Bias Corrections of GIC for Ridge Logistic Regression

Masahiko Ono[1]*, Kaito Iriyama[1]†, Hirokazu Yanagihara[3]

[1]*Department of Mathematics Graduate School of Science Hiroshima University, Higashi-Hiroshima, Hiroshima 739-8526, Japan*

[2]*Osaka Central Advanced Mathematical Institute, Osaka Metropolitan University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan*

November 4, 2025

Abstract

Ridge regression is a technique used to avoid the problem of unstable estimation of regression coefficients when there are many explanatory variables or when strong correlations exist among them. Since the estimates from ridge regression depend on the value of the ridge parameter, appropriate selection becomes a critical issue. One common method for ridge parameter selection is based on minimizing GIC. However, because GIC is derived from asymptotic theory, it introduces non-negligible bias when the sample size is insufficiently large, leading to ineffective ridge parameter selection. Therefore, we propose the Modified GIC (MGIC), a bias-corrected GIC specifically designed for ridge parameter selection, with a particular focus on logistic regression models.

***2020 Mathematics Subject Classification*:** 62J07, 62J12
***Keywords*:** Ridge logistic regression, GIC, Bias correction

1 Introduction

When assuming that a binary response variable follows a Bernoulli distribution (binomial distribution), analysis is performed using a logistic regression model (see e.g., [2, 11, 16]). Logistic regression models are widely used across various fields as statistical models for modeling binary response variables. For example, in the medical field, Fleiss [6] analyzed the choice between two psychotherapies in psychiatry, and Moss *et al.* [22] analyzed whether the drug digitalis increases the risk of cardiac death within four months after discharge in patients who have suffered a myocardial infarction. In economics, Grzelak *et al.* [8] analyzed the profitability of transport services using light commercial vehicles with a maximum load capacity of 3.5 tons by small and medium-sized enterprises in Poland. Abid [1] analyzed the impact of corporate financial indicators and macroeconomic factors on credit risk using data from the Central Bank of Tunisia.

In logistic regression models, analysis is conducted by estimating regression coefficients using maximum likelihood estimation (MLE), which maximizes the log-likelihood function (in this paper, MLE is formulated as minimizing the negative log-likelihood function, obtained by multiplying the log-likelihood function by -1). However, when the number of explanatory variables is large or when strong correlations exist among them, the variance of the maximum likelihood estimator (MLE) increases significantly. Consequently, the estimates become unstable, potentially leading to a significant decline in prediction accuracy. A method to address such problems is ridge regression. It was originally proposed by Hoerl & Kennard [10] to improve the instability of estimators in multiple regression models. Subsequently, it was extended to logistic regression models by Schaefer *et al.* [23] and Cessie & van Houwelingen [4], among others. Other methods addressing estimation instability in multiple regression models include the Liu estimator [19] and the K-L estimator [15], which combines the ridge estimators and Liu estimators. These estimators were extended to logistic regression models by Mansson *et al.* [21] and Lukman *et al.* [20], respectively. Furthermore, various other methods have been studied to address estimation instability (see e.g., [24, 25]).

In the present study, choosing from among the various ridge logistic regression methods, we employ the approach proposed by Cessie & van Houwelingen [4], in which the regression coefficients are estimated by minimizing a penalized negative log-likelihood function. This method reduces the variability of the estimated coefficients through the penalty term, enabling the derivation of a more stable model. The magnitude of the penalty term relative to the negative log-likelihood function is controlled by a tuning parameter known as the ridge parameter. Since the estimated values are influenced by the size of the penalty term, selecting an appropriate ridge parameter is critical. One approach to selecting the ridge parameter involves using an information criterion. A representative information criterion is the Akaike Information Criterion (AIC), proposed by Akaike [3]. AIC is an asymptotically unbiased estimator of a risk function based on the Kullback-Leibler (KL) divergence [18], which measures the distance between the unknown true model and the estimated statistical model. However,

*Corresponding author (E-mail address:masahiko-ono@hiroshima-u.ac.jp)
†Current address: Ube Frontier University Kagawa High School 1-25 Bunkyocho, Ube, Yanaguchi 755-8560, Japan

because AIC is a criterion derived within the framework of maximum likelihood estimation, it cannot be directly applied to model selection in penalized estimation methods such as ridge logistic regression. An information criterion extended to accommodate penalized estimation is the Generalized Information Criterion (GIC), proposed by Konishi & Kitagawa [17]. Like AIC, GIC is an asymptotically unbiased estimator of the risk function based on KL divergence. Consequently, it may introduce non-negligible bias when sample sizes are small, potentially leading to inappropriate ridge parameter selection. Therefore, a bias-corrected GIC suitable for small-sample settings is needed. Specifically, the present paper refers to the bias-corrected GIC not as Corrected GIC, but as Modified GIC (MGIC), adopting the terminology "modified" from the Modified AIC (MAIC) proposed by Fujikoshi & Satoh [7]. This reflects that MAIC is an information criterion that corrects AIC bias even for underspecified models that do not contain the true model. In ridge regression, when the ridge parameter is non-zero, the regression coefficient estimates are biased; thus, the statistical model is considered underspecified. Consequently, MGIC denotes that it corrects bias even for underspecified models. The bias of GIC was derived by Yanagihara & Fujisawa [27], and using their result allows us to formally define MGIC. However, the GIC bias derived by Yanagihara & Fujisawa [27] assumes that the response variables are mutually independent and identically distributed. Since a logistic regression model involves response variables that are independent but follow different distributions, their result cannot be applied directly. Therefore, the present study proposes MGIC without relying on existing results concerning bias.

Bias correction for GIC can be achieved by explicitly calculating the bias term in the risk function for GIC. This paper focuses on correcting the bias up to an order of $O(n^{-1})$. In this context, the bias term involves the ridge estimator, but since the ridge estimator for the logistic regression model cannot be expressed in closed form, it is necessary to derive the probability expansion formula for the estimator (e.g., [13, 14, 29]). However, when calculating the bias using the probability expansion of the estimator, particularly for terms up to $O(n^{-1})$, the computational load increases significantly, and the bias term becomes complex. This issue can be resolved by assuming a constant order for the ridge parameter. Under this assumption, the probability expansion formula of Yanagihara *et al.* [13] can be applied, dramatically reducing the amount of new computation required. On the other hand, it is evident that this assumption cannot accommodate large ridge parameters. Therefore, the present study derives the bias up to the $O(n^{-1})$ term for both cases: assuming a constant order for the ridge parameter to reduce computational complexity, and without assuming any order for the ridge parameter to accommodate large ridge parameters despite increased computational complexity. We propose an MGIC that corrects for these biases.

When deriving bias without assuming an order in the ridge parameter, the expected value of the ridge regression estimator does not asymptotically match the true regression coefficient. This discrepancy is known as the noncentrality parameter. In bias calculations, the expected value must be computed under the distribution that incorporates this noncentrality parameter. Fujikoshi & Satoh [7] computed the bias under the noncentral Wishart distribution. Calculating the expected value under this distribution, which includes the noncentrality parameter, complicates the bias derivation. Indeed, Fujikoshi & Satoh [7] derived the bias up to the $O(n^{-1})$ term for overspecified models, for which the expectation can be obtained using a distribution without the noncentrality parameter. However, for underspecified models, they only derived the bias up to the $O(1)$ term and provided a correction for it. As noted above, deriving bias up to $O(n^{-1})$ is generally very challenging for underspecified models. However, in the present study, by performing a probability expansion around the parameters that minimize the expected value of the penalized negative log-likelihood function, we can propose a practical MGIC that, despite being an underspecified model, avoids unnecessary expectation calculations involving distributions with noncentral parameters, and relatively easily derives and corrects the bias up to the $O(n^{-1})$ term.

The structure of this paper is as follows. Chapter 2 explains ridge logistic regression and GIC. Chapter 3 derives MGIC using the bias of GIC as derived by Yanagihara & Fujisawa [27], and also derives MGIC using a simplified method that assumes a constant order for the ridge parameter. Chapter 4 derives MGIC using a rigorous method that does not assume a constant order for the ridge parameter. Chapter 5 presents numerical experiments based on simulations and real data.

2 Ridge Logistic Regression and GIC

2.1 Ridge Logistic Regression

Let $y_1, \ldots, y_n$ be mutually independent binary variables, and consider fitting the following logistic regression model to each y_i $(i = 1, \ldots, n)$:

$$y_i \sim B(1, p_i(\boldsymbol{\beta})), \quad p_i(\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})},$$

where $\boldsymbol{x}_i$ denotes a k-dimensional vector of explanatory variables, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is an unknown k-dimensional parameter vector. In the present study, we consider a model with an intercept, and thus the first component of $\boldsymbol{x}_i$ is set to 1. Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$. Then the negative log-likelihood function of the logistic regression model is given by

$$\ell(\boldsymbol{\beta} \mid \boldsymbol{y}) = -\sum_{i=1}^{n} y_i \boldsymbol{x}_i'\boldsymbol{\beta} + \sum_{i=1}^{n} \log(1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})),$$

and the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is estimated by minimizing this log-likelihood function, expressed as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}|\boldsymbol{y}).$$

In ridge logistic regression, estimation is performed using the function obtained by adding a penalty term to the negative log-likelihood function $\ell(\boldsymbol{\beta} \mid \boldsymbol{y})$, namely

$$\ell_\lambda(\boldsymbol{\beta} \mid \boldsymbol{y}) = \ell(\boldsymbol{\beta} \mid \boldsymbol{y}) + \frac{\lambda}{2}\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta},$$

where λ (> 0) is the ridge parameter, and $\boldsymbol{D}$ is the $k \times k$ diagonal matrix given by $\boldsymbol{D} = \mathrm{diag}(0, 1, \ldots, 1)$. This form is used because the intercept is generally not penalized. In this case, the ridge logistic regression estimator $\hat{\boldsymbol{\beta}}_\lambda = (\hat{\beta}_{\lambda,1}, \ldots, \hat{\beta}_{\lambda,k})'$ is obtained by minimizing $\ell_\lambda(\boldsymbol{\beta} \mid \boldsymbol{y})$. That is,

$$\hat{\boldsymbol{\beta}}_\lambda = \arg\min_{\boldsymbol{\beta}} \ell_\lambda(\boldsymbol{\beta}|\boldsymbol{y}).$$

The ridge parameter λ controls the strength of the penalty term. When $\lambda = 0$, the ridge estimator $\hat{\boldsymbol{\beta}}_\lambda$ coincides with the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$, that is, $\hat{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}}$. As λ increases, the estimates shrink toward zero, and in the limit $\lambda \to \infty$, we have $\hat{\beta}_{\lambda,j} \to 0$ for $j = 2, \ldots, k$, meaning that the estimated coefficients, excluding the intercept, converge to the zero vector. We also define the parameter that minimizes the expected value of $\ell_\lambda(\boldsymbol{\beta} \mid \boldsymbol{y})$ as

$$\boldsymbol{\beta}^*_\lambda = \arg\min_{\boldsymbol{\beta}} E[\ell_\lambda(\boldsymbol{\beta} \mid \boldsymbol{y})].$$

Substituting $\lambda = 0$ into $\boldsymbol{\beta}^*_\lambda$ yields $\boldsymbol{\beta}^*_0$, which is the true parameter. The maximum likelihood estimator $\hat{\boldsymbol{\beta}}_0$ converges in probability to $\boldsymbol{\beta}^*_0$, whereas the ridge estimator $\hat{\boldsymbol{\beta}}_\lambda$ is a shrinkage estimator; therefore, when $\lambda \neq 0$, its convergence point deviates from $\boldsymbol{\beta}^*_0$ and instead converges in probability to $\boldsymbol{\beta}^*_\lambda$.

2.2 Preliminaries

In this section, we define the expressions necessary to define GIC and to correct its bias for ridge logistic regression. First, we define the matrices required to express the stochastic expansions of the estimators in matrix form. For $a \in \mathbb{R}$, we define

$$\kappa_2(a) = a(1-a), \quad \kappa_3(a) = a(1-a)(1-2a), \quad \kappa_4(a) = a(1-a)\{1-6a(1-a)\}.$$

Then, $\kappa_j(p_i(\boldsymbol{\beta}^*_0))$ ($j = 2, 3, 4$) correspond to the cumulants of $u_i = y_i - p_i(\boldsymbol{\beta}^*_0)$. That is,

$$E[u_i^2] = \kappa_2(p_i(\boldsymbol{\beta}^*_0)), \quad E[u_i^3] = \kappa_3(p_i(\boldsymbol{\beta}^*_0)), \quad E[u_i^4] - 3E[u_i^2]^2 = \kappa_4(p_i(\boldsymbol{\beta}^*_0)).$$

Using $\kappa_2(a)$, $\kappa_3(a)$, and $\kappa_4(a)$, we define the following $k \times k$, $k \times k^2$, and $k^2 \times k^2$ matrices:

$$\boldsymbol{\Psi}_2(\boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^n \kappa_2(a_i)\boldsymbol{x}_i\boldsymbol{x}_i', \quad \boldsymbol{\Psi}_3(\boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^n \kappa_3(a_i)\boldsymbol{x}_i(\boldsymbol{x}_i \otimes \boldsymbol{x}_i)', \quad \boldsymbol{\Psi}_4(\boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^n \kappa_4(a_i)(\boldsymbol{x}_i \otimes \boldsymbol{x}_i)(\boldsymbol{x}_i \otimes \boldsymbol{x}_i)',$$

where $\otimes$ denotes the Kronecker product (see, e.g., [9, Chapter 16]). Then the first, second, and third derivatives of $p_i(\boldsymbol{\beta})$ are

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} p_i(\boldsymbol{\beta}) &= p_i(\boldsymbol{\beta})\{1 - p_i(\boldsymbol{\beta})\}\boldsymbol{x}_i = \kappa_2(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i, \\
\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} p_i(\boldsymbol{\beta}) &= p_i(\boldsymbol{\beta})\{1 - p_i(\boldsymbol{\beta})\}\{1 - 2p_i(\boldsymbol{\beta})\}\boldsymbol{x}_i\boldsymbol{x}_i' = \kappa_3(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i', \\
\left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right) p_i(\boldsymbol{\beta}) &= p_i(\boldsymbol{\beta})\{1 - p_i(\boldsymbol{\beta})\}[1 - 6p_i(\boldsymbol{\beta})\{1 - p_i(\boldsymbol{\beta})\}]\boldsymbol{x}_i \otimes \boldsymbol{x}_i\boldsymbol{x}_i' = \kappa_4(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i \otimes \boldsymbol{x}_i\boldsymbol{x}_i'.
\end{aligned}$$

It follows that, for $\boldsymbol{p}(\boldsymbol{\beta}) = (p_1(\boldsymbol{\beta}), \ldots, p_n(\boldsymbol{\beta}))'$,

$$\begin{aligned}
\frac{1}{n}\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\ell(\boldsymbol{\beta}|\boldsymbol{y}) &= \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}} p_i(\boldsymbol{\beta})\boldsymbol{x}_i' = \frac{1}{n}\sum_{i=1}^n \kappa_2(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i' = \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta})), \\
\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \frac{\partial}{\partial \boldsymbol{\beta}'}\right)\ell(\boldsymbol{\beta}|\boldsymbol{y}) &= \frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} p_i(\boldsymbol{\beta}) \otimes \boldsymbol{x}_i' = \frac{1}{n}\sum_{i=1}^n \kappa_3(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i' \otimes \boldsymbol{x}_i' = \boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta})), \\
\frac{1}{n}\left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \frac{\partial}{\partial \boldsymbol{\beta}'}\right)\ell(\boldsymbol{\beta}|\boldsymbol{y}) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right) p_i(\boldsymbol{\beta}) \otimes \boldsymbol{x}_i' \\
&= \frac{1}{n}\sum_{i=1}^n \left\{\kappa_4(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i' \otimes \boldsymbol{x}_i'\right\} \otimes \boldsymbol{x}_i' = \boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta})).
\end{aligned}$$

The above expressions imply that, as in Yanagihara *et al.* [28], both the bias expansion and its expected value can be expressed using $\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}))$, $\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}))$, and $\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}))$, enabling a simplified representation of the bias expansion.

Next, we define the risk function and, based on it, the information criterion as its estimator. In logistic regression, the loss function based on the Kullback-Leibler (KL) divergence is given by

$$\mathcal{L}(\boldsymbol{\beta}) = E[2\ell(\boldsymbol{\beta}|\boldsymbol{y})] = -2\sum_{i=1}^{n} p_i(\boldsymbol{\beta}_0^*)\boldsymbol{x}_i'\boldsymbol{\beta} + 2\sum_{i=1}^{n}\log(1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta})).$$

The risk function is defined as the expected value of the loss function. Specifically, the risk function $R(\lambda)$ is defined as

$$R(\lambda) = E[\mathcal{L}(\hat{\boldsymbol{\beta}}_\lambda)]. \tag{2.1}$$

After some straightforward calculations, the risk function can be expressed as

$$\begin{aligned} R(\lambda) &= E\left[-2\sum_{i=1}^{n} y_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\lambda + 2\sum_{i=1}^{n}\log(1+\exp(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\lambda)) + 2\sum_{i=1}^{n}(y_i - p_i(\boldsymbol{\beta}_0^*))\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\lambda\right] \\ &= 2E[\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y})] + 2E\left[\sum_{i=1}^{n}(y_i - p_i(\boldsymbol{\beta}_0^*))\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\lambda\right] = 2E[\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y})] + B, \end{aligned} \tag{2.2}$$

where B corresponds to the bias when the risk function is estimated by $2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y})$. In general, an information criterion is defined using an estimator $\hat{B}$ of B as

$$\mathrm{IC}(\lambda) = 2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y}) + \hat{B}. \tag{2.3}$$

Each information criterion can be systematically characterized by the form of the bias estimator $\hat{B}$ that it employs.

2.3 GIC in Ridge Logistic Regression

In ridge logistic regression, GIC is typically defined by approximating B in equation (2.2) using its first-order estimate. The estimated Fisher information matrix in ridge logistic regression is given by $\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))$, and the estimated Hessian matrix is denoted by $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)$, which is defined by

$$\boldsymbol{M}(\boldsymbol{a}|\lambda) = \boldsymbol{\Psi}_2(\boldsymbol{a}) + \frac{\lambda}{n}\boldsymbol{D}.$$

Using these matrices, the GIC for ridge logistic regression is expressed as

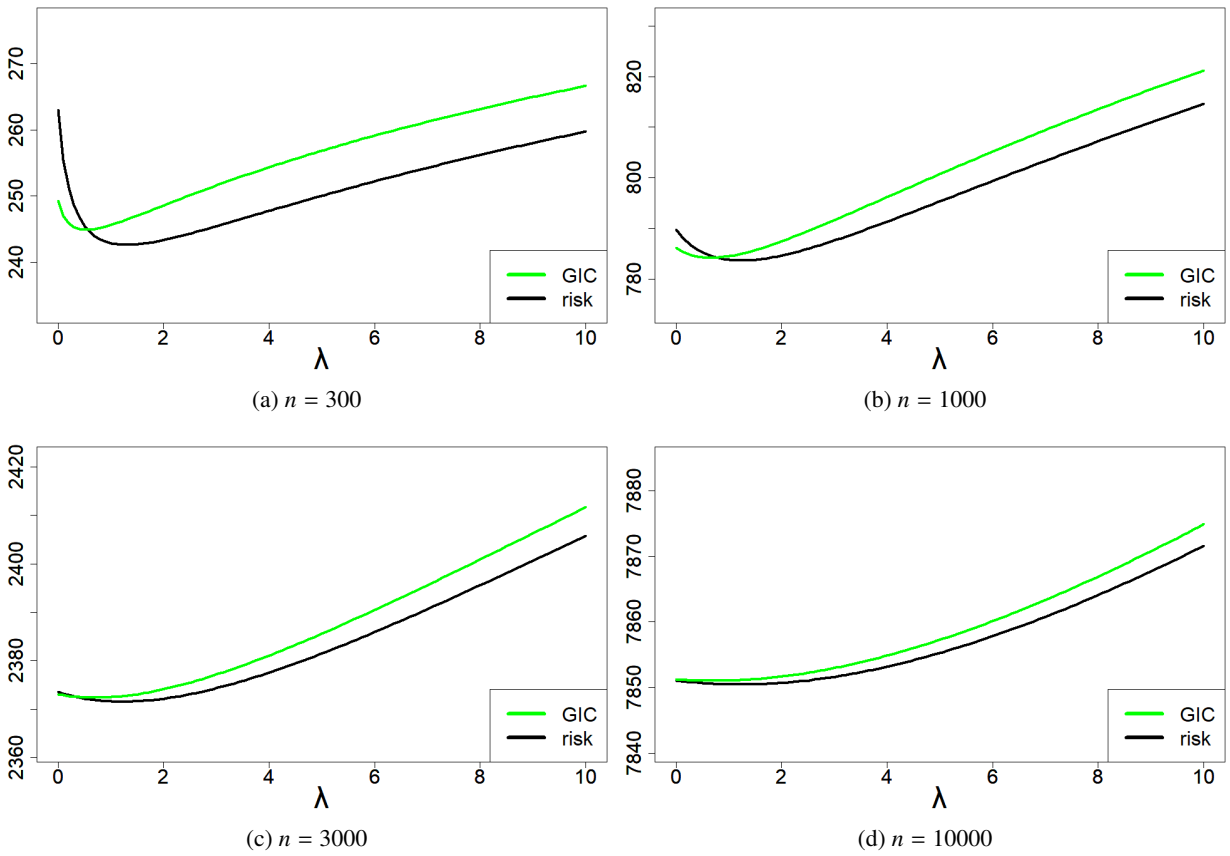
$$\mathrm{GIC}(\lambda) = 2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y}) + 2\mathrm{tr}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\right\}. \tag{2.4}$$

As mentioned earlier, GIC is an asymptotically unbiased estimator of the risk function based on the KL information. Therefore, when the sample size is small, it may not approximate the risk function accurately. We verify this fact through numerical analysis. We set the number of explanatory variables to $k = 30$ and increase the sample size n as 300, 1,000, 3,000, and 10,000 to examine how well the risk function can be approximated. The true regression coefficients $\boldsymbol{\beta}_0^*$ are generated from a standard normal distribution $N(0,1)$ for all 30 variables. The explanatory variables are generated with an autocorrelation structure with a correlation coefficient of 0.8, using the autocorrelation matrix $\boldsymbol{\Phi}_{\mathrm{A}}(\rho) = (\rho^{|i-j|})_{ij}$ and defining $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)' = (\boldsymbol{1}_n, \boldsymbol{W}\boldsymbol{\Phi}_{\mathrm{A}}(0.8)^{1/2})$, where $\boldsymbol{1}_m$ is an m-vector of ones, $\rho \in \mathbb{R}$ is the correlation coefficient, and $\boldsymbol{W}$ is an $n \times (k-1)$ matrix with each element independently drawn from a uniform distribution $U(-1,1)$. Using these, the true probabilities are set as

$$p_i(\boldsymbol{\beta}_0^*) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0^*)}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0^*)}.$$

In this simulation model, $\boldsymbol{y}$ is generated 5,000 times, and the expected values of the risk function and GIC are compared.

Figure 2.1 shows graphs of the risk function and the expected GIC values as λ ranges from 0 to 10. From Figure 2.1, it can be seen that for $n = 300$, the risk function is not well approximated; however, as n increases, the approximation improves and becomes closer to the true risk function. These results indicate that when there are 30 explanatory variables, even with $n = 300$ the estimator of the risk function is substantially biased and inadequate. Considering that both the risk function and GIC depend on the sample size n, a sufficiently large sample size, such as $n = 10{,}000$, is required for the GIC to accurately approximate the risk function. These numerical experiments clearly demonstrate that if n is not sufficiently large, the bias in estimating the risk function remains considerable, which may adversely affect the selection of the ridge parameter.



(a) $n = 300$ (b) $n = 1000$

(c) $n = 3000$ (d) $n = 10000$

Figure 2.1: Difference between risk function and expected value of GIC

3 Bias Correction using Existing Methods

High-dimensional bias correction requires substantial computational effort, involving a lengthy sequence of complex calculations. Any approach that can simplify this cumbersome process would be highly beneficial. Even if the method does not achieve complete bias correction theoretically, as long as it can be demonstrated numerically to reduce bias to some extent, it can be considered practically valuable. Therefore, in this section, we propose a simplified bias correction by adapting existing methods.

3.1 Formal Bias Correction

First, using the bias of the GIC derived by Yanagihara & Fujisawa [27], we develop a bias-corrected GIC for ridge logistic regression. In the present paper, since the MGIC obtained in this section is formally derived using existing results, we denote it by $\mathrm{MGIC}_{\mathrm{F}}$. Here, the bias of the GIC derived by Yanagihara & Fujisawa [27] assumes that the response variables $y_1, \ldots, y_n$ are independent and identically distributed. In contrast, in the ridge logistic regression considered in the present paper, the response variables are independent but not identically distributed. Therefore, although $\mathrm{MGIC}_{\mathrm{F}}$ does not fully correct the bias, it may still reduce it to some extent.

To utilize the results of Yanagihara & Fujisawa [27], it is necessary to explicitly define both the evaluation function used for model assessment and the estimation function used for model fitting. Here, let us define

$$\gamma_i(y_i|\boldsymbol{\beta}) = -y_i\boldsymbol{x}_i'\boldsymbol{\beta} + \log(1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})), \quad \psi_i(y_i|\boldsymbol{\beta}) = -y_i\boldsymbol{x}_i'\boldsymbol{\beta} + \log(1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})) + \frac{\lambda}{2n}\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta}.$$

Then the evaluation function $\ell(\boldsymbol{\beta}|\boldsymbol{y})$ and the estimating function $\ell_\lambda(\boldsymbol{\beta}|\boldsymbol{y})$ can be written as

$$\ell(\boldsymbol{\beta}|\boldsymbol{y}) = \sum_{i=1}^{n} \gamma_i(y_i|\boldsymbol{\beta}), \quad \ell_\lambda(\boldsymbol{\beta}|\boldsymbol{y}) = \sum_{i=1}^{n} \psi_i(y_i|\boldsymbol{\beta}).$$

In Yanagihara & Fujisawa [27], the bias is expressed in terms of the derivatives of $\gamma_i(y_i|\boldsymbol{\beta})$ and $\psi_i(y_i|\boldsymbol{\beta})$. Next, let us define the following:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \mathrm{diag}\left\{y_1 - p_1(\boldsymbol{\beta}), \ldots, y_n - p_n(\boldsymbol{\beta})\right\}, \quad \boldsymbol{Q}(\boldsymbol{\beta}) = \boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}))\left\{\boldsymbol{M}(\boldsymbol{\beta}|\lambda)^{-1} \otimes \boldsymbol{M}(\boldsymbol{\beta}|\lambda)^{-1}\right\},$$

$$\boldsymbol{g}_i(\boldsymbol{\beta}|\lambda) = (y_i - p_i(\boldsymbol{\beta}))\boldsymbol{x}_i + \frac{\omega}{n}\boldsymbol{D}\boldsymbol{\beta}, \quad \bar{\boldsymbol{g}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(\boldsymbol{\beta}|0), \quad \boldsymbol{H}_i(\boldsymbol{\beta}|\lambda) = \kappa_2(p_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i' + \frac{\lambda}{n}\boldsymbol{D}.$$

Then the formally bias-corrected MGIC, denoted $\mathrm{MGIC_F}$, is given by

$$\mathrm{MGIC_F}(\lambda) = 2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y}) + 2\mathrm{tr}\left\{\frac{1}{n}\left(\boldsymbol{X}'\boldsymbol{U}(\hat{\boldsymbol{\beta}}_\lambda)\boldsymbol{X} - \frac{\lambda^2}{n}\boldsymbol{D}\hat{\boldsymbol{\beta}}_\lambda\hat{\boldsymbol{\beta}}_\lambda'\boldsymbol{D}\right)\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\right\} - \frac{1}{n}\{2(\hat{\eta}_1 - \hat{\eta}_4) - (\hat{\eta}_2 - \hat{\eta}_5) + (\hat{\eta}_3 - \hat{\eta}_6)\},$$

where coefficients $\hat{\eta}_1$, $\hat{\eta}_2$, $\hat{\eta}_3$, $\hat{\eta}_4$, $\hat{\eta}_5$ and $\hat{\eta}_6$ are given by

$$\hat{\eta}_1 = \frac{1}{n}\bar{\boldsymbol{g}}(\hat{\boldsymbol{\beta}}_\lambda)'\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\sum_{i=1}^n\{\boldsymbol{H}_i(\hat{\boldsymbol{\beta}}_\lambda|0)\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|\lambda)\}, \quad \hat{\eta}_2 = \frac{1}{n}\bar{\boldsymbol{g}}(\hat{\boldsymbol{\beta}}_\lambda)'\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{Q}(\hat{\boldsymbol{\beta}}_\lambda)\mathrm{vec}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{U}(\hat{\boldsymbol{\beta}}_\lambda)\boldsymbol{X}\right),$$

$$\hat{\eta}_3 = \mathrm{tr}\left\{\frac{1}{n}\boldsymbol{X}'\boldsymbol{U}(\hat{\boldsymbol{\beta}}_\lambda)\boldsymbol{X}\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\right\}, \quad \hat{\eta}_4 = \frac{1}{n}\sum_{i=1}^n\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|0)'\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{H}_i(\hat{\boldsymbol{\beta}}_\lambda|0)\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|\lambda),$$

$$\hat{\eta}_5 = \frac{1}{n}\sum_{i=1}^n\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|0)'\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{Q}(\hat{\boldsymbol{\beta}}_\lambda)\mathrm{vec}\left\{\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|\lambda)\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|\lambda)'\right\}, \quad \hat{\eta}_6 = \frac{1}{n}\sum_{i=1}^n\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|0)'\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{H}_i(\hat{\boldsymbol{\beta}}_\lambda|\lambda)\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}_\lambda|\lambda).$$

Here, $\mathrm{vec}(\cdot)$ denotes the vec operator. For a matrix $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$, $\mathrm{vec}(\boldsymbol{A}) = (\boldsymbol{a}_1', \ldots, \boldsymbol{a}_m')'$ (see, e.g., [9, Chapter 16]).

3.2 Bias Correction under $\lambda = O(1)$

Assuming that the ridge parameter λ is of constant order, the ridge logistic regression estimator $\hat{\boldsymbol{\beta}}_\lambda$ can be expressed in terms of $\hat{\boldsymbol{\beta}}_0$. A bias correction using the stochastic expansion of $\hat{\boldsymbol{\beta}}_0$ has been carried out by Yanagihara *et al.* [29], so bias correction under $\lambda = O(1)$ can be readily obtained by leveraging their previous work. Since the MGIC derived in this section is obtained in a simplified manner, we denote it by $\mathrm{MGIC_S}$.

First, to explicitly correct the bias term of the GIC, an asymptotic expansion of $\hat{\boldsymbol{\beta}}_\lambda$ is required. Since $\ell_\lambda(\boldsymbol{\beta})$ attains its minimum at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_\lambda$, its first derivative vanishes at that point:

$$\frac{\partial}{\partial\boldsymbol{\beta}}\ell_\lambda(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_\lambda} = -\sum_{i=1}^n(y_i - p_i(\hat{\boldsymbol{\beta}}_\lambda))\boldsymbol{x}_i + \lambda\boldsymbol{D}\hat{\boldsymbol{\beta}}_\lambda = \boldsymbol{0}_k. \tag{3.1}$$

Expanding this equation around $\hat{\boldsymbol{\beta}}_\lambda = \hat{\boldsymbol{\beta}}_0$ gives

$$\begin{aligned}
&-\frac{\lambda}{n}\sum_{i=1}^n(y_i - p_i(\hat{\boldsymbol{\beta}}_0))\boldsymbol{x}_i + \frac{\lambda}{n}\boldsymbol{D}\hat{\boldsymbol{\beta}}_\lambda + \boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)(\hat{\boldsymbol{\beta}}_\lambda - \hat{\boldsymbol{\beta}}_0) + O_p(\|\hat{\boldsymbol{\beta}}_\lambda - \hat{\boldsymbol{\beta}}_0\|^2) = \boldsymbol{0}_k \\
\Leftrightarrow\quad &\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)(\hat{\boldsymbol{\beta}}_\lambda - \hat{\boldsymbol{\beta}}_0) = -\frac{\lambda}{n}\boldsymbol{D}\hat{\boldsymbol{\beta}}_\lambda + O_p(\|\hat{\boldsymbol{\beta}}_\lambda - \hat{\boldsymbol{\beta}}_0\|^2) \\
\Leftrightarrow\quad &\hat{\boldsymbol{\beta}}_\lambda = \hat{\boldsymbol{\beta}}_0 - \frac{\lambda}{n}\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1}\boldsymbol{D}\hat{\boldsymbol{\beta}}_0 + O_p(n^{-2}).
\end{aligned} \tag{3.2}$$

Next, let $\boldsymbol{z}$ be the k-dimensional vector defined by

$$\boldsymbol{z} = \frac{1}{\sqrt{n}}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\sum_{i=1}^n u_i\boldsymbol{x}_i.$$

By using $\boldsymbol{z}$, we obtain the following expansions:

$$\begin{aligned}
\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1} &= \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1} + \frac{1}{\sqrt{n}}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\left\{\boldsymbol{z}\otimes\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\right\} + O_p(n^{-1}), \\
\hat{\boldsymbol{\beta}}_0 &= \boldsymbol{\beta}_0^* + \frac{1}{\sqrt{n}}\boldsymbol{z} + O_p(n^{-1}),
\end{aligned} \tag{3.3}$$

where the details of the expansion of $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1}$ are given in Appendix 3, and those of the expansion of $\hat{\boldsymbol{\beta}}_0$ can be found in Yanagihara *et al.* [29]. Thus, $\hat{\boldsymbol{\beta}}_\lambda$ can be expanded as

$$\hat{\boldsymbol{\beta}}_\lambda = \hat{\boldsymbol{\beta}}_0 - \frac{1}{n}\lambda\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{D}\boldsymbol{\beta}_0^* + \frac{1}{n\sqrt{n}}\lambda\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\left[\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\left\{\boldsymbol{z}\otimes\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{D}\boldsymbol{\beta}_0^*\right\} - \boldsymbol{D}\boldsymbol{z}\right] + O_p(n^{-2}). \tag{3.4}$$

From equation (2.2), the bias term is

$$B = 2\sqrt{n}E\left[\boldsymbol{z}'\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\hat{\boldsymbol{\beta}}_\lambda\right].$$

Substituting equation (3.4) into this expression gives

$$B = 2\left\{\sqrt{n}E\left[\boldsymbol{z}'\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\hat{\boldsymbol{\beta}}_0\right] + E\left[-\frac{1}{\sqrt{n}}\lambda\boldsymbol{z}'\boldsymbol{D}\boldsymbol{\beta}_0^*\right] + E\left[\frac{1}{n}\boldsymbol{z}'\left[\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\left\{\boldsymbol{z}\otimes\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{D}\boldsymbol{\beta}_0^*\right\} - \boldsymbol{D}\boldsymbol{z}\right]\right]\right\} + O(n^{-2}). \tag{3.5}$$

Here, the last expectation term in the bias is of order $O(n^{-1})$, and the subsequent remainder term is $O(n^{-2})$ rather than $O(n^{-3/2})$. This occurs because the $O(n^{-3/2})$ term involves an odd-order polynomial of z, which is asymptotically normal; taking its expectation reduces the order by $O(n^{-1/2})$. Using the results of Yanagihara *et al.* [13], we have

$$\sqrt{n}E\left[z'\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\hat{\boldsymbol{\beta}}_0\right] = k + \frac{1}{2n}\{d_1(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_2(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_3(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_4(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0)\},$$

where for $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^k$, $d_1(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$, $d_2(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$, $d_3(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$, and $d_4(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$ are given by

$$\begin{aligned}
d_1(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =& -\operatorname{tr}\left\{\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\left(\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\right)\Psi_3(\boldsymbol{p}(\boldsymbol{a}_2))'\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\right\},\\
d_2(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =& 2\operatorname{tr}\left[\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))'\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\{\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\}\right],\\
d_3(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =& \operatorname{vec}(\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda))\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))'\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\operatorname{vec}(\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda))',\\
d_4(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =& -\operatorname{tr}\left[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{a}_2))\{\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\}\right].
\end{aligned}$$

Here, we have $E[u_i] = 0$ and $E[zz'] = \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}$ (proof in Appendix 2), and

$$\boldsymbol{a}_1'\boldsymbol{\Psi}_3(\boldsymbol{a})(\boldsymbol{a}_1\otimes\boldsymbol{a}_2) = \boldsymbol{a}_2'\boldsymbol{\Psi}_3(\boldsymbol{a})(\boldsymbol{a}_1\otimes\boldsymbol{a}_1) \tag{3.6}$$

holds (proof in Appendix 1). Using these results, we obtain

$$\begin{aligned}
&E\left[-\frac{1}{\sqrt{n}}\lambda z'\boldsymbol{D}\boldsymbol{\beta}_0^*\right] = 0,\\
&E\left[\frac{1}{n}z'\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\left\{z\otimes\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{D}\boldsymbol{\beta}_0^*\right\} - z'\boldsymbol{D}z\right]\\
=&\frac{\lambda}{2n}\boldsymbol{\beta}_0^{*\prime}\boldsymbol{D}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\operatorname{vec}(E[zz']) - \frac{\lambda}{2n}\operatorname{tr}(\boldsymbol{D}E[zz'])\frac{\lambda}{2n}d_5(\boldsymbol{\beta}_0^*) - \frac{\lambda}{2n}\operatorname{tr}(\boldsymbol{D}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}),
\end{aligned}$$

where $d_5(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$ is given by

$$d_5(\boldsymbol{a}_1) = \boldsymbol{a}_1'\boldsymbol{D}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_1))^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\operatorname{vec}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_1))^{-1}\right\}.$$

From the above, the bias term is given by

$$\begin{aligned}
B =& 2k - \frac{2\lambda}{n}\operatorname{tr}(\boldsymbol{D}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}) + \frac{1}{n}\{d_1(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_2(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_3(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_4(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + \lambda d_5(\boldsymbol{\beta}_0^*)\} + O(n^{-2})\\
=& 2\operatorname{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_0^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right) + \frac{1}{n}\{d_1(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_2(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_3(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + d_4(\boldsymbol{\beta}_0^*,\boldsymbol{\beta}_0^*|0) + \lambda d_5(\boldsymbol{\beta}_0^*)\} + O(n^{-2})
\end{aligned} \tag{3.7}$$

By substituting the estimate $\hat{\boldsymbol{\beta}}_\lambda$ for $\boldsymbol{\beta}_0^*$ in equation (3.7), we can obtain an estimate of the bias term. Accordingly, the bias-corrected GIC, denoted as $\mathrm{MGIC_S}$, is defined as

$$\mathrm{MGIC_S}(\lambda) = 2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y}) + 2\operatorname{tr}\left\{\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))\right\} + \frac{1}{n}\{d_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + d_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + d_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + d_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + \lambda d_5(\hat{\boldsymbol{\beta}}_\lambda)\}.$$

Under the assumption that $\lambda = O(1)$, the following holds for the risk function $R(\lambda)$:

$$R(\lambda) - E[\mathrm{MGIC_S}(\lambda)] = O(n^{-2}).$$

In particular, when $\lambda = 0$, $\mathrm{MGIC_S}$ naturally coincides with the bias-corrected AIC (CAIC) for the standard logistic regression model derived by Yanagihara *et al.* [13].

4 MGIC

In Subsection 3.2, we derived $\mathrm{MGIC_S}$ under the assumption that the ridge parameter has a constant order. However, in practice, the magnitude of λ is unknown when performing ridge regression, and assuming a constant order prevents accommodating large values of λ. Therefore, in this section, we perform bias correction for GIC without imposing any order assumptions on the ridge parameter.

We expand $\boldsymbol{\beta}_\lambda^*$ around $\hat{\boldsymbol{\beta}}_\lambda$ as follows:

$$\hat{\boldsymbol{\beta}}_\lambda = \boldsymbol{\beta}_\lambda^* + \frac{1}{\sqrt{n}}\boldsymbol{b}_\lambda^{(1)} + \frac{1}{n}\boldsymbol{b}_\lambda^{(2)} + \frac{1}{n\sqrt{n}}\boldsymbol{b}_\lambda^{(3)} + O_p(n^{-2}), \tag{4.1}$$

where $\boldsymbol{b}_\lambda^{(j)}$ $(j = 1,2,3)$ are the coefficient vectors corresponding to each term. To determine $\boldsymbol{b}_\lambda^{(j)}$ $(j = 1,2,3)$, we perform a Taylor expansion of $p_i(\hat{\boldsymbol{\beta}}_\lambda) = \exp(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\lambda)/\{1+\exp(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\lambda)\}$ around $\boldsymbol{\beta}_\lambda^*$ and substitute equation (4.1), that is,

$$\begin{aligned}
p_i(\hat{\boldsymbol{\beta}}_\lambda) - p_i(\boldsymbol{\beta}_\lambda^*) =& \frac{1}{\sqrt{n}}\kappa_2(p_i(\boldsymbol{\beta}_\lambda^*))\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(1)} + \frac{1}{n}\left\{\kappa_2(p_i(\boldsymbol{\beta}_\lambda^*))\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(2)} + \frac{1}{2}\kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))\left(\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(1)}\right)^2\right\}\\
&+ \frac{1}{n\sqrt{n}}\left\{\kappa_2(p_i(\boldsymbol{\beta}_\lambda^*))\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(3)} + \kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(1)}\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(2)} + \kappa_4(p_i(\boldsymbol{\beta}_\lambda^*))\left(\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(1)}\right)^3\right\} + O_p(n^{-2}).
\end{aligned} \tag{4.2}$$

Hence, we rewrite equation (3.1) as follows:

$$\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(p_i(\hat{\boldsymbol{\beta}}_\lambda)-p_i(\boldsymbol{\beta}_\lambda^*)\right)\boldsymbol{x}_i+\lambda\boldsymbol{D}(\hat{\boldsymbol{\beta}}_\lambda-\boldsymbol{\beta}_\lambda^*)\right\}=\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(y_i-p_i(\boldsymbol{\beta}_\lambda^*)\right)\boldsymbol{x}_i\right\}-\lambda\boldsymbol{D}\boldsymbol{\beta}_\lambda^*. \tag{4.3}$$

Since the ridge estimator does not have an expected value equal to the true regression coefficient, the term $y_i-p_i(\boldsymbol{\beta}_\lambda^*)$ satisfies $E[y_i-p_i(\boldsymbol{\beta}_\lambda^*)]=p_i(\boldsymbol{\beta}_0^*)-p_i(\boldsymbol{\beta}_\lambda^*)\neq 0$ and thus becomes a noncentral parameter. As mentioned in the Introduction, correcting the bias term up to $O(n^{-1})$ under a distribution with a noncentral parameter is difficult. Using the definition of $\boldsymbol{\beta}_\lambda^*$, we can rewrite equation (4.3) as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(p_i(\hat{\boldsymbol{\beta}}_\lambda)-p_i(\boldsymbol{\beta}_\lambda^*)\right)\boldsymbol{x}_i+\lambda\boldsymbol{D}(\hat{\boldsymbol{\beta}}_\lambda-\boldsymbol{\beta}_\lambda^*)\right\}=\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(y_i-p_i(\boldsymbol{\beta}_0^*)\right)\boldsymbol{x}_i\right\}+\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(p_i(\boldsymbol{\beta}_0^*)-p_i(\boldsymbol{\beta}_\lambda^*)\right)\boldsymbol{x}_i\right\}-\lambda\boldsymbol{D}\boldsymbol{\beta}_\lambda^*\\ \Leftrightarrow&\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\left(p_i(\hat{\boldsymbol{\beta}}_\lambda)-p_i(\boldsymbol{\beta}_\lambda^*)\right)\boldsymbol{x}_i+\lambda\boldsymbol{D}(\hat{\boldsymbol{\beta}}_\lambda-\boldsymbol{\beta}_\lambda^*)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i\boldsymbol{x}_i.\end{aligned}$$

This transformation removes the noncentral parameter, and thus that difficulty. Substituting equations (4.2) and (4.1) into this expression gives

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i\boldsymbol{x}_i=&\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\boldsymbol{b}_\lambda^{(1)}+\frac{1}{\sqrt{n}}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\boldsymbol{b}_\lambda^{(2)}+\frac{1}{2}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\right\}\\&+\frac{1}{n}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\boldsymbol{b}_\lambda^{(3)}+\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(2)})+\frac{1}{6}(\boldsymbol{I}_k\otimes\boldsymbol{b}_\lambda^{(1)})'\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\right\}\\&+\frac{\lambda}{\sqrt{n}}\left(\boldsymbol{\beta}_\lambda^*+\frac{1}{\sqrt{n}}\boldsymbol{b}_\lambda^{(1)}+\frac{1}{n}\boldsymbol{b}_\lambda^{(2)}+\frac{1}{n\sqrt{n}}\boldsymbol{b}_\lambda^{(3)}\right)+O_p(n^{-2}).\end{aligned}$$

By comparing both sides of this equation order by order, we obtain

$$\begin{aligned}\boldsymbol{b}_\lambda^{(1)}&=\frac{1}{\sqrt{n}}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^{n}u_i\boldsymbol{x}_i,\\ \boldsymbol{b}_\lambda^{(2)}&=-\frac{1}{2}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)}),\\ \boldsymbol{b}_\lambda^{(3)}&=-\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left\{\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(2)})-\frac{1}{6}(\boldsymbol{I}_k\otimes\boldsymbol{b}_\lambda^{(1)})'\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\right\}.\end{aligned}$$

Now, noting from equation (2.2) that the bias term satisfies $n^{-1/2}\sum_{i=1}^{n}u_i\boldsymbol{x}_i'=\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)$, we have

$$B=2\sqrt{n}E[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\hat{\boldsymbol{\beta}}_\lambda].$$

Substituting equation (4.1) into this expression gives

$$2\left\{\sqrt{n}E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{\beta}_0^*\right]+E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(1)}\right]+\frac{1}{\sqrt{n}}E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(2)}\right]+\frac{1}{n}E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(3)}\right]\right\}+O(n^{-2}).$$

Since $\boldsymbol{b}_\lambda^{(1)}$ also follows an asymptotic normal distribution, the expectation needs to be computed only up to terms of order $O(n^{-1})$, for the same reason as in equation (3.5). Calculating each term yields

$$\begin{aligned}E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{\beta}_0^*\right]=&0,\qquad E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(1)}\right]=\mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right]\right),\\ E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(2)}\right]=&-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))E\left[(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\boldsymbol{b}_\lambda^{(1)\prime}\right]\right),\\ E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(3)}\right]=&\frac{1}{6}\mathrm{tr}\left(\left\{3\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))-\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\right\}E\left[\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\otimes\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\right]\right).\end{aligned} \tag{4.4}$$

A proof of equation (4.4) is given in Appendix 1. Here, the expectations are

$$E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right]=\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda), \tag{4.5}$$

$$E\left[(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\boldsymbol{b}_\lambda^{(1)\prime}\right]=\frac{1}{\sqrt{n}}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}, \tag{4.6}$$

$$E\left[\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\otimes\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\right]=(\boldsymbol{I}_{k^2}+\boldsymbol{K}_k)\{\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\}+\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda))\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda))'+O(n^{-1}), \tag{4.7}$$

(calculations of the above expectations are presented in Appendix 2) where $\boldsymbol{\Omega}(\boldsymbol{a}_1, \boldsymbol{a}_2|\lambda)$ is given by

$$\boldsymbol{\Omega}(\boldsymbol{a}_1, \boldsymbol{a}_2|\lambda) = \boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_2))\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1},$$

and $\boldsymbol{K}_k$ is the $k^2 \times k^2$ matrix such that for any $k \times k$ matrix $\boldsymbol{A}$, $\mathrm{vec}(\boldsymbol{A}) = \boldsymbol{K}_k\mathrm{vec}(\boldsymbol{A}')$ (see e.g., [9, Chapter 16]). From the above, we have

$$\begin{aligned}
E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(1)}\right] =& \mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right),\\
\frac{1}{\sqrt{n}}E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(2)}\right] =& -\frac{1}{2n}\mathrm{tr}\{\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\},\\
\frac{1}{n}E\left[\boldsymbol{b}_\lambda^{(1)\prime}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)\boldsymbol{b}_\lambda^{(3)}\right] =& \frac{1}{n}\mathrm{tr}\left[\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\{\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda)\}\right]\\
&+\frac{1}{2n}\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda))\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda))'\\
&-\frac{1}{2n}\mathrm{tr}\left[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\{\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda)\}\right].
\end{aligned}$$

Hence, the bias term is

$$B = 2\mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right) + \frac{1}{n}\{d_1(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda) + d_2(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda) + d_3(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda) + d_4(\boldsymbol{\beta}_\lambda^*, \boldsymbol{\beta}_0^*|\lambda)\} + O(n^{-2}). \tag{4.8}$$

By substituting the estimators $\hat{\boldsymbol{\beta}}_\lambda$ and $\hat{\boldsymbol{\beta}}_0$ for $\boldsymbol{\beta}_\lambda^*$ and $\boldsymbol{\beta}_0^*$, we can obtain an estimate of the bias term. However, taking the expectation of the first term in the bias estimator yields

$$E\left[\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right)\right] - \mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right) = O(n^{-1}). \tag{4.9}$$

Hence, an $O(n^{-1})$ term remains. Therefore, it is necessary to expand $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}$ and $\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))$ in the first term of the bias estimator and perform the calculation. These expansions of $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}$ and $\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))$ are given by

$$\begin{aligned}
\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1} =& \left\{\boldsymbol{I}_k - \boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_\lambda^{(1)} + \frac{1}{n}\boldsymbol{G}_\lambda^{(2)}\right) + \frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\right\}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1} + O_p(n^{-3/2}),\\
\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0)) =& \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*)) + \frac{1}{\sqrt{n}}\boldsymbol{G}_0^{(1)} + \frac{1}{n}\boldsymbol{G}_0^{(2)} + O_p(n^{-3/2}).
\end{aligned}$$

Proofs of the above expansions are given in Appendix 3, and $\boldsymbol{G}_\lambda^{(1)}$ and $\boldsymbol{G}_\lambda^{(2)}$ here are given by

$$\boldsymbol{G}_\lambda^{(1)} = \frac{1}{n}\sum_{i=1}^n \kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(1)}\boldsymbol{x}_i\boldsymbol{x}_i', \quad \boldsymbol{G}_\lambda^{(2)} = \frac{1}{n}\sum_{i=1}^n\left\{\kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(2)} + \frac{1}{2}\kappa_4(p_i(\boldsymbol{\beta}_\lambda^*))\left(\boldsymbol{x}_i'\boldsymbol{b}_\lambda^{(1)}\right)^2\right\}\boldsymbol{x}_i\boldsymbol{x}_i'.$$

When taking expectation, the expansions of $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}$ and $\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))$ lower the order by $O(n^{-1/2})$ for the same reason as in equation (3.5), so it suffices to consider terms up to $O(n^{-1})$. Hence, the first term of the estimator can be written as

$$\begin{aligned}
&\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right)\\
=&\mathrm{tr}\left[\left\{\boldsymbol{I}_k - \boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_\lambda^{(1)} + \frac{1}{n}\boldsymbol{G}_\lambda^{(2)}\right) + \frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\right\}\right.\\
&\left.\times\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*)) + \frac{1}{\sqrt{n}}\boldsymbol{G}_0^{(1)} + \frac{1}{n}\boldsymbol{G}_0^{(2)}\right\}\right] + O_p(n^{-3/2})\\
=&\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\} + \mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_0^{(1)} + \frac{1}{n}\boldsymbol{G}_0^{(2)})\right)\right\} - \mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_\lambda^{(1)} + \frac{1}{n}\boldsymbol{G}_\lambda^{(2)}\right)\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\}\\
&-\mathrm{tr}\left(\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(0)}\right) + \mathrm{tr}\left(\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right) + O_p(n^{-3/2}).
\end{aligned}$$

Taking expectation term by term gives

$$
\begin{aligned}
&E\left[\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\}\right] = \mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\},\\
&E\left[\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_0^{(1)}+\frac{1}{n}\boldsymbol{G}_0^{(2)})\right)\right\}\right]\\
=&-\frac{1}{2n}\mathrm{tr}\left[\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\mathrm{vec}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\}\right\}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\right]\\
&+\frac{1}{2n}\mathrm{tr}\left[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right\}\right], \qquad (4.10)\\
&E\left[\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_\lambda^{(1)}+\frac{1}{n}\boldsymbol{G}_\lambda^{(2)}\right)\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\}\right]\\
=&-\frac{1}{2n}\mathrm{tr}\left[\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\mathrm{vec}\left(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\right)\right\}\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\right]\\
&+\frac{1}{2n}\mathrm{tr}\left[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\left\{\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\right\}\right], \qquad (4.11)\\
&E\left[\mathrm{tr}\left(\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(0)}\right)\right]\\
=&-\frac{1}{n}\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\left(\boldsymbol{M}_\lambda(\boldsymbol{\beta}_\lambda^*)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\right\}, \qquad (4.12)\\
&E\left[\mathrm{tr}\left(\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right)\right]\\
=&\frac{1}{n}\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\left(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\}, \qquad (4.13)
\end{aligned}
$$

where the proof is given in Appendix 4. From the above, equation (4.9) becomes

$$
\begin{aligned}
&E\left[\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right)\right]-\mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right)\\
=&\frac{1}{2n}\{h_1(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)+h_2(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)+h_3(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)+h_4(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\}+O(n^{-2}),
\end{aligned}
$$

where $h_1(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$, $h_2(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$, $h_3(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$, and $h_4(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)$ are given by

$$
\begin{aligned}
h_1(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =&-\mathrm{tr}\left[\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_2))^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_2))\mathrm{vec}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_2))\right\}\right\}\otimes\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_2))'\right]\\
&+\mathrm{tr}\left[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{a}_2))\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_2))\otimes\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\right\}\right],\\
h_2(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =&-\mathrm{tr}\left[\left\{\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\mathrm{vec}\left(\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\right)\right\}\otimes\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_2))'\right]\\
&+\mathrm{tr}\left[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{a}_2))\left\{\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\right\}\right],\\
h_3(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =&-2\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\left(\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\right)\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_2))'\right\},\\
h_4(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda) =&2\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))\left(\boldsymbol{\Omega}(\boldsymbol{a}_1,\boldsymbol{a}_2|\lambda)\otimes\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\right)\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{a}_1))'\boldsymbol{M}(\boldsymbol{a}_1|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{a}_2))\right\}.
\end{aligned}
$$

From the above, we define the MGIC as follows.

Definition 4.1 *The bias-corrected GIC without assuming an order for the ridge parameter is called the MGIC and is defined by*

$$
\begin{aligned}
\mathrm{MGIC}(\lambda) =&2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y})+2\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right)\\
&+\frac{1}{n}\{d_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+d_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+d_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+d_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)\}\\
&-\frac{1}{n}\{h_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+h_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+h_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+h_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)\}.
\end{aligned}
$$

Then the bias of MGIC against the risk function $R(\lambda)$ is

$$
R(\lambda)-E[\mathrm{MGIC}(\lambda)]=O(n^{-2}).
$$

When $\lambda=0$, MGIC coincides with CAIC in logistic regression as derived by Yanagihara *et al.* [13], similarly to $\mathrm{MGIC_S}$. Moreover, while MGIC does not assume any order for λ, if a constant order is assumed, it becomes asymptotically equivalent to $\mathrm{MGIC_S}$ (proof in Appendix 5).

5 Numerical Studies

5.1 Simulation Study

In this section, we examine the performance of the proposed MGIC method through simulations. We consider two scenarios for sample size and number of explanatory variables: $n = 300$, $k = 30$ and $n = 500$, $k = 50$. The true regression coefficients $\boldsymbol{\beta}_0^*$ are generated in the same manner as in the numerical experiments of Section 3.3, producing k coefficients. In addition to the autocorrelated design used in Section 3.3, $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{W}\boldsymbol{\Phi}_{\mathrm{A}}(\rho)^{1/2})$, we also consider a uniformly correlated structure by using the uniform correlation matrix $\boldsymbol{\Phi}_{\mathrm{U}}(\rho) = (1-\rho)\boldsymbol{I}_{k-1} + \rho\boldsymbol{J}_{k-1}$, resulting in $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{W}\boldsymbol{\Phi}_{\mathrm{U}}(\rho)^{1/2})$, where $\boldsymbol{J}_{k-1} = \mathbf{1}_{k-1}\mathbf{1}_{k-1}'$. The correlation coefficient ρ is set to four levels: 0.2, 0.5, 0.8, and 0.99. Using these settings, we calculate the true probabilities $\boldsymbol{p}(\boldsymbol{\beta}_0^*)$ and generate $\boldsymbol{y}$ 5,000 times based on each.

First, we compare four information criteria - MGIC, $\mathrm{MGIC_S}$, $\mathrm{MGIC_F}$, and GIC - to evaluate how accurately they approximate the risk function. Figures 5.1–5.4 show the differences between the risk function and the expected values of each information criterion as λ ranges from 0 to 10. For $\rho = 0.2, 0.5$, and 0.8, the results indicate that GIC deviates considerably from the risk function. Although $\mathrm{MGIC_F}$ approximates the risk function better than GIC, it performs worse than both MGIC and $\mathrm{MGIC_S}$. Between MGIC and $\mathrm{MGIC_S}$, although both approximate the risk function well, MGIC achieves a smaller deviation. In particular, under uniform correlation, $\mathrm{MGIC_S}$ becomes less accurate as the correlation coefficient increases. For $\rho = 0.99$, $\mathrm{MGIC_S}$ performs poorly, while the other three criteria show relatively similar results. Therefore, MGIC provides the closest overall approximation to the risk function overall.

Next, we calculate the λ^* that minimizes the risk and the $\hat{\lambda}$ that minimizes each information criterion, and then compare the mean, MSE, variance, and risk of $\hat{\lambda}$, where the MSE is defined as the mean squared error over 5,000 repetitions. In addition to the four information criteria used earlier, we also consider $\mathrm{AIC_c}$ [12], an information criterion based on the extended quasi-likelihood. For ridge logistic regression, $\mathrm{AIC_c}$ is defined as follows:

$$\mathrm{AIC_c}(\lambda) = n\log(\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_\lambda) + 1) + \frac{2n\left\{df(\hat{\boldsymbol{\beta}}_\lambda|\lambda) + 1\right\}}{n - df(\hat{\boldsymbol{\beta}}_\lambda|\lambda) - 2},$$

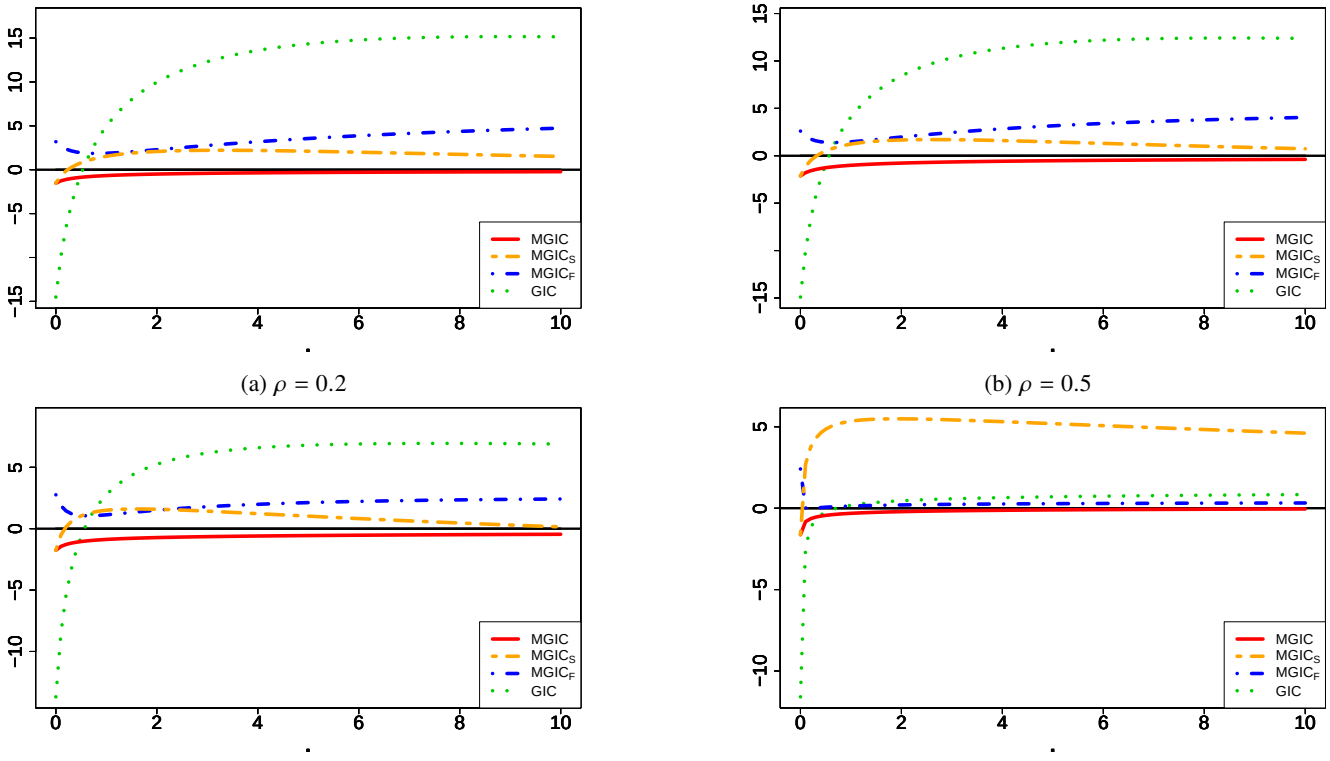
where $\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_\lambda)$ and $df(\hat{\boldsymbol{\beta}}_\lambda|\lambda)$ are given by

$$\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_\lambda) = -\frac{2}{n}\left[\hat{\boldsymbol{\beta}}_\lambda'\boldsymbol{X}'\boldsymbol{y} - \sum_{i=1}^{n}\log\{1 - p_i(\hat{\boldsymbol{\beta}}_\lambda)\}\right], \quad df(\hat{\boldsymbol{\beta}}_\lambda|\lambda) = \mathrm{tr}\left\{\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))\right\}.$$

Since the $\hat{\lambda}$ that minimizes each information criterion cannot be obtained in closed form, they were determined using a line search over the range 0 to 100.

The results for each setting are summarized in Tables 5.1–5.4. The smallest values of the MSE, variance, and risk of $\hat{\lambda}$ are highlighted in bold. First, focusing on the variance, we observe that it tends to be larger for the bias-corrected criteria, indicating somewhat inferior performance compared to GIC and $\mathrm{AIC_c}$. However, the MSE results show that for $\rho = 0.99, 0.8$, and 0.5, the bias-corrected criteria - MGIC, $\mathrm{MGIC_S}$, and $\mathrm{MGIC_F}$ - perform favorably. In particular, for $\rho = 0.99$, although the MSE of $\mathrm{MGIC_S}$ is slightly larger, considering that MSE is the sum of the variance and squared bias, it can be inferred that MGIC and $\mathrm{MGIC_F}$ select λ values closer to the true optimal λ^*. Furthermore, when comparing the risk values, MGIC achieves the lowest risk in most cases, indicating that it performs best from the perspective of risk. MGIC and $\mathrm{MGIC_F}$ show similar results, and together with the figures, this suggests that although $\mathrm{MGIC_F}$ has somewhat limited approximation accuracy, it still performs well in selecting the ridge parameter.

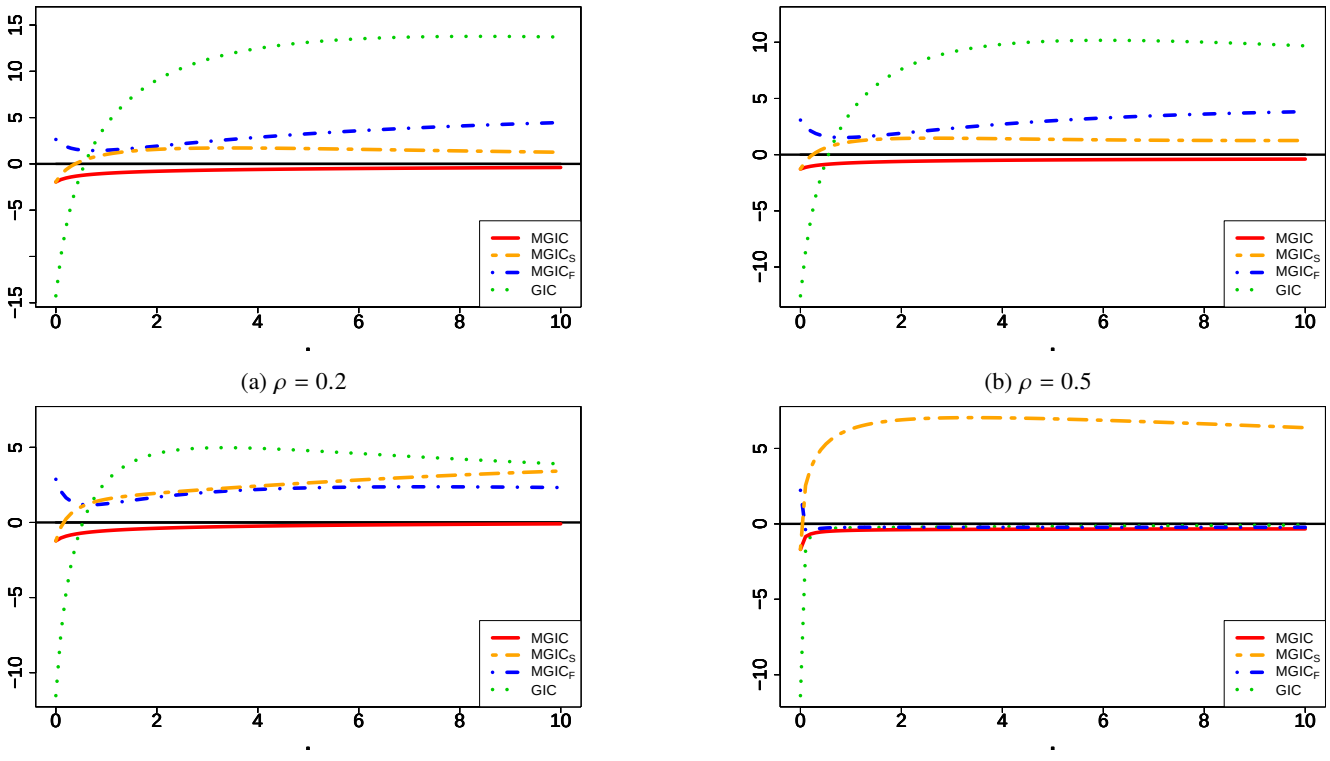
From the above results, the formally bias-corrected $\mathrm{MGIC_F}$ does not necessarily achieve high approximation accuracy, but it does perform well in selecting the ridge parameter. In contrast, $\mathrm{MGIC_S}$, which was derived for ridge logistic regression under the assumption of a constant-order ridge parameter, shows better approximation accuracy (except in the case of $\rho = 0.99$) than $\mathrm{MGIC_F}$, but it performs worse in selecting the ridge parameter that minimizes the risk. Therefore, $\mathrm{MGIC_S}$ and $\mathrm{MGIC_F}$ each have their own strengths and weaknesses. Overall, MGIC performs well in terms of both approximation accuracy and parameter selection, demonstrating that it is the most effective criterion.



(a) $\rho = 0.2$ (b) $\rho = 0.5$

(c) $\rho = 0.8$ (d) $\rho = 0.99$

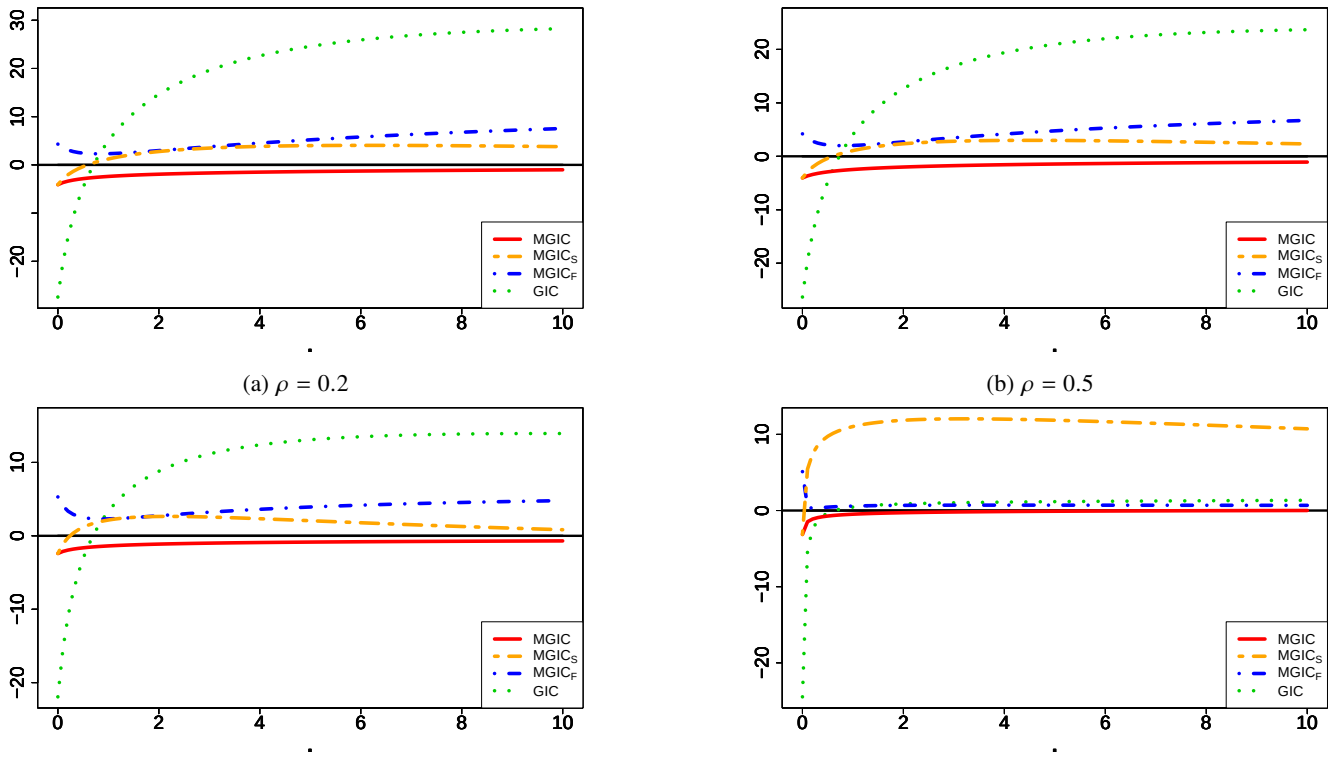
Figure 5.1: Difference between risk function and expected value of IC under autocorrelation ($n = 300, k = 30$)



(a) $\rho = 0.2$ (b) $\rho = 0.5$

(c) $\rho = 0.8$ (d) $\rho = 0.99$

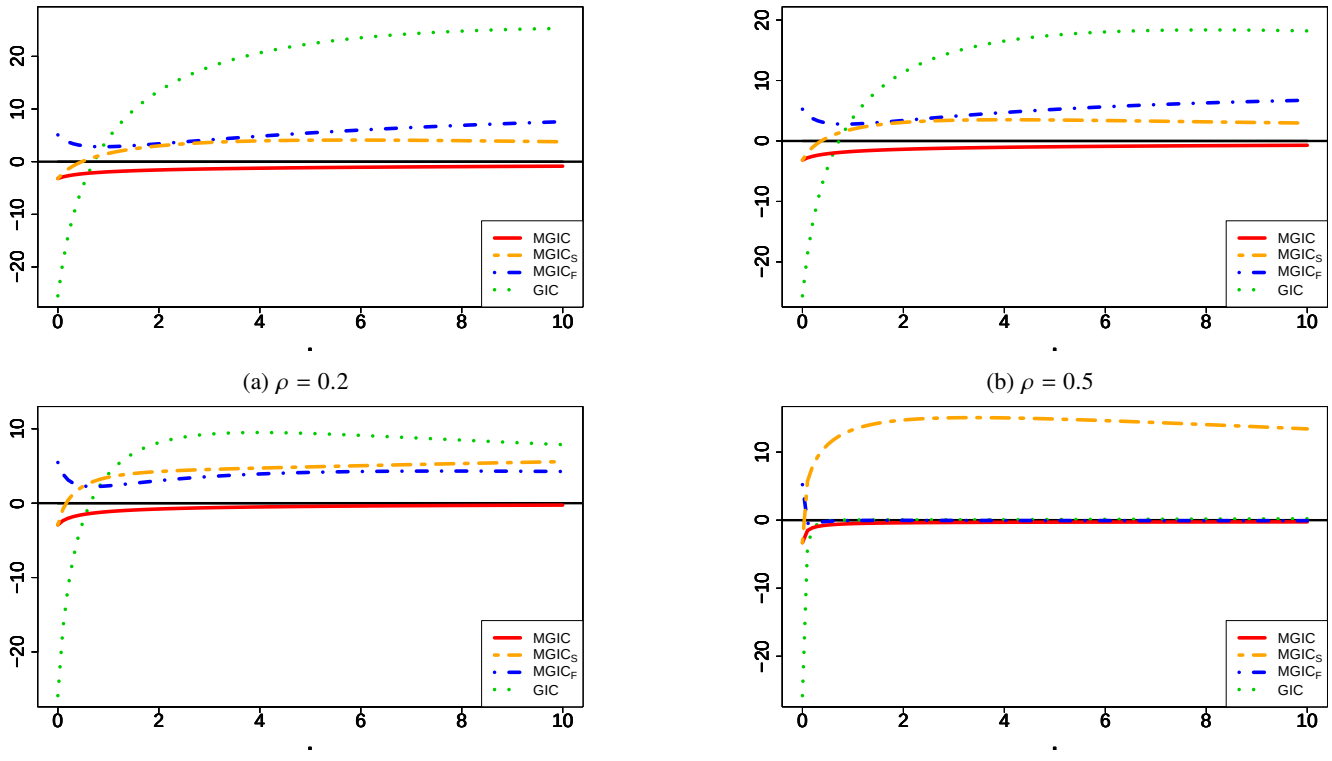
Figure 5.2: Difference between risk function and expected value of IC under uniform correlation ($n = 300, k = 30$)



(a) $\rho = 0.2$ (b) $\rho = 0.5$

(c) $\rho = 0.8$ (d) $\rho = 0.99$

Figure 5.3: Difference between risk function and expected value of IC under autocorrelation ($n = 500, k = 50$)



(a) $\rho = 0.2$ (b) $\rho = 0.5$

(c) $\rho = 0.8$ (d) $\rho = 0.99$

Figure 5.4: Difference between risk function and expected value of IC under uniform correlation ($n = 500, k = 50$)

Table 5.1: Results for $(n, k) = (300, 30)$ in the presence of autocorrelation

IC	$\rho = 0.2$				$\rho = 0.5$			
	Risk-minimizing λ^* =1.333				Risk-minimizing λ^* =1.320			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.422	0.161	0.153	**232.425**	1.399	0.193	0.187	**235.135**
$MGIC_S$	1.340	0.161	0.161	232.575	1.353	0.214	0.213	235.296
$MGIC_F$	1.388	0.158	0.155	232.468	1.359	**0.192**	0.190	235.199
GIC	0.273	**1.146**	0.023	241.255	0.362	0.963	0.045	242.490
AIC_c	0.191	1.322	**0.017**	243.500	0.248	1.180	**0.032**	245.179

IC	$\rho = 0.8$				$\rho = 0.99$			
	Risk-minimizing λ^* =1.301				Risk-minimizing λ^* =1.676			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.418	0.369	0.355	**243.777**	2.025	4.474	4.352	**268.457**
$MGIC_S$	1.453	0.491	0.468	243.938	4.702	96.112	86.958	269.531
$MGIC_F$	1.374	**0.359**	0.353	243.851	2.076	4.762	4.602	268.479
GIC	0.677	0.586	0.197	247.405	1.724	3.222	3.220	268.922
AIC_c	0.454	0.855	**0.136**	250.498	1.297	**2.820**	**2.675**	270.555

Table 5.2: Results for $(n, k) = (300, 30)$ in the presence of uniform correlation

IC	$\rho = 0.2$				$\rho = 0.5$			
	Risk-minimizing λ^* =1.330				Risk-minimizing λ^* =1.300			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.409	0.153	0.147	**242.015**	1.391	0.165	0.157	**256.544**
$MGIC_S$	1.337	0.156	0.156	242.157	1.350	0.167	0.164	256.631
$MGIC_F$	1.377	**0.151**	0.149	242.054	1.357	**0.160**	0.157	256.576
GIC	0.303	1.082	0.027	250.054	0.439	0.797	0.056	262.214
AIC_c	0.224	1.246	**0.023**	252.075	0.350	0.958	**0.055**	264.073

IC	$\rho = 0.8$				$\rho = 0.99$			
	Risk-minimizing λ^* =1.256				Risk-minimizing λ^* =11.328			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.356	0.298	0.288	272.948	5.808	68.297	37.826	**271.449**
$MGIC_S$	1.289	0.235	**0.234**	**272.927**	65.733	3022.780	62.907	273.278
$MGIC_F$	1.315	**0.291**	0.288	272.991	5.969	**65.539**	36.814	271.449
GIC	0.819	0.460	0.269	275.530	3.990	72.170	18.315	271.534
AIC_c	0.709	0.601	0.302	277.095	2.868	87.161	**15.588**	272.749

Table 5.3: Results for $(n, k) = (500, 50)$ in the presence of autocorrelation

IC	$\rho = 0.2$				$\rho = 0.5$			
	Risk-minimizing λ^* =1.677				Risk-minimizing λ^* =1.673			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.697	**0.136**	0.135	**358.836**	1.694	**0.162**	0.162	**369.988**
$MGIC_S$	1.532	0.155	0.134	359.120	1.575	0.184	0.174	370.235
$MGIC_F$	1.646	0.140	0.139	358.919	1.640	0.165	0.164	370.065
GIC	0.218	2.138	0.010	376.904	0.311	1.877	0.023	384.708
AIC_c	0.132	2.393	**0.005**	380.918	0.189	2.213	**0.013**	389.410

IC	$\rho = 0.8$				$\rho = 0.99$			
	Risk-minimizing λ^* =1.653				Risk-minimizing λ^* =2.346			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.712	0.258	0.254	**401.653**	2.700	8.843	8.718	**358.287**
$MGIC_S$	1.709	0.327	0.324	401.841	20.995	825.931	478.147	363.072
$MGIC_F$	1.660	**0.256**	0.256	401.718	2.879	9.906	9.622	358.342
GIC	0.663	1.094	0.114	408.515	2.291	6.249	6.246	358.834
AIC_c	0.423	1.580	**0.067**	413.589	0.496	**4.288**	**0.865**	368.230

Table 5.4: Results for $(n, k) = (500, 50)$ in the presence of uniform correlation

IC	$\rho = 0.2$				$\rho = 0.5$			
	Risk-minimizing λ^* =1.651				Risk-minimizing λ^* =1.680			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.686	**0.144**	0.142	**371.195**	1.717	0.195	0.194	**367.805**
$MGIC_S$	1.517	0.159	0.141	371.462	1.605	0.214	0.209	368.034
$MGIC_F$	1.634	0.145	0.145	371.267	1.656	**0.195**	0.194	367.891
GIC	0.245	1.992	0.013	387.143	0.372	1.751	0.039	380.944
AIC_c	0.154	2.250	**0.008**	390.891	0.212	2.173	**0.017**	386.404

IC	$\rho = 0.8$				$\rho = 0.99$			
	Risk-minimizing λ^* =1.658				Risk-minimizing λ^* =6.670			
	Mean	MSE	Variance	Risk	Mean	MSE	Variance	Risk
MGIC	1.723	0.329	0.325	**354.587**	4.199	25.202	19.097	337.510
$MGIC_S$	1.649	**0.293**	0.293	354.631	71.867	4361.373	110.778	343.391
$MGIC_F$	1.648	0.323	0.323	354.708	4.452	23.675	18.751	337.488
GIC	0.868	0.913	0.289	360.918	3.470	**21.058**	10.814	**337.431**
AIC_c	0.358	1.756	**0.066**	371.057	0.429	39.910	**0.955**	348.447

5.2 Example Study

In this section, we compare the performance of MGIC, $MGIC_S$, $MGIC_F$, GIC, and AIC_c using the Breast Cancer dataset (Wisconsin Breast Cancer Database) included in the mlbench package in R. This dataset, collected by Dr. W. H. Wolberg at the University of Wisconsin, contains cytological measurements of breast tumors and is publicly available in the UCI Machine Learning Repository [5, 26]. The dataset contains 699 cases, each with 10 cytological features: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses. Each feature is recorded as an integer from 1 to 10. In this study, we use these variables along with all second-order interactions and the intercept, resulting in a total of 46 explanatory variables.

The response variable is the tumor diagnosis, labeled as either "benign" or "malignant." Note that the variable bare nuclei contains some missing values (16 out of 699 cases). We excluded these cases, leaving n = 683 observations for analysis.

To evaluate the effectiveness of ridge parameter selection using information criteria, we employed the leave-one-out (LOO) method. Specifically, for the i-th observation, we excluded it and selected the ridge parameter $\hat{\lambda}^{[-i]}$ that minimizes each information criterion based on the remaining data. Then, using this $\hat{\lambda}^{[-i]}$, we estimated the parameter vector $\hat{\boldsymbol{\beta}}^{[-i]}_{\hat{\lambda}^{[-i]}}$ excluding the i-th observation, and calculated the following quantity for comparison:

$$\text{LOO} = 2\sum_{i=1}^{n}\left\{-y_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}^{[-i]}_{\hat{\lambda}^{[-i]}} + \log\left(1+\exp\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}^{[-i]}_{\hat{\lambda}^{[-i]}}\right)\right)\right\}.$$

Table 5.5 presents the LOO values for each information criterion. The proposed MGIC and MGIC_S show smaller values compared to the other criteria, with MGIC in particular being the smallest. This indicates that, even for real data, MGIC performs the best.

Table 5.5: LOO values for information criteria

IC	MGIC	MGIC_S	MGIC_F	GIC	AIC_c
LOO	**145.165**	147.484	210.620	166.336	170.284

Appendices

Appendix 1: Proof of Equation (4.4)

Let $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2 \in \mathbb{R}^k$. Then the left-hand side of equation (4.4) can be rewritten as a scalar product:

$$\begin{aligned}\boldsymbol{a}_1'\boldsymbol{\Psi}_3(\boldsymbol{a})(\boldsymbol{a}_1\otimes\boldsymbol{a}_2) =& \frac{1}{n}\sum_{i=1}^{n}\kappa_3(a_i)\boldsymbol{a}_1'\boldsymbol{x}_i(\boldsymbol{x}_i\otimes\boldsymbol{x}_i)'(\boldsymbol{a}_1\otimes\boldsymbol{a}_2)\\ =& \frac{1}{n}\sum_{i=1}^{n}\kappa_3(a_i)(\boldsymbol{a}_1'\boldsymbol{x}_i)(\boldsymbol{a}_1'\boldsymbol{x}_i)(\boldsymbol{a}_2'\boldsymbol{x}_i).\end{aligned}$$

Therefore, the order of $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ does not matter, and we have

$$\boldsymbol{a}_1'\boldsymbol{\Psi}_3(\boldsymbol{a})(\boldsymbol{a}_1\otimes\boldsymbol{a}_2) = \boldsymbol{a}_2'\boldsymbol{\Psi}_3(\boldsymbol{a})(\boldsymbol{a}_1\otimes\boldsymbol{a}_1). \tag{A.1}$$

Appendix 2: Computation of Expectations of Polynomials Involving $\boldsymbol{b}_\lambda^{(1)}$ and $\boldsymbol{z}$

First, we show equation (4.5).

$$\begin{aligned}E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right] =& \frac{1}{n}E\left[\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^{n}u_i\boldsymbol{x}_i\sum_{j=1}^{n}u_j\boldsymbol{x}_j'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right]\\ =& \frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[u_iu_j\right]\boldsymbol{x}_i\boldsymbol{x}_j'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}.\end{aligned}$$

Here, the expected value of the product of u_i and u_j is given by

$$E\left[u_iu_j\right] = \begin{cases}\kappa_2(p_i(\boldsymbol{\beta}_0^*)) & (i=j)\\ 0 & (i\neq j)\end{cases}.$$

Hence, The expected value of $\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}$ is given by

$$\begin{aligned}E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right] =& \frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^{n}\kappa_2(p_i(\boldsymbol{\beta}_0^*))\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\\ =& \boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}.\end{aligned}$$

For $E\left[\boldsymbol{z}\boldsymbol{z}'\right]$, substituting $\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1} = \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}$, we obtain

$$E\left[\boldsymbol{z}\boldsymbol{z}'\right] = \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}.$$

Next, we show equation (4.6).

$$E\left[(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\boldsymbol{b}_\lambda^{(1)\prime}\right]=\frac{1}{n\sqrt{n}}E\left[\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^n u_i\boldsymbol{x}_i\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{j=1}^n u_j\boldsymbol{x}_j\right)\sum_{m=1}^n u_m\boldsymbol{x}_m'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right]$$
$$=\frac{1}{n\sqrt{n}}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\sum_{i=1}^n\sum_{j=1}^n\sum_{m=1}^n E\left[u_iu_ju_m\right](\boldsymbol{x}_i\otimes\boldsymbol{x}_j)\boldsymbol{x}_m'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}.$$

Here, the expected value of the product of u_i, u_j, and u_m is given by

$$E\left[u_iu_ju_m\right]=\begin{cases}\kappa_3(p_i(\boldsymbol{\beta}_0^*)) & (i=j=m)\\ 0 & (\text{other})\end{cases}.$$

Hence, The expected value of $(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\boldsymbol{b}_\lambda^{(1)\prime}$ is given by

$$E\left[(\boldsymbol{b}_\lambda^{(1)}\otimes\boldsymbol{b}_\lambda^{(1)})\boldsymbol{b}_\lambda^{(1)\prime}\right]=\frac{1}{n\sqrt{n}}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\sum_{i=1}^n\kappa_3(p_i(\boldsymbol{\beta}_0^*))(\boldsymbol{x}_i\otimes\boldsymbol{x}_i)\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}$$
$$=\frac{1}{\sqrt{n}}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\right)\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}.$$

Finally, we show equation (4.7). Let $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3,\boldsymbol{a}_4\in\mathbb{R}^k$. Then the following holds:

$$(\boldsymbol{a}_1\boldsymbol{a}_2')\otimes(\boldsymbol{a}_3\boldsymbol{a}_4')=\mathrm{vec}(\boldsymbol{a}_1\boldsymbol{a}_3')\mathrm{vec}(\boldsymbol{a}_2\boldsymbol{a}_4')'=\boldsymbol{K}_k(\boldsymbol{a}_1\boldsymbol{a}_3')\otimes(\boldsymbol{a}_2\boldsymbol{a}_4').$$

Noting that

$$E\left[u_iu_ju_mu_l\right]=\begin{cases}\kappa_4(p_i(\boldsymbol{\beta}_0^*))+2\kappa_3(p_i(\boldsymbol{\beta}_0^*))^2 & (i=j=m=l)\\ \kappa_2(p_i(\boldsymbol{\beta}_0^*))\kappa_2(p_m(\boldsymbol{\beta}_0^*)) & (i=j,m=l\ \text{ or }\ i=l,m=j)\\ \kappa_2(p_i(\boldsymbol{\beta}_0^*))\kappa_2(p_j(\boldsymbol{\beta}_0^*)) & (i=m,j=l)\\ 0 & (\text{other})\end{cases},$$

we can expand equation (4.7) by letting $\boldsymbol{V}_i=\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}$ as

$$\begin{aligned}&E\left[\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\otimes\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\right]\\ =&\frac{1}{n^2}E\left[\sum_{i=1}^n u_i^4\boldsymbol{V}_i\otimes\boldsymbol{V}_i\right]+\frac{1}{n^2}E\left[\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right]\otimes E\left[\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right]\\ &+\frac{1}{n^2}\boldsymbol{K}_kE\left[\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right]\otimes E\left[\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right]+\frac{1}{n^2}E\left[\mathrm{vec}\left(\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right)\right]\otimes E\left[\mathrm{vec}\left(\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right)\right]\\ =&\frac{1}{n^2}E\left[\sum_{i=1}^n u_i^4\boldsymbol{V}_i\otimes\boldsymbol{V}_i\right]+(\boldsymbol{I}_{k^2}+\boldsymbol{K}_k)\frac{1}{n^2}E\left[\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right]\otimes E\left[\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right]+\frac{1}{n^2}E\left[\mathrm{vec}\left(\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right)\right]\otimes E\left[\mathrm{vec}\left(\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right)\right].\end{aligned}$$

Calculating each term yields the following:

$$\begin{aligned}\frac{1}{n^2}E\left[\sum_{i=1}^n u_i^4\boldsymbol{V}_i\otimes\boldsymbol{V}_i\right]=&O(n^{-1}),\\ \frac{1}{n^2}E\left[\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right]\otimes E\left[\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right]=&\frac{1}{n^2}E\left[\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right]\otimes E\left[\sum_{i=1}^n u_j^2\boldsymbol{V}_j\right]+O(n^{-1})=\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)+O(n^{-1})\\ \frac{1}{n^2}E\left[\mathrm{vec}\left(\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right)\right]\otimes E\left[\mathrm{vec}\left(\sum_{j\neq i}^{n-1}u_j^2\boldsymbol{V}_j\right)\right]=&\frac{1}{n^2}E\left[\mathrm{vec}\left(\sum_{i=1}^n u_i^2\boldsymbol{V}_i\right)\right]\otimes E\left[\mathrm{vec}\left(\sum_{i=1}^n u_j^2\boldsymbol{V}_j\right)\right]+O(n^{-1})\\ =&\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda))\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda))'+O(n^{-1}).\end{aligned}$$

Hence, The expected value of $(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime})\otimes(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime})$ is given by

$$E\left[\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\otimes\left(\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right)\right]=(\boldsymbol{I}_{k^2}+\boldsymbol{K}_k)\{\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\}+\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda))\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda))'+O(n^{-1}).$$

Appendix 3: Expansions of $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1}$, $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}$, and $\Psi_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))$

By expanding $\kappa_2(p_i(\hat{\boldsymbol{\beta}}_\lambda))$ around $\hat{\boldsymbol{\beta}}_\lambda = \boldsymbol{\beta}^*_\lambda$, we obtain

$$\kappa_2(p_i(\hat{\boldsymbol{\beta}}_\lambda)) = \kappa_2(p_i(\boldsymbol{\beta}^*_\lambda)) + \kappa_3(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*_\lambda) + \frac{1}{2}\kappa_4(p_i(\boldsymbol{\beta}^*_\lambda))\left\{\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*_\lambda)\right\}^2 + O_p(\|\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*_\lambda\|^3). \tag{A.2}$$

We first derive the expansion of $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1}$. Substituting $\lambda = 0$ and equation (3.3) into equation (A.2), we have

$$\kappa_2(p_i(\hat{\boldsymbol{\beta}}_0)) = \kappa_2(p_i(\boldsymbol{\beta}^*_0)) + \frac{1}{\sqrt{n}}\kappa_3(p_i(\boldsymbol{\beta}^*_0))\boldsymbol{x}_i'\boldsymbol{z} + O_p(n^{-1}).$$

Substituting this result into $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1}$ yields

$$\begin{aligned}\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1} &= \left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}^*_0))\left(\boldsymbol{I}_k + \frac{1}{n\sqrt{n}}\sum_{i=1}^n \kappa_3(p_i(\boldsymbol{\beta}^*_0))\boldsymbol{x}_i'\boldsymbol{z}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}^*_0))^{-1}\boldsymbol{x}_i\boldsymbol{x}_i', + O_p(n^{-1})\right)\right\}^{-1} \\ &= \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}^*_0))^{-1} + \frac{1}{\sqrt{n}}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}^*_0))^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}^*_0))\left\{\boldsymbol{z}\otimes\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}^*_0))^{-1}\right\} + O_p(n^{-1}).\end{aligned}$$

Next, we derive the expansions of $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}$ and $\Psi_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))$. Substituting equation (4.1) into (A.2), we obtain

$$\kappa_2(p_i(\hat{\boldsymbol{\beta}}_\lambda)) = \kappa_2(p_i(\boldsymbol{\beta}^*_\lambda)) + \kappa_3(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'\left(\frac{1}{\sqrt{n}}\boldsymbol{b}^{(1)}_\lambda + \frac{1}{n}\boldsymbol{b}^{(2)}_\lambda\right) + \frac{1}{2n}\kappa_4(p_i(\boldsymbol{\beta}^*_\lambda))\left\{\boldsymbol{x}_i'\boldsymbol{b}^{(1)}_\lambda\right\}^2 + O_p(n^{-3/2}).$$

Then, substituting this result into $\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}$ and $\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))$, we have

$$\begin{aligned}\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1} &= \left\{\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)\left(\boldsymbol{I}_k + \frac{1}{\sqrt{n}}\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\boldsymbol{G}^{(1)}_\lambda + \frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\boldsymbol{G}^{(2)}_\lambda + O_p(n^{-3/2})\right)\right\}^{-1} \\ &= \left\{\boldsymbol{I}_k - \boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}^{(1)}_\lambda + \frac{1}{n}\boldsymbol{G}^{(2)}_\lambda\right) + \frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\boldsymbol{G}^{(1)}_\lambda\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\boldsymbol{G}^{(1)}_\lambda\right\}\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1} + O_p(n^{-3/2}), \\ \boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0)) &= \boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}^*_0)) + \frac{1}{\sqrt{n}}\boldsymbol{G}^{(1)}_0 + \frac{1}{n}\boldsymbol{G}^{(2)}_0 + O_p(n^{-3/2}).\end{aligned}$$

Appendix 4: Computation of Expected Value of Expansion of $\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right)$

First, we compute the expected values of $\boldsymbol{G}^{(1)}_\lambda$ and $\boldsymbol{G}^{(2)}_\lambda$:

$$\begin{aligned}E\left[\boldsymbol{G}^{(1)}_\lambda\right] &= \frac{1}{n}\sum_{i=1}^n \kappa_3(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'E\left[\boldsymbol{b}^{(1)}_\lambda\right]\boldsymbol{x}_i\boldsymbol{x}_i' = \boldsymbol{O}_{k\times k}, \\ E\left[\boldsymbol{G}^{(2)}_\lambda\right] &= \frac{1}{n}\sum_{i=1}^n\left\{\kappa_3(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'E\left[\boldsymbol{b}^{(2)}_\lambda\right] + \frac{1}{2}\kappa_4(p_i(\boldsymbol{\beta}^*_\lambda))E\left[\left(\boldsymbol{x}_i'\boldsymbol{b}^{(1)}_\lambda\right)^2\right]\right\}\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \frac{1}{2n}\sum_{i=1}^n\left\{-\kappa_3(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}^*_\lambda))\mathrm{vec}\left(E\left[\boldsymbol{b}^{(1)}_\lambda\boldsymbol{b}^{(1)\prime}_\lambda\right]\right) + \kappa_4(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'E\left[\boldsymbol{b}^{(1)}_\lambda\boldsymbol{b}^{(1)\prime}_\lambda\right]\boldsymbol{x}_i\right\}\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \frac{1}{2n}\sum_{i=1}^n\left\{-\kappa_3(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}^*_\lambda))\mathrm{vec}\left(\boldsymbol{\Omega}(\boldsymbol{\beta}^*_\lambda,\boldsymbol{\beta}^*_0|\lambda)\right) + \kappa_4(p_i(\boldsymbol{\beta}^*_\lambda))\boldsymbol{x}_i'\boldsymbol{\Omega}(\boldsymbol{\beta}^*_\lambda,\boldsymbol{\beta}^*_0|\lambda)\boldsymbol{x}_i\right\}\boldsymbol{x}_i\boldsymbol{x}_i'.\end{aligned}$$

Thus, equation (4.11) becomes

$$\begin{aligned}&E\left[\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}^{(1)}_0 + \frac{1}{n}\boldsymbol{G}^{(2)}_0)\right)\right\}\right] \\ &= \frac{1}{2n^2}\mathrm{tr}\left[\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\sum_{i=1}^n\{-\kappa_3(p_i(\boldsymbol{\beta}^*_0))\boldsymbol{x}_i'\boldsymbol{M}_\lambda(\boldsymbol{\beta}^*_0)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}^*_0))\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{\beta}^*_\lambda,\boldsymbol{\beta}^*_0|\lambda)\right) + \kappa_4(p_i(\boldsymbol{\beta}^*_0))\boldsymbol{x}_i'\boldsymbol{\Omega}(\boldsymbol{\beta}^*_\lambda,\boldsymbol{\beta}^*_0|\lambda)\boldsymbol{x}_i\}\boldsymbol{x}_i\boldsymbol{x}_i'\Bigg] \\ &= -\frac{1}{2n}\mathrm{tr}\Bigg[\{\boldsymbol{M}_0(\boldsymbol{\beta}^*_0)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}^*_0))\mathrm{vec}\left(\boldsymbol{\Omega}(\boldsymbol{\beta}^*_\lambda,\boldsymbol{\beta}^*_0|\lambda)\right)\otimes\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}^*_0))'\Bigg] + \frac{1}{2n}\mathrm{tr}\Big[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}^*_0))\{\boldsymbol{\Omega}(\boldsymbol{\beta}^*_\lambda,\boldsymbol{\beta}^*_0|\lambda)\otimes\boldsymbol{M}(\boldsymbol{\beta}^*_\lambda|\lambda)^{-1}\}\Big].\end{aligned}$$

Similarly, we have

$$
\begin{aligned}
&E\left[\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{G}_\lambda^{(1)}+\frac{1}{n}\boldsymbol{G}_\lambda^{(2)}\right)\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right\}\right]\\
&=-\frac{1}{2n}\mathrm{tr}\left[\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))\mathrm{vec}\left(\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\right)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\}\boldsymbol{\Psi}_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\right]+\frac{1}{2n}\mathrm{tr}\Big[\boldsymbol{\Psi}_4(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\{\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\otimes\boldsymbol{\Omega}(\boldsymbol{\beta}_\lambda^*,\boldsymbol{\beta}_0^*|\lambda)\}\Big].
\end{aligned}
$$

Therefore, we obtain equation (4.10). Moreover, the expectation of $\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_0^{(1)\prime}$ is given by

$$
\begin{aligned}
E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_0^{(1)\prime}\right]=&\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^n\sum_{j=1}^n E\left[u_iu_j\right]\boldsymbol{x}_i\boldsymbol{x}_j'\Psi_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}\\
=&\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\Psi_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\Psi_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))^{-1}=\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}.
\end{aligned}
$$

Then equation (4.12) becomes

$$
\begin{aligned}
&E\left[\mathrm{tr}\left(\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_0^{(1)}\right)\right]\\
=&\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))\kappa_3(p_j(\boldsymbol{\beta}_0^*))\mathrm{tr}\left(\boldsymbol{x}_i'E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_0^{(1)\prime}\right]\boldsymbol{x}_j\right)\mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j'\right)\\
=&\frac{1}{n^3}\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\sum_{i=1}^n\kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{x}_i'\otimes\boldsymbol{x}_i\boldsymbol{x}_i')(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1})\sum_{j=1}^n\kappa_3(p_j(\boldsymbol{\beta}_0^*))(\boldsymbol{x}_j\otimes\boldsymbol{x}_j\boldsymbol{x}_j')\right\}\\
=&\frac{1}{n}\mathrm{tr}\Big\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1})\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_0^*))'\Big\}.
\end{aligned}
$$

Similarly, we obtain

$$
\begin{aligned}
&E\left[\mathrm{tr}\left(\frac{1}{n}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{G}_\lambda^{(1)}\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right)\right]\\
=&\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\left\{\kappa_3(p_i(\boldsymbol{\beta}_\lambda^*))\kappa_3(p_j(\boldsymbol{\beta}_\lambda^*))\mathrm{tr}\left(\boldsymbol{x}_i'E\left[\boldsymbol{b}_\lambda^{(1)}\boldsymbol{b}_\lambda^{(1)\prime}\right]\boldsymbol{x}_j\right)\mathrm{tr}\left(\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\right)\right\}\\
=&\frac{1}{n}\mathrm{tr}\Big\{\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))(\boldsymbol{\Omega}_\lambda(\boldsymbol{\beta}_\lambda^*)\otimes\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1})\Psi_3(\boldsymbol{p}(\boldsymbol{\beta}_\lambda^*))'\boldsymbol{M}(\boldsymbol{\beta}_\lambda^*|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0^*))\Big\}.
\end{aligned}
$$

Thus, we obtain equation (4.13).

Appendix 5: Proof of Relationship between MGIC_S and MGIC

Assume that $\lambda = O(1)$, and expand $\hat{\boldsymbol{\beta}}_0$ around $\hat{\boldsymbol{\beta}}_\lambda$ and substitute it into MGIC to investigate the relationship between MGIC_S and MGIC. First, using equation (3.2), $\hat{\boldsymbol{\beta}}_0$ can be expressed in terms of $\hat{\boldsymbol{\beta}}_\lambda$. Since $\boldsymbol{M}(\boldsymbol{\beta}|\lambda)^{-1}$ can be expanded as

$$
\boldsymbol{M}(\boldsymbol{\beta}|\lambda)^{-1}=\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}))^{-1}-\frac{\lambda}{n}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}))^{-1}\boldsymbol{D}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}))^{-1}+O_p(n^{-2}),\tag{A.3}
$$

we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_0&=\hat{\boldsymbol{\beta}}_\lambda+\frac{\lambda}{n}\boldsymbol{M}(\hat{\boldsymbol{\beta}}_0|\lambda)^{-1}\boldsymbol{D}\hat{\boldsymbol{\beta}}_0+O_p(n^{-2})\\
\Leftrightarrow\hat{\boldsymbol{\beta}}_0&=\hat{\boldsymbol{\beta}}_\lambda+\frac{\lambda}{n}\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0))^{-1}\boldsymbol{D}\hat{\boldsymbol{\beta}}_0+O_p(n^{-2}).
\end{aligned}\tag{A.4}
$$

Next, substitute equation (A.4) into MGIC. Using equations (A.4) and (A.3), we obtain

$$
\begin{aligned}
&\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))=\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))+\frac{1}{n}\boldsymbol{L}+O_p(n^{-2}),\quad \boldsymbol{\Psi}_3(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))=\boldsymbol{\Psi}_3(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))+O_p(n^{-1}),\\
&\boldsymbol{\Psi}_4(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))=\boldsymbol{\Psi}_4(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))+O_p(n^{-1}),\quad \boldsymbol{\Omega}(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0)=\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))^{-1}+O_p(n^{-1}),
\end{aligned}
$$

where $\boldsymbol{L}=n^{-1}\sum_{i=1}^n\kappa_3(p_i(\hat{\boldsymbol{\beta}}_\lambda))\boldsymbol{x}_i'\boldsymbol{z}\boldsymbol{x}_i\boldsymbol{x}_i'$, $\boldsymbol{z}=\boldsymbol{\Psi}_2(\boldsymbol{p}(\boldsymbol{\beta}_0))^{-1}\boldsymbol{D}\hat{\boldsymbol{\beta}}_0$. Using these results, we have

$$
\begin{aligned}
&\frac{1}{n}\{d_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+d_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+d_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+d_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)\}\\
=&\frac{1}{n}\{d_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0)+d_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0)+d_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0)+d_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0)\}+O_p(n^{-2}),\\
&\frac{1}{n}\{h_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+h_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+h_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)+h_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)\}=O_p(n^{-2}).
\end{aligned}
$$

Moreover, $\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right)$ can be expanded as follows:

$$\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right) = \mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))\right) + \frac{1}{n}\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{L}\right) + O_p(n^{-2}).$$

Calculating the second term yields

$$\begin{aligned}
\frac{1}{n}\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{L}\right) &= \frac{1}{n^2}\mathrm{tr}\left(\sum_{i=1}^{n}\kappa_3(p_i(\hat{\boldsymbol{\beta}}_\lambda))\boldsymbol{z}'\boldsymbol{x}_i\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}))^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\right) + O_p(n^{-2}) \\
&= \frac{1}{n}\boldsymbol{z}'\frac{1}{n}\sum_{i=1}^{n}\kappa_3(p_i(\hat{\boldsymbol{\beta}}_\lambda))(\boldsymbol{x}_i\boldsymbol{x}_i'\otimes\boldsymbol{x}_i)\mathrm{vec}\left\{\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}))^{-1}\right\} + O_p(n^{-2}) \\
&= \lambda d_5(\hat{\boldsymbol{\beta}}_\lambda) + O_p(n^{-2}).
\end{aligned}$$

Therefore, assuming $\lambda = O(1)$, we have

$$\begin{aligned}
\mathrm{MGIC}(\lambda) =& 2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y}) + 2\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_0))\right) \\
&+ \frac{1}{n}\{d_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda) + d_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda) + d_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda) + d_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)\} \\
&- \frac{1}{n}\{h_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda) + h_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda) + h_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda) + h_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_0|\lambda)\} \\
=& 2\ell(\hat{\boldsymbol{\beta}}_\lambda|\boldsymbol{y}) + 2\mathrm{tr}\left(\boldsymbol{M}(\hat{\boldsymbol{\beta}}_\lambda|\lambda)^{-1}\boldsymbol{\Psi}_2(\boldsymbol{p}(\hat{\boldsymbol{\beta}}_\lambda))\right) \\
&+ \frac{1}{n}\{d_1(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + d_2(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + d_3(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + d_4(\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{\beta}}_\lambda|0) + \lambda d_5(\hat{\boldsymbol{\beta}}_\lambda)\} + O(n^{-2}) \\
=& \mathrm{MGIC}_\mathrm{S}(\lambda) + O_p(n^{-2}).
\end{aligned}$$

Therefore, MGIC_S and MGIC are asymptotically equivalent.

Acknowledgements

The authors would like to express their deepest gratitude to Associate Professor Ryoya Oda of Hiroshima University for their generous guidance and valuable advice despite their busy schedule. The authors are also deeply grateful to Assistant Professor Mineaki Ohishi of Tohoku University and Mr. Koki Kirishima of Osaka Metropolitan University for their helpful suggestions and kind support in many aspects of this study. This work was supported by JST, Establishment of University Fellowships towards the Creation of Science Technology Innovation, Grant Number JPMJFS2129. The third author's research was supported by JSPS KAKENHI Grant Number 23H00809. The authors thank FORTE Science Communications (https://www.forte-science.co.jp/) for English language editing.

References

[1] Abid, L. (2022). A logistic regression model for credit risk of companies in the service sector. *Int. Res. Econ. Finance*, **6**, 1179–1189.

[2] Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Wiley.

[3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. Inf. Theory*, (eds. B. N. Petrov and F. Csaki), 267–281, Akademiai Kiado, Budapest.

[4] Cessie, L. S. & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Appl. Statist.*, **41**, 191–201.

[5] Dua, D. & Graff, C. (2019). *UCI Machine Learning Repository*. University of California, Irvine. Retrieved from http://archive.ics.uci.edu/ml.

[6] Fleiss, J. L., Williams, J. B. & Dubro, A. F. (1986). The logistic regression analysis of psychiatric data. *J. Psychiatr. Res.*, **20**, 195–209.

[7] Fujikoshi, Y. & Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression. *Biometrika*, **84**, 707–716.

[8] Grzelak, M., Owczarek, P., Stoica, R.-M., Voicu, D. & Vilau, R. (2024). Application of logistic regression to analyze the economic efficiency of vehicle operation in terms of the financial security of enterprises. *Logistics*, **8**, 46.

[9] Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.

[10] Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.

[11] Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons.

[12] Hurvich, M. C. & Tsai, C. (1995). Model selection for extended quasi-likelihood models in small samples. *Biometrics*, **51**, 1077–1084.

[13] Imori, S., Yanagihara, H. & Wakaki, H. (2014). Simple formula for calculating bias-corrected AIC in generalized linear models. *Scand. J. Statist.*, **41**, 535–555.

[14] Kamo, K., Yanagihara, H. & Satoh, K. (2013). Bias-corrected AIC for selecting variables in Poisson regression models. *Commun. Statist. Theory Methods*, **42**, 1911–1921.

[15] Kibria, B. M. G. & Lukman, A. F. (2020). A new ridge type estimator for the linear regression model: Simulations and applications. *Comput. Math. Methods Med.*, **2020**, 9758378.

[16] Kleinbaum, D. G. & Klein, M. (2002). *Logistic Regression: A Self-Learning Text* (2nd ed.). Springer.

[17] Konishi, S. & Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, **83**, 875–890.

[18] Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.

[19] Liu, K. (1993). A new class of biased estimate in linear regression. *Commun. Statist. Theory Methods*, **22**, 393–402.

[20] Lukman, A. F., Kibria, B. M. G., Nziku, C. K., Amin, M., Adewuyi, E. T. & Farghali, R. (2023). K-L Estimator: Dealing with multicollinearity in the logistic regression model. *Mathematics*, **11**, 340.

[21] Mansson, K., Kibria, B. M. G. & Shukur, G. (2012). On Liu estimators for the logit regression model. *Econ. Model.*, **29**, 1483–1488.

[22] Moss, A. J., Davis, H. T., Conard, D. L., DeCamilla, J. J. & Odoroff, C. L. (1981). Digitalis-associated cardiac mortality after myocardial infarction. *Circulation*, **64**, 1150–1158.

[23] Schaefer, R. L., Roi, D. L. & Wolfe, A. R. (1984). A ridge logistic estimator. *Commun. Statist. Theory Methods*, **13**, 99–113.

[24] Schaefer, R. L. (1986). Alternative estimators in logistic regression when the data are collinear. *J. Stat. Comput. Simul.*, **25**, 75–91.

[25] Sinkovec, H., Heinze, G., Blagus, R. & Geroldinger, A. (2021). To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. *BMC Med. Res. Methodol.*, **21**, 199.

[26] Wolberg, W. H. & Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 9193–9196.

[27] Yanagihara, H. & Fujisawa, H. (2012). Iterative bias correction of the cross-validation criterion. *Scand. J. Stat.*, **39**, 116–130.

[28] Yanagihara, H., Kamo, K., Imori, S. & Satoh, K. (2012). Bias-corrected AIC for selecting variables in multinomial logistic regression models. *Linear Algebra Appl.*, **436**, 4329–4341.

[29] Yanagihara, H., Sekiguchi, R. & Fujikoshi, Y. (2003). Bias correction of AIC in logistic regression models. *J. Statist. Plann. Inference*, **115**, 349–360.