegun, I., 1972. Handbook of Mathematical Functions, New York: Dover Publications.

Skrondal, A., Rabe-Hesketh, S., 2009. Prediction in multilevel generalized linear models. Journal of the Royal Statistical Society, Series B 172, 659–687.

Booth, J.G., Hobert, J.P., 1998. Standard errors of prediction in generalized linear mixed models. Journal of the American Statistical Association 93, 262-272.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, 9-25.

Calyton, D., Kaldor, J., 1987. Empiraical bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43, 671-681.

de Bruijn 1981. Asyptotic Methods in Analysis, New York: Dover.

Goldberger, A.S., 1962. Best linear unbiased prediction in the generalized regression model. Journal of the American Statistical Association 57, 369-375.

Golub, G.H., Welsch, J.H., 1969. Calculation of gaussian quadrature rules. Mathematical Computing 23, 221–230.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423-447.

Harville, D.A., 1985. Decomposition of the prediction error. Journal of the American Statistical Association 80, 132-138.

Harville, D.A., Jeske, D.R., 1992. Mean Squared Error of Estimation or Prediction Under a General Linear Model. Journal of the American Statistical Association 87, 724- 731.

Harville, D.A., 2008. Accounting for the estimation of variances and covariances in prediction under a general linear model: an overview. Tatra Mt. Math 39, 1–15.

Jeske, D.R., Harville, D.A., 1988. Prediction-interval procedures and (fixed effects) confidence-interval procedures for mixed linear models. Communications in Statistics - Theory and Methods 17, 1053-1087.

Kackar, R.N., Harville, D.A., 1984. Approximations for standard errors of estimators of fixed and random effect in mixed linear models. Journal of the American Statistical Association 79, 853-862.

Laird, N.M., Louis, T.A., 1982. Approximate posterior distributions for incomplete data problems. Journal of the Royal Statistical Society, Series B 44, 190-200.

Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46, 673-687.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., Schabenberger, O., 2006. SAS for Mixed Models, Second Edition, Cary, NC: SAS Institute Inc.

McCulloch, C.E., Searle, S.R., Neuhaus, J.M., Generalized, Linear and Mixed Models. 2nd edition. Wiley; New York: 2008.

Pinheiro, J.C., Chao, E.C., 2006. Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. Journal of Computational and Graphical Statistics 15, 58–81.

Wolfinger, R., O'Connell, M., 1993. Generalized linear mixed models: a pseudo-likelihood approach. Journal of Statistical Computation and Simulation 4, 233–243.

Yang, C., 2013. "Prediction Intervals in Generalized Linear Mixed Models". PhD Dissertation, Department of Applied Statistics, University of California, Riverside.

Zhao, Y., Staudenmayer J., Coull, B.A., Wand, M.P., 2006. General design bayesian generalized linear mixed models. Statistical Science 21, 35–51.