Appendix A

In this section, the asymptotic properties of QAICb1 and QAICb2 in equations (3.10) and (3.18) are established under linear mixed models. We will prove that the proposed selection criteria are asymptotically equivalent when the number of individuals $n \to \infty$. In addition, we will show that QAICb1 and QAICb2 are asymptotically unbiased estimators of the expected K-L discrepancy $d(\beta_0, \phi, k)$ in expression (3.2).

Let $\mathcal{B}$ be the parameter space of the unknown parameter β. Let $\bar{\beta}$ be the pseudo true parameter of β, which is the estimator of β obtained by solving the corresponding quasi-score equation. It should be pointed out that $\bar{\beta}$ is not necessarily a global minimizer and is assumed to be existing and unique.

A.1 Asymptotic Equivalence of QAICb1 and QAICb2

In this section, we will focus on proving the asymptotic equivalence of QAICb1 and QAICb2. It is the same to prove the asymptotic equivalence of two bias correction terms b1 and b2. Following Shang and Cavanaugh (2008), we will establish the consistency of the two estimators $\hat{\beta}$ and $\hat{\beta}^*$ from the original and bootstrap sample, respectively. We will show that as $n \to \infty$, we have

$$\hat{\beta} \to \bar{\beta} \text{ and } \hat{\beta}^* \to \bar{\beta}$$

based on **Assumptions 1** to **4**.
Assumption 1.

1. The parameter space $\mathcal{B}$ is a compact subset of k-dimensional Euclidean space.
2. The first, second, and third derivatives of the log quasi-likelihood with respect to β exist, and are continuous and bounded over $\mathcal{B}$.
3. $\bar{\beta}$ is an interior point of $\mathcal{B}$.

With **Assumption 1**, let $Q_i(y_i|\beta)$ and $Q_i(y_i|\tilde{\beta})$ be the two marginal densities for the ith individual given β and $\tilde{\beta}$ in a neighborhood U of $\mathcal{B}$. The log quasi-likelihood ratio is defined by

$$R_i(y_i, \beta, U) = \inf_{\tilde{\beta} \in U} \{Q_i(y_i|\beta) - Q_i(y_i|\tilde{\beta})\} = \inf_{\tilde{\beta} \in U} \frac{Q_i(y_i|\beta)}{Q_i(y_i|\tilde{\beta})}.$$

Assume that the following limit

$$\bar{R}(\bar{\beta}, U) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E_0\{R_i(y_i, \bar{\beta}, U)\}$$

exists and is finite in a neighborhood $U = U_\beta$ for any β in $\mathcal{B}$.

Note that the log quasi-likelihood function $Q(\beta; Y)$ is continuous with

$$\lim_{\tilde{\beta} \to \beta} Q(\tilde{\beta}; Y) = Q(\beta; Y)$$

for any $\tilde{\beta}$ in the space $\mathcal{B}$. Based on the Lebesgue monotone convergence theorem, we have the following limits

$$\lim_{q\to\infty} \bar{R}(\bar{\beta}, U_{\beta}^{(q)}) = \bar{R}(\bar{\beta}, \beta) = \lim_{n\to\infty} \frac{1}{n} E_0\{Q(\bar{\beta}; Y) - Q(\beta; Y)\}$$

hold true for a monotone decreasing sequence of neighborhoods $U_{\beta}^{(q)}$, $q = 1, 2, 3, ...,$ converging to a parameter β.

For a bootstrap sample $Y^* = (y_1^{*'}, ..., y_n^{*'})'$, we can similarly define

$$\bar{R}_B(\bar{\beta}, U) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_0 E_*\{R_i(y_i^*, \bar{\beta}, U)\}$$

and

$$\lim_{q\to\infty} \bar{R}(\bar{\beta}, U_{\beta}^{(q)}) = \bar{R}_B(\bar{\beta}, \beta) = \lim_{n\to\infty} \frac{1}{n} E_0 E_*\{Q(\bar{\beta}; Y^*) - Q(\beta; Y^*)\}.$$

Then we provide the next assumption:

Assumption 2.

1. $\frac{1}{n}\sum\limits_{i=1}^{n} R_i(y_i, \bar{\beta}, U_{\beta})$ converges almost surely to $\bar{R}(\bar{\beta}, U_{\beta})$ in a neighborhood U_{β} for any β in $\mathcal{B}$ and $\bar{R}(\bar{\beta}, U_{\beta}) > 0$ for β in $\mathcal{B}$, where $\beta \neq \bar{\beta}$.
2. $\frac{1}{n}\sum\limits_{i=1}^{n} R_i(y_i^*, \bar{\beta}, U_{\beta})$ converges almost surely to $\bar{R}_B(\bar{\beta}, U_{\beta})$ in a neighborhood U_{β} for any β in $\mathcal{B}$ and $\bar{R}_B(\bar{\beta}, U_{\beta}) > 0$ for β in $\mathcal{B}$, where $\beta \neq \bar{\beta}$.

Next, we will provide **Assumption 3** as well as the proof using the nonparametric and semiparemetric approaches under the linear mixed model. **Assumption 3** and its special case **Assumption 3b** lay the foundation of the proposed QAICb1 and QAICb2. We can extend **Assumption 3** and **Assumption 3b** to the generalized linear model with random effects.

Assumption 3. $E_*\{Q(\beta; Y^*)\} = Q(\beta; Y)$.

Proof. To prove the **Assumption 3**, we take advantage of the independent model structure by assuming the data are not correlated. Shang and Cavanaugh (2008) proved a assumption similar to **Assumption 3** under log likelihood and MLE estimators using parametric, semiparemetric, and nonparametric settings. As both an MLE estimator and a quasi-score-based estimator are obtained by solving the first derivative of the log likelihood functions, **Assumption 3** can be extended using quasi-likelihood functions. Moreover, the log quasi-likelihood and log likelihood functions take the same form other than a constant term arisen from the integral when the linear mixed model is used.

We will first show the proof of **Assumption 3** using the semiparametric approach.

Semiparametric bootstrap

Notations from models (2.1) and (2.2) will be used here. The semiparemetric bootstrap depends on resampling over the residuals obtained through the fitted model by

$$\hat{\xi}_i = y_i - X_i\hat{\beta}, \quad i = 1, ..., n,$$

where $\hat{\xi}_i$ is the vector of residuals corresponding to the ith individual. Let $\hat{\xi} = (\hat{\xi}_1, ..., \hat{\xi}_n)'$ and the equation of obtaining the residuals $\hat{\xi}$ in the matrix form is denoted by

$$\hat{\xi} = Y - X\hat{\beta}.$$

Then a bootstrap sample Y^* using the semiparametric bootstrap is achieved by

$$Y^* = X\hat{\beta} + \hat{\xi}^*, \tag{A1}$$

where $\hat{\xi}^*$, denoted by $(\hat{\xi}_1^*, ..., \hat{\xi}_n^*)'$, is selected with replacement based on the empirical distribution of the residuals $\hat{\xi}$.

Note that the overdispersion parameter is not included because it is set to be 1 when a linear mixed model is used.

Let V be the variance covariance matrix of Y which is independent of β. With respect to β, X and Y, the first derivative of the quasi-likelihood function under the independent model setting can be written by

$$\frac{\partial Q(\beta; Y)}{\partial \beta} = X'V^{-1}(Y - X\beta). \tag{A2}$$

Actually, the derivative in equation (A2) is the same as the quasi-score equation defined under the uncorrelated observations. By integrating equation (A2) and ignoring the constant, the log quasi-likelihood function of the original sample is calculated as

$$Q(\beta; X, Y) = \frac{1}{2}X'(Y - X\beta)'V^{-1}(Y - X\beta), \tag{A3}$$

and the log quasi-likelihood of the bootstrap sample Y^*, similar to equation (A3), under the semiparametric bootstrap setting can be written as

$$Q(\beta; X, Y^*) = \frac{1}{2}X'(Y^* - X\beta)'V^{-1}(Y^* - X\beta). \tag{A4}$$

By taking the expectation with respect to the bootstrap distribution over expression (A4) and applying equation (A1), we have

$$\begin{aligned}
E_*\{Q(\beta; X, Y^*)\} &= E_*\{\frac{1}{2}X'(Y^* - X\beta)'V^{-1}(Y^* - X\beta)\} \\
&= \frac{1}{2}X'E_*\{(Y^* - X\beta)'V^{-1}(Y^* - X\beta)\} \\
&= \frac{1}{2}X'E_*\{(X\hat{\beta} + \hat{\xi}^* - X\beta)'V^{-1}(X\hat{\beta} + \hat{\xi}^* - X\beta)\} \\
&= \frac{1}{2}X'E_*\{(Y - X\beta + \hat{\xi}^* - (Y - X\hat{\beta}))'V^{-1}(Y - X\beta + \hat{\xi}^* - (Y - X\hat{\beta}))\} \\
&= \frac{1}{2}X'E_*\{(\xi - (\hat{\xi} - \hat{\xi}^*))'V^{-1}(\xi - (\hat{\xi} - \hat{\xi}^*))\} \\
&= \frac{1}{2}X'E_*\{\xi'V^{-1}\xi - 2\xi'V^{-1}(\hat{\xi} - \hat{\xi}^*) + (\hat{\xi} - \hat{\xi}^*)'V^{-1}(\hat{\xi} - \hat{\xi}^*)\} \\
&= \frac{1}{2}X'\{\xi'V^{-1}\xi - 2\xi'V^{-1}E_*(\hat{\xi} - \hat{\xi}^*) + E_*||V^{-1}(\hat{\xi} - \hat{\xi}^*)||^2\}.
\end{aligned} \tag{A5}$$

As the distribution of $\hat{\xi}^*$ is the same as the empirical distribution of $\hat{\xi}$ and E_* can also be viewed as the expectation taken under the empirical distribution, then we have

$$E_*\{(\hat{\xi} - \hat{\xi}^*)\} = E_*(\hat{\xi}) - E_*(\hat{\xi}^*) = \frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_i - \frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_i = 0.$$

Thus, the expectation of equation (A5) can be simplified to

$$\begin{aligned} E_*Q(\beta; X, Y^*) &= \frac{1}{2}X'\{\xi' V^{-1}\xi\} \\ &= \frac{1}{2}X'\{(Y - X\beta)' V^{-1}(Y - X\beta)\} \\ &= Q(\beta; X, Y), \end{aligned}$$

which has validated the proof of **Assumption 3** under semi-parametric bootstrap.

Nonparametric bootstrap

When implementing the nonparametric bootstrap approach, we should select data points with replacement at the individual level when there exists correlation in the data. For the subject indexes $i = 1, ..., n$, let the original pairs of sample to be (X_i, y_i) with y_i denoting the response vector of the ith subject and X_i the corresponding covariate matrix. Then a bootstrap sample is represented as (X_i^*, y_i^*). According to the linear mixed model setting (2.1), let the vector $\xi_i = Z_i b_i + \epsilon_i$, $i = 1, ..., n$, then model (2.1) becomes

$$y_i = X_i\beta + \xi_i, \quad i = 1, ..., n,$$

and the associated bootstrap sample (X_i^*, y_i^*) can be expressed as

$$y_i^* = X_i^*\beta + \xi_i^*, \quad i = 1, ..., n.$$

With respect to the matrix notation (2.2), we have

$$\begin{aligned} Y &= X\beta + \xi \quad \text{and} \\ Y^* &= X^*\beta + \xi^*, \end{aligned} \tag{A6}$$

where ξ and ξ^* are $N \times 1$ vectors with $\xi = (\xi_1', ..., \xi_n')'$ and $\xi^* = (\xi_1^{*\prime}, ..., \xi_n^{*\prime})'$, respectively. We can see from expression (A6) that, similar to the semiparametric bootstrap, the bootstrap distribution of ξ^* is the same as the empirical distribution of $\hat{\xi}$. Let V^* be the variance covariance matrix of ξ^* and then V^* is a positive definite block diagonal matrix of n blocks with each block being $\frac{1}{n}\sum_{i=1}^{n}(\xi - \bar{\xi})(\xi - \bar{\xi})'$ with $\bar{\xi}$ to be the mean of ξ. As we have

$$\bar{\xi} = E(\xi) = E(Z_i b_i + \epsilon_i) = 0,$$

the block of the matrix V^* can be denoted by $\frac{1}{n}\sum_{i=1}^{n}\xi\xi'$.

Referring to the integrated log quasi-likelihood function in equation (A3), the corresponding log quasi-likelihood function of a bootstrap sample with respect to X^* and Y^* via nonparametric bootstrap can be represented as

$$Q(\beta; X^*Y^*) = \frac{1}{2}X'(Y^* - X^*\beta)' V^{-1}(Y^* - X^*\beta).$$

By plugging in the variance covariance V, we then have

$$\begin{aligned}
E_*\{Q(\beta; X^*, Y^*)\} &= \frac{1}{2}E_*\{X'(Y^* - X^*\beta)'V^{-1}(Y^* - X^*\beta)\} \\
&= \frac{1}{2}X'E_*||V^{-\frac{1}{2}}(Y^* - X^*\beta)||^2 \\
&= \frac{1}{2}X'E_*||V^{-\frac{1}{2}}(\xi^*)||^2 \\
&= \frac{1}{2}X'\{\text{trace}(V^{-1}V^*) + E_*(\xi^*)'V^{-1}E_*(\xi^*)\} \\
&= \frac{1}{2}X'\text{trace}(V^{-1}V^*) \\
&= \frac{1}{2}X'\xi'V^{-1}\xi \\
&= \frac{1}{2}X'||V^{-\frac{1}{2}}\xi||^2 \\
&= \frac{1}{2}X'||V^{-\frac{1}{2}}(Y - X\beta)||^2 \\
&= \frac{1}{2}X'(Y - X\beta)'V^{-1}(Y - X\beta) \\
&= Q(\beta; X, Y).
\end{aligned}$$

Thus, the **Assumption 3** holds true under nonparametric bootstrap framework.

To summary, we have showed that the **Assumption 3** is valid for any β of $\mathcal{B}$ under both the semiparametric and nonparametric bootstrap approaches. Specially, when $\hat{\beta}$ is an estimator of β by solving the quasi-score equations, we have the following assumption:

Assumption 3b. $E_*\{Q(\hat{\beta}; Y^*)\} = Q(\hat{\beta}; Y)$, where $\hat{\beta}$ is the estimator by solving the quasi-score equations.

The last assumption is provided to establish the asymptotic property of the second derivatives of the quasi-likelihood functions of both the original sample and bootstrap samples. The proof of this assumption is also given.

Assumption 4. Let $\hat{\varphi}(Y, \beta) = -\frac{\partial^2}{\partial\beta\partial\beta'}\, Q(\beta; Y)$ and $\hat{\varphi}(Y^*, \beta) = -\frac{\partial^2}{\partial\beta\partial\beta'}\, Q(\beta; Y^*)$ be the observed fisher information over the original sample Y and bootstrap sample Y^*, respectively. Then as $n \to \infty$,

$$\begin{aligned}
\hat{\varphi}(Y, \beta)/n &\longrightarrow \bar{\varphi}(\beta) \quad \text{a.s.} \quad \text{and} \\
\hat{\varphi}(Y^*, \beta)/n &\longrightarrow \bar{\varphi}_B(\beta) \quad \text{a.s.},
\end{aligned}$$

where $\bar{\varphi}(\beta)$ and $\bar{\varphi}_B(\beta)$ are two positive definite matrices. Moreover, $\bar{\varphi}(\bar{\beta}) = \bar{\varphi}_B(\bar{\beta})$.

Proof. From **Assumption 1**, we know that both quantities $\hat{\varphi}(Y, \beta)$ and $\hat{\varphi}(Y^*, \beta)$ exist and so do the corresponding limits $\hat{\varphi}(Y, \beta)/n$ and $\hat{\varphi}(Y^*, \beta)/n$. Let

$$\bar{\varphi}(\beta) = \lim_{n\to\infty} \frac{1}{n} E_0\{\hat{\varphi}(Y, \beta)\} \quad \text{and} \quad \bar{\varphi}_B(\beta) = \lim_{n\to\infty} \frac{1}{n} E_0 E_*\{\hat{\varphi}(Y^*, \beta)\}.$$

We claim that

$$\lim_{n\to\infty} \frac{1}{n}\hat{\varphi}(Y, \beta) = \bar{\varphi}(\beta) \quad \text{a.s.} \quad \text{and} \lim_{n\to\infty} \frac{1}{n}\hat{\varphi}_B(Y^*, \beta) = \bar{\varphi}(\beta) \quad \text{a.s.}$$

Next, we will show that $\bar{\varphi}(\beta) = \bar{\varphi}_B(\beta)$ under the semiparametric and nonparametric bootstrapping approaches. Based on **Assumptions 1** and **3**, we have

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\hat{\varphi}(Y^*, \beta) = \bar{\varphi}_B(\beta) \quad \text{a.s.} \\
&= \lim_{n\to\infty} \frac{1}{n} E_0 E_* \{\hat{\varphi}(Y^*, \beta)\} \\
&= \lim_{n\to\infty} \frac{1}{n} E_0 E_* \{-\frac{\partial^2}{\partial\beta\partial\beta'} Q(\beta; Y^*)\} \\
&= \lim_{n\to\infty} \frac{1}{n} E_0 \{-\frac{\partial^2}{\partial\beta\partial\beta'} E_*\{Q(\beta; Y^*)\}\} \\
&= \lim_{n\to\infty} \frac{1}{n} E_0 \{-\frac{\partial^2}{\partial\beta\partial\beta'} Q(\beta; Y)\} \\
&= \lim_{n\to\infty} \frac{1}{n} E_0 \{\hat{\varphi}(Y, \beta)\} \\
&= \bar{\varphi}(\beta) \quad \text{a.s.}
\end{aligned}
$$

As a special case, when $\beta = \bar{\beta}$ which is the global maximum of the log quasi-likelihood function $Q(\beta; Y)$, we have $\bar{\varphi}(\bar{\beta}) = \bar{\varphi}_B(\bar{\beta})$. Note that the $\bar{\beta}$ is not necessarily the global maximum under the model and can be the estimators over quasi-likelihood under any correlation structures. As **Assumption 3** is validated based on the independent model as well as all the calculations involving quasi-likelihood functions for all the β in the parameter space $\mathcal{B}$, the global maximum of β under the model should also satisfy **Assumption 4**.

Proof of Asymptotic Equivalence of QAICb1 and QAICb2. After establishing **Assumptions 1 - 4**, we can show the asymptotic equivalence of QAICb1 and QAICb2 defined by equations (3.10) and (3.18).

Consider a second-order expansion of the $-2Q(\hat{\beta}; Y^*)$ about $\hat{\beta}^*$ as

$$
-2Q(\hat{\beta}; Y^*) = -2Q(\hat{\beta}; Y^*) + (\hat{\beta} - \hat{\beta}^*)' \hat{\varphi}(Y^*, \beta^*)(\hat{\beta} - \hat{\beta}^*), \tag{A7}
$$

where β^* is a vector whose value is between $\hat{\beta}$ and $\hat{\beta}^*$.

By taking expectations with respect to the bootstrap distribution on both sides of equation (A7), and referring to the consistency of $\hat{\beta}$ and $\hat{\beta}^*$ along with **Assumption 4**, we have

$$
\begin{aligned}
&E_*\{-2Q(\hat{\beta}; Y^*)\} \\
&= E_*\{-2Q(\hat{\beta}^*; Y^*) + (\hat{\beta} - \hat{\beta}^*)' \hat{\varphi}(Y^*, \beta^*)(\hat{\beta} - \hat{\beta}^*)\} \\
&= E_*\{-2Q(\hat{\beta}^*; Y^*)\} + E_*\{(\hat{\beta} - \hat{\beta}^*)' \hat{\varphi}(Y^*, \beta^*)(\hat{\beta} - \hat{\beta}^*)\} \\
&= E_*\{-2Q(\hat{\beta}^*; Y^*)\} + E_*\{m(\hat{\beta} - \hat{\beta}^*)' \bar{\varphi}_B(\bar{\beta})(\hat{\beta} - \hat{\beta}^*)\}(1 + o(1)) \quad \text{a.s.}
\end{aligned}
$$

and then we have

$$
\begin{aligned}
&E_*\{-2Q(\hat{\beta}; Y^*)\} - E_*\{-2Q(\hat{\beta}^*; Y^*)\} \\
&= E_*\{m(\hat{\beta} - \hat{\beta}^*)' \bar{\varphi}_B(\bar{\beta})(\hat{\beta} - \hat{\beta}^*)\}(1 + o(1)) \quad \text{a.s.}
\end{aligned} \tag{A8}
$$

Similarly, consider a second-order expansion of the $-2Q(\hat{\beta}^*; Y)$ about $\hat{\beta}$ as

$$
-2Q(\hat{\beta}^*; Y) = -2Q(\hat{\beta}; Y) + (\hat{\beta}^* - \hat{\beta})' \hat{\varphi}(Y, \beta^{**})(\hat{\beta}^* - \hat{\beta}), \tag{A9}
$$

where β^{**} is a vector lying between $\hat{\beta}^*$ and $\hat{\beta}$.

By taking expectations with respect to the bootstrap distribution on both sides of equation (A9), and referring to the consistency of $\hat{\beta}$ and $\hat{\beta}^*$ and **Assumption 4**, we have

$$\begin{aligned}
&E_*\{-2Q(\hat{\beta}^*;Y)\}\\
&= E_*\{-2Q(\hat{\beta};Y) + (\hat{\beta}-\hat{\beta}^*)'\hat{\varphi}(Y,\beta^{**})(\hat{\beta}-\hat{\beta}^*)\}\\
&= E_*\{-2Q(\hat{\beta};Y)\} + E_*\{(\hat{\beta}-\hat{\beta}^*)'\hat{\varphi}(Y,\beta^{**})(\hat{\beta}-\hat{\beta}^*)\}\\
&= E_*\{-2Q(\hat{\beta};Y)\} + E_*\{m(\hat{\beta}-\hat{\beta}^*)'\bar{\varphi}(\bar{\beta})(\hat{\beta}-\hat{\beta}^*)\}(1+o(1)) \quad \text{a.s.}
\end{aligned}$$

and we have

$$\begin{aligned}
&E_*\{-2Q(\hat{\beta}^*;Y)\} - E_*\{-2Q(\hat{\beta};Y)\}\\
&= E_*\{m(\hat{\beta}-\hat{\beta}^*)'\bar{\varphi}(\bar{\beta})(\hat{\beta}-\hat{\beta}^*)\}(1+o(1)) \quad \text{a.s.}
\end{aligned} \tag{A10}$$

From **Assumption 4**, we know that $\bar{\varphi}_B(\bar{\beta}) = \bar{\varphi}(\bar{\beta})$, we therefore have

$$E_*\{m(\hat{\beta}-\hat{\beta}^*)'\bar{\varphi}_B(\bar{\beta})(\hat{\beta}-\hat{\beta}^*)\} = E_*\{m(\hat{\beta}-\hat{\beta}^*)'\bar{\varphi}(\bar{\beta})(\hat{\beta}-\hat{\beta}^*)\}.$$

Combining two asymptotic results in equations (A8) and (A10), we have

$$\begin{aligned}
&E_*\{-2Q(\hat{\beta};Y^*)\} - E_*\{-2Q(\hat{\beta}^*;Y^*)\}\\
&= E_*\{-2Q(\hat{\beta}^*;Y)\} - E_*\{-2Q(\hat{\beta};Y)\}(1+o(1)) \quad a.s.
\end{aligned}$$

By the definitions of b1, b2 as well as **Assumption 3**, as $n \to \infty$, we have

$$\begin{aligned}
\text{b1} &\to E_*\{-2Q(\hat{\beta}^*;Y)\} - E_*\{-2Q(\hat{\beta}^*;Y^*)\}\\
&= E_*\{-2Q(\hat{\beta}^*;Y,)\} - E_*\{-2Q(\hat{\beta};Y^*)\}\\
&\quad + E_*\{-2Q(\hat{\beta};Y^*)\} - E_*\{-2Q(\hat{\beta}^*;Y^*)\}\\
&= E_*\{-2Q(\hat{\beta}^*;Y)\} - \{-2Q(\hat{\beta};Y)\}\\
&\quad + E_*\{-2Q(\hat{\beta};Y^*)\} - E_*\{-2Q(\hat{\beta}^*;Y^*)\}\\
&= E_*\{-2Q(\hat{\beta}^*;Y) - \{-2Q(\hat{\beta};Y)\}\}\\
&\quad + E_*\{-2Q(\hat{\beta}^*;Y)\} - E_*\{-2Q(\hat{\beta};Y\}(1+o(1)) \quad a.s.\\
&= 2E_*\{-2Q(\hat{\beta}^*;Y) - \{-2Q(\hat{\beta};Y)\}\}(1+o(1)) \quad a.s.\\
&= \text{b2}(1+o(1)) \quad a.s.
\end{aligned}$$

Therefore, two variants b1 and b2 are asymptotically equivalent, which leads to the asymptotic equivalence of QAICb1 and QAICb2.

A.2 Consistency of QAICb1 and QAICb2

We will show that b1 and b2 defined in equations (3.9) and (3.17) are consistent estimators of the bias correction term

$$E_0[E_0\{-2Q(\beta;Y)\}|_{\beta=\hat{\beta}}] - E_0[-2Q(\hat{\beta};Y)]. \tag{A11}$$

To prove the consistency of QAICb1 and QAICb2, **Lemma 1** and **Lemma 2** will be established.

Let the quantity $\mathcal{H}_B$ be defined as

$$\mathcal{H}_B = E_*\{n(\hat{\beta} - \hat{\beta}^*)'\bar{\varphi}_B(\bar{\beta})(\hat{\beta} - \hat{\beta}^*)\} = E_*\{n(\hat{\beta} - \hat{\beta}^*)'\bar{\varphi}(\bar{\beta})(\hat{\beta} - \hat{\beta}^*)\}.$$

From Section A.1, we know that b1 and b2 are asymptotically equivalent to $2\mathcal{H}_B$. In other words, to prove the consistence of b1 and b2 is the same to show that $2\mathcal{H}_B$ is the consistent estimator of bias term (A11). To start, we first define the term $\hat{\tau}(\beta, Y)$ using the first derivative of log quasi-likelihood functions by

$$\hat{\tau}(\beta, Y) = \{\frac{\partial}{\partial\beta}Q(\beta; Y)\frac{\partial}{\partial\beta'}Q(\beta; Y)\}.$$

By **Assumption 1**, similarly to $\hat{\varphi}(\beta, Y)$ in **Assumption 4**, the limit of $\hat{\tau}(\beta, Y)/n$ exists and if the limit is defined by $\bar{\tau}(\beta)$, we have

$$\bar{\tau}(\beta) = \lim_{n\to\infty}\frac{1}{n}E_0\{\hat{\tau}(\beta, Y)\} \quad \text{and}$$
$$\lim_{n\to\infty}\frac{1}{n}\hat{\tau}(\beta, Y) = \bar{\tau}(\beta) \quad \text{a.s.} \tag{A12}$$

Lemma 1.

$$\lim_{m\to\infty}\mathcal{H}_B = \text{tr}\{\bar{\tau}(\bar{\beta})\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.}$$

Proof of Lemma 1. Consider a first-order Taylor expansion of $\frac{\partial}{\partial\beta}Q(\hat{\beta}; Y)$ around $\hat{\beta}^*$. With $\hat{\beta}$ being the estimator that maximizes $Q(\beta; Y)$, we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial\beta}Q(\hat{\beta}; Y) \\ &= \frac{\partial}{\partial\beta}Q(\hat{\beta}^*; Y) + \frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta_\alpha; Y)(\hat{\beta} - \hat{\beta}^*), \end{aligned} \tag{A13}$$

where β_α is a random vector between $\hat{\beta}$ and $\hat{\beta}^*$. By solving equation (A13), we have

$$\begin{aligned} \hat{\beta} - \hat{\beta}^* &= -\{\frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta_\alpha; Y)\}^{-1}\frac{\partial}{\partial\beta}Q(\hat{\beta}^*; Y) \\ &= -\{\hat{\varphi}(Y, \beta_\alpha)\}^{-1}\frac{\partial}{\partial\beta}Q(\hat{\beta}^*; Y). \end{aligned} \tag{A14}$$

Furthermore, by **Assumption 4**, we know

$$\begin{aligned} \lim_{n\to\infty} n(\hat{\beta} - \hat{\beta}^*) &= \lim_{n\to\infty} -\{\frac{1}{n}\hat{\varphi}(Y, \beta_\alpha)\}^{-1}\frac{\partial}{\partial\beta}Q(\hat{\beta}^*; Y) \\ &= \{-\bar{\varphi}(\bar{\beta})\}^{-1}\frac{\partial}{\partial\beta}Q(\hat{\beta}^*; Y) \quad \text{a.s.} \end{aligned} \tag{A15}$$

By substituting $\hat{\beta} - \hat{\beta}^*$ of $\mathcal{H}_B$ using (A14) and applying the asymptotic property in expression (A15) combined with the consistency $\hat{\beta}$ and $\hat{\beta}^*$, we have

$$\begin{aligned}
\lim_{n\to\infty} \mathcal{H}_B &= \lim_{n\to\infty} E_*\{n(\hat{\beta}-\hat{\beta}^*)'\bar{\varphi}(\bar{\beta})(\hat{\beta}-\hat{\beta}^*)\} \\
&= \lim_{n\to\infty} E_*\{\frac{1}{n}\{n(\hat{\beta}-\hat{\beta}^*)'\}\bar{\varphi}(\bar{\beta})\{n(\hat{\beta}-\hat{\beta}^*)\}\} \\
&= \lim_{n\to\infty} E_*\{\frac{1}{n}\frac{\partial}{\partial\beta'}Q(\hat{\beta}^*;Y)\{-\bar{\varphi}(\bar{\beta})\}^{-1}\bar{\varphi}(\bar{\beta})\{-\bar{\varphi}(\bar{\beta})\}^{-1}\frac{\partial}{\partial\beta}Q(\hat{\beta}^*;Y)\} \quad \text{a.s.} \\
&= \lim_{n\to\infty} E_*\text{tr}\{\{\frac{1}{n}\{\frac{\partial}{\partial\beta}Q(\hat{\beta}^*;Y,I)\frac{\partial}{\partial\beta'}Q(\hat{\beta}^*;Y)\}\}\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.} \\
&= \lim_{n\to\infty} E_*\text{tr}\{\{\frac{1}{n}\hat{\psi}(\hat{\beta}^*,Y)\}\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.} \qquad \text{(A16)}
\end{aligned}$$

Utilizing limit (A12) and the consistency of $\hat{\beta}^*$, we have

$$\lim_{n\to\infty}\frac{1}{n}\hat{\tau}(\hat{\beta}^*,Y) = \bar{\tau}(\bar{\beta}). \qquad \text{(A17)}$$

Combining (A16) and (A17), it can be shown that

$$\lim_{n\to\infty}\mathcal{H}_B = \text{tr}\{\bar{\tau}(\bar{\beta})\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.},$$

which validates **Lemma 1**.

Next, we will provide **Lemma 2** with its proof.

Lemma 2. $2\mathcal{H}_B$ is a consistent estimator of the bias adjustment (A11).

Proof of Lemma 2. Similar to the proof of the asymptotic equivalence of two variants b1 and b2, Taylor expansion is used to construct the equivalence of two expectations. First, we take a second order expansion of $E_0\{-2Q(\beta;Y)\}|_{\beta=\hat{\beta}}$ about $\bar{\beta}$ to obtain

$$\begin{aligned}
E_0\{-2Q(\beta;Y)\}|_{\beta=\hat{\beta}} &= E_0\{-2Q(\bar{\beta};Y)\} + E_0\{-\frac{\partial}{\partial\beta}2Q(\bar{\beta};Y)\}(\hat{\beta}-\bar{\beta}) \\
&\quad + (\hat{\beta}-\bar{\beta})'E_0\{-\frac{\partial^2}{\partial\beta\partial\beta'}2Q(\beta_\gamma;Y)\}(\hat{\beta}-\bar{\beta}) \\
&= E_0\{-2Q(\bar{\beta};Y)\} \\
&\quad + (\hat{\beta}-\bar{\beta})'E_0\{-\frac{\partial^2}{\partial\beta\partial\beta'}2Q(\beta_\gamma;Y)\}(\hat{\beta}-\bar{\beta}) \\
&= E_0\{-2Q(\bar{\beta};Y)\} + (\hat{\beta}-\bar{\beta})'E_0\{\hat{\varphi}(Y,\beta_\gamma)\}(\hat{\beta}-\bar{\beta}),
\end{aligned}$$

where β_γ is a random vector between $\hat{\beta}$ and $\bar{\beta}$. Then we have

$$E_0\{-2Q(\beta;Y)\}|_{\beta=\hat{\beta}} - E_0\{-2Q(\bar{\beta};Y)\} = (\hat{\beta}-\bar{\beta})'E_0\{\hat{\varphi}(Y,\beta_\gamma)\}(\hat{\beta}-\bar{\beta}). \qquad \text{(A18)}$$

Next, considering a second order expansion of $-2Q(\bar{\beta};Y)$ about $\hat{\beta}$ and $\hat{\beta}$ being obtained by maximizing $Q(\beta;Y)$, we have

$$\begin{aligned}
-2Q(\bar{\beta};Y) &= -2Q(\hat{\beta};Y) - \frac{\partial}{\partial\beta}2Q(\hat{\beta};Y)(\bar{\beta}-\hat{\beta}) \\
&\quad + (\bar{\beta}-\hat{\beta})\{-\frac{\partial^2}{\partial\beta\partial\beta'}2Q(\beta_\delta;Y)\}(\bar{\beta}-\hat{\beta}) \\
&= -2Q(\hat{\beta};Y) + (\bar{\beta}-\hat{\beta})\{-\frac{\partial^2}{\partial\beta\partial\beta'}2Q(\beta_\delta;Y)\}(\bar{\beta}-\hat{\beta}) \\
&= -2Q(\hat{\beta};Y) + (\hat{\beta}-\bar{\beta})'\hat{\varphi}(Y,\beta_\delta)(\hat{\beta}-\bar{\beta}),
\end{aligned}$$

where β_δ is a random vector between $\hat{\beta}$ and $\bar{\beta}$. Similarly, we have

$$-2Q(\bar{\beta}; Y) - \{-2Q(\hat{\beta}; Y)\} = (\hat{\beta} - \bar{\beta})' \hat{\varphi}(Y, \beta_\delta)(\hat{\beta} - \bar{\beta}). \tag{A19}$$

We take the expectation on both sides of equation (A19) with respect to the true model and have

$$E_0\{-2Q(\bar{\beta}; Y)\} - E_0\{-2Q(\hat{\beta}; Y)\} = E_0\{(\hat{\beta} - \bar{\beta})' \hat{\varphi}(Y, \beta_\delta)(\hat{\beta} - \bar{\beta})\}, \tag{A20}$$

and we obtain the summation of equations (A18) and (A20), which gives us

$$\begin{aligned} &E_0\{-2Q(\beta; Y)\}|_{\beta=\hat{\beta}} - E_0\{-2Q(\hat{\beta}; Y)\} \\ &\quad = (\hat{\beta} - \bar{\beta})' E_0\{\hat{\varphi}(Y, \beta_\gamma)\}(\hat{\beta} - \bar{\beta}) + E_0\{(\hat{\beta} - \bar{\beta})' \hat{\varphi}(Y, \beta_\delta)(\hat{\beta} - \bar{\beta})\}. \end{aligned} \tag{A21}$$

By taking the expectation on both sides of equation (A21) with respect to the true model, we can obtain

$$\begin{aligned} &E_0\{E_0\{-2Q(\beta; Y)\}|_{\beta=\hat{\beta}}\} - E_0\{-2Q(\hat{\beta}; Y)\} \\ &\quad = E_0\{(\hat{\beta} - \bar{\beta})' E_0\{\hat{\varphi}(Y, \beta_\gamma)\}(\hat{\beta} - \bar{\beta}) + E_0\{(\hat{\beta} - \bar{\beta})' \hat{\varphi}(Y, \beta_\delta)(\hat{\beta} - \bar{\beta})\}\}, \end{aligned} \tag{A22}$$

and expression (A22) is asymptotically equivalent to

$$\lim_{n\to\infty} \ (\hat{\beta} - \bar{\beta})' E_0\{\hat{\varphi}(Y, \beta_\gamma)\}(\hat{\beta} - \bar{\beta}) + E_0\{(\hat{\beta} - \bar{\beta})' \hat{\varphi}(Y, \beta_\delta)(\hat{\beta} - \bar{\beta})\}. \tag{A23}$$

So, if we need to show that $2\mathcal{H}_B$ is asymptotically equivalent to bias term (A11), it is identical to establish the asymptotic equivalence of $2\mathcal{H}_B$ and expression (A23).

According to **Assumption 4** and the consistency of $\hat{\beta}$, we have

$$\lim_{n\to\infty} \{\frac{1}{n} \hat{\varphi}(Y, \beta_\delta)\} = \lim_{n\to\infty} E_0\{\frac{1}{n} \hat{\varphi}(Y, \beta_\gamma)\} = \bar{\varphi}(\bar{\beta}) \quad \text{a.s.} \tag{A24}$$

Moreover, by taking the first order expansion of $\frac{\partial}{\partial\beta} Q(\hat{\beta}; Y)$ about $\bar{\beta}$, we have

$$\begin{aligned} 0 = \frac{\partial}{\partial\beta} Q(\hat{\beta}; Y) &= \frac{\partial}{\partial\beta} Q(\bar{\beta}; Y) + \frac{\partial^2}{\partial\beta\partial\beta'} Q(\beta_\epsilon; Y)(\hat{\beta} - \bar{\beta}) \\ \Longrightarrow \hat{\beta} - \bar{\beta} &= -\{\frac{\partial^2}{\partial\beta\partial\beta'} Q(\beta_\epsilon; Y)\}^{-1} \frac{\partial}{\partial\beta} Q(\bar{\beta}; Y) \\ &= -\hat{\varphi}(Y, \beta_\epsilon)^{-1} \frac{\partial}{\partial\beta} Q(\bar{\beta}; Y), \end{aligned} \tag{A25}$$

where β_ϵ is a random vector between $\hat{\beta}$ and $\bar{\beta}$.

Finally, by the previously established (A12), (A24), (A25), and **Lemma 1** along with the consistency of $\hat{\beta}$, the limit of (A23) can be reduced as

$$\begin{aligned}
&\lim_{n\to\infty} (\hat{\beta}-\bar{\beta})' E_0\{\hat{\varphi}(Y,\beta_\gamma)\}(\hat{\beta}-\bar{\beta}) + E_0\{(\hat{\beta}-\bar{\beta})'\hat{\varphi}(Y,\beta_\delta)(\hat{\beta}-\bar{\beta})\} \\
&= 2\lim_{n\to\infty}\{\frac{1}{n}\{n(\hat{\beta}-\bar{\beta})'\}\bar{\varphi}(\bar{\beta})\{n(\hat{\beta}-\bar{\beta})\}\} \\
&= 2\lim_{n\to\infty}\{\frac{1}{n}\{\frac{\partial}{\partial\beta'}Q(\bar{\beta};Y)\}\{\frac{1}{n}\hat{\varphi}(Y,\beta_\epsilon)\}^{-1}\bar{\varphi}(\bar{\beta})\{\frac{1}{n}\hat{\varphi}(Y,\beta_\epsilon)\}^{-1}\}\{\frac{\partial}{\partial\beta}Q(\bar{\beta};Y)\} \\
&= 2\lim_{n\to\infty}\{\frac{1}{n}\{\{\frac{\partial}{\partial\beta'}Q(\bar{\beta};Y)\}\bar{\varphi}(\bar{\beta})^{-1}\}\bar{\varphi}(\bar{\beta})\{\bar{\varphi}(\bar{\beta})^{-1}\{\frac{\partial}{\partial\beta}Q(\bar{\beta};Y)\}\}\} \quad \text{a.s.} \\
&= 2\lim_{n\to\infty}\text{tr}\{\{\frac{1}{n}\{\frac{\partial}{\partial\beta}Q(\bar{\beta};Y)\frac{\partial}{\partial\beta'}Q(\bar{\beta};Y)\}\}\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.} \\
&= 2\lim_{n\to\infty}\text{tr}\{\{\frac{1}{n}\hat{\tau}(Y,\bar{\beta})\}\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.} \\
&= 2\text{tr}\{\bar{\tau}(\bar{\beta})\bar{\varphi}(\bar{\beta})^{-1}\} \quad \text{a.s.} \\
&= 2\lim_{n\to\infty}\mathcal{H}_B \quad \text{a.s.}
\end{aligned}$$

Therefore, we have established the asymptotic equivalence of $2\mathcal{H}_B$ and limit (A23), which completes the proof of **Lemma 2**.

Based on **Lemma 2**, we know that two variants b1 and b2 in equations (3.9) and (3.17) are asymptotically equivalent to bias term (A11) and more specifically, they are consistent estimators of bias correction term (A11).

References

Shang, J. and Cavanaugh, J. E. (2008). Bootstrap variants of the akaike information criterion for mixed model selection. *Computational Statistics & Data Analysis*, 52(4):2004–2021.