

Semiparametric Regression Splines in Matched Case-Control Studies

Inyoung Kim,^{1,*} Noah D. Cohen,^{2,**} and Raymond J. Carroll^{3,***}

¹Department of Statistics, Texas A&M University, College Station, Texas, U.S.A.

²Department of Large Animal Medicine and Surgery, Texas A&M University, College Station, Texas, U.S.A.

³Department of Statistics, Texas A&M University, College Station, Texas, U.S.A.

* *email:* inyoung@stat.tamu.edu

** *email:* ncohen@cvm.tamu.edu

*** *email:* carroll@stat.tamu.edu

SUMMARY. We develop semiparametric methods for matched case-control studies using regression splines. Three methods are developed: 1) an approximate cross-validation scheme to estimate the smoothing parameter inherent in regression splines, as well as 2) Monte Carlo expectation maximization (MCEM) and 3) Bayesian methods to fit the regression spline model. We compare the approximate cross-validation approach, MCEM, and Bayesian approaches using simulation, showing that they appear approximately equally efficient; the approximate cross-validation method is computationally the most convenient. An example from equine epidemiology that motivated the work is used to demonstrate our approaches.

KEY WORDS: Bayesian method; Cross-validation; EM algorithm; Matched case-control; Monte Carlo EM; Penalized regression splines; Semiparametric regression splines.

1. Introduction

This article concerns semiparametric regression in matched case-control studies. Matched case-control studies are based upon the classical prospective logistic regression model, with a binary outcome Y (case-control status), covariate (Z, X) , stratum level S and the model

$$\Pr(Y = 1 | Z, X) = H\{Z^T \beta_0 + \beta_1 X + q(S)\}, \quad (1)$$

where $H(\bullet)$ is the logistic distribution function and $q(\bullet)$ is an arbitrary function including the intercept and unknown effects of the strata. The classical matched study begins with the model (1), but by conditioning on the fixed number of cases and controls in the stratum, any stratum effect is removed, i.e. $q(\bullet)$ disappears (Hosmer and Lemeshow, 1989).

The purpose of this article is to generalize model (1) into the matched study, to allow the effect of X to be modeled nonparametrically. The resulting prospective model is

$$\Pr(Y = 1 | Z, X) = H\{Z^T \beta_0 + m(X) + q(S)\}. \quad (2)$$

We model the function $m(X)$ via penalized regression splines (Eilers and Marx, 1996; Ruppert 2002). The class of models proposed is a special case of generalized additive models; see Hastie and Tibshirani (1990), who developed a penalized conditional likelihood method for our problem. While the model itself is not new, the methods that we develop for fitting that model and making inference about it do appear to be new.

This article is motivated by an example from equine epidemiology. The syndrome of equine colic (i.e., abdominal pain) is an important cause of disease and death in horses

(Cohen, 1997), frightening for the animal's human companions, and of considerable financial import. Despite the high prevalence of colic, epidemiologic studies of equine colic are limited (Reeves, Gay and Hilbert, 1989; Uhlinger, 1990; Proudman, 1991; Proudman and Edwards, 1993; Cohen, Matejka, and Honnas, 1995; Reeves, Salman, and Smith, 1996; Kaneene, Miller, and Ross, 1997; Tinker, White, and Lessard, 1997a, 1997b; Cohen, Gibbs, and Woods, 1999; Hillyer, Taylor, and French, 2001; Hudson, Cohen, and Gibbs, 2001; Traub-Dargatz, Koprak, and Seitzinger, 2001). In particular, the association with colic of various characteristics of affected horses (e.g., age of the horse) and management practices for these horses has been poorly characterized.

This report utilizes data from a case-control study of the association of various management practices with equine colic. As previously reported (Cohen et al., 1999), veterinarians in private practice in Texas were recruited to participate in a study of the association of equine colic with dietary and other management factors. A letter soliciting participation was mailed to 774 veterinarians. Of 244 willing respondents, 145 ultimately contributed data to the study, representing 104 private practices. Participating veterinarians were asked to provide data monthly for one horse treated for colic and one horse that received emergency treatment for any condition other than colic, between March 1, 1997, and February 28, 1998. A case of colic was defined as the first horse treated during a given month for signs of intra-abdominal pain. A control horse was defined as the next horse that received emergency treatment for any condition other than colic and was treated by the veterinarian who treated the horse with colic. Controls

were examined no more than 30 days after the corresponding colic case, to avoid any seasonal bias in choosing the comparison population, both because incidence of colic is considered to be seasonal, and because feeding and other management practices of horses vary by season. Horses were functionally matched on the basis of month, veterinarian, and region of Texas. Because some participating veterinarians were employed in the same practice, matched case-control pairs were often contributed by practices, rather than cases and controls matched by individual veterinarian. For purposes of this study, we examined only the 498 matched pairs (996 horses) contributed by the 145 participating veterinarians in which the case-control sets were matched by individual veterinarian. Data collected for cases and controls included information regarding identifiers for the horse, farm, veterinarian, date of examination, age, breed, sex, and various management factors (including dietary practices).

While the case-control study itself was performed to assess the effects of management strategy, a related question is whether and how the age of the horse is associated with the risk of colic. There is some evidence in the literature of an age effect (Reeves et al., 1989; Uhlinger, 1990; Proudman, 1991; Cohen, et al., 1995; Reeves et al., 1996; Kaneene et al., 1997; Cohen et al., 1999; Traub-Dargatz et al., 2001), so that age is a potential confounding factor that needs to be accounted for in our analysis.

Our goal was to understand the shape of the age effect in a detailed way. Previous analysis of data such as this have either been with a linear age effect, which we found *a priori* implausible, or categorical. Our approach, based on model (2), is to fit a quadratic regression spline to the data. The analysis is given in Section 5, where we found that the shape of the age effect was clearly not linear, and suggested instead a rapid increase in risk for young horses followed by a near flattening out of the risk shape for older horses.

We consider several methods to fit our semiparametric regression spline model (2) in the matched case-control study. These methods will be described in detail in Section 2, but for now, we note that, in general, they depend on a smoothing parameter, which can be estimated either by cross-validation or using a mixed model formulation (Coull, Ruppert, and Wand, 2001). In the latter case, certain of the regression spline parameters are treated as if they are random, with a variance related to the smoothing parameter.

Our first approach is to use cross-validation to choose the smoothing parameter. Rather than use the computationally expensive direct cross-validation, we instead develop a computationally convenient approximate method, based on an expansion for the leave-one-out method. Previously, O'Sullivan (1988) proposed an approximate cross-validation method in a different context, while Joly, Commenges, and Letenneur (1998) applied and changed the form of the O'Sullivan method to fit their problem. There are technical differences in our approaches, largely based on the idea that we did not approximate the Hessian for the leave-one-out estimator with the Hessian from using all data, as it is done in O'Sullivan's article (see the Appendix). At least in our simulations, these extra details lead to a significant improvement in efficiency, albeit at the cost of slightly more programming.

The second approach is to start with a mixed model formulation, where, as described above, certain parameters of the regression spline are treated as if they were random. We first develop a Monte Carlo EM (MCEM) algorithm in this context; see McCulloch (1997). Our third approach is the natural Bayesian counterpart to the MCEM algorithm, using Gibbs sampling.

The article is organized as follows. In Section 2, we describe the matched case-control study and regression splines. In Section 3, we discuss our methods to fit this semiparametric regression spline model: penalized regression splines, MCEM, and a Bayesian method using MCMC. In Section 4, we report the results of simulations comparing the various methods. Our simulations suggest that they appear approximately efficient in terms of mean squared error, with the approximate cross-validation method being computationally the fastest, while the Bayesian method carries with it posterior credible intervals. In Section 5, we apply our approaches to the equine epidemiology example that motivated this work. Section 6 contains concluding remarks.

2. Regression Splines

We now define a p th-order regression spline with a truncated power series basis. Define $C_1(X) = \{X, \dots, X^p\}^T$ and $C_2(X) = \{(X - \xi_1)_+^p, \dots, (X - \xi_k)_+^p\}^T$, where $(u)_+^p = u^p I(u \geq 0)$. The knots $\xi_1 < \dots < \xi_k$ can be selected *a priori*, e.g., at sample quantiles of observed X 's which are centered; see Ruppert (2002), who suggests that the number of knots is not crucial. We let $\beta_1 = (\beta_{11}, \beta_{12}, \dots, \beta_{1p})^T$ be the regression coefficients of the polynomial function, $\beta_2 = (\beta_{21}, \dots, \beta_{2k})^T$ be regression coefficients of the truncated power series basis, and $\mathcal{B} = (\beta_0^T, \beta_1^T, \beta_2^T)^T$ be a vector of regression coefficients. With this setup, model (2) becomes

$$\Pr(Y = 1 | Z, X, S) = H \{ Z^T \beta_0 + C_1^T(X) \beta_1 + C_2^T(X) \beta_2 + q(S) \}.$$

We will assume 1-1 matching; the results are easily generalized to 1- m matching. Retrospectively, the conditional likelihood is that of (Y_1, Y_2) given $Y_1 + Y_2 = 1$. Define $L_1(X_1, X_2) = \{(X_1 - X_2), \dots, (X_1^p - X_2^p)\}^T$ and $L_2(X_1, X_2) = \{(X_1 - \xi_1)_+^p - (X_2 - \xi_1)_+^p, \dots, (X_1 - \xi_k)_+^p - (X_2 - \xi_k)_+^p\}^T$. Then, the retrospective likelihood is

$$\begin{aligned} & \Pr\{Y_1 = 1, Y_2 = 0 | X_1, X_2, Z_1, Z_2, Y_1 + Y_2 = 1, S\} \\ &= H \{ (Z_1 - Z_2)^T \beta_0 + L_1^T(X_1, X_2) \beta_1 + L_2^T(X_1, X_2) \beta_2 \}. \end{aligned} \quad (3)$$

Note how the function $q(\bullet)$ disappears in this formulation.

3. Methods

3.1 Introduction

This section describes our methods. Section 3.2 develops an approximate cross-validation method for a penalized regression splines. Section 3.3 develops the Monte Carlo EM algorithm, while Section 3.4 develops the complementary Bayesian MCMC approach.

3.2 Penalized Regression Splines

Here, we treat β_2 as a fixed parameter and start from model (3). Naturally, as for all splines, some penalty must be imposed in order to achieve smoothness. We penalize the likelihood as follows.

Let α be a smoothing parameter and D be the $(1 + p + k) \times (1 + p + k)$ matrix with $1 + p$ zeros followed by k ones along the diagonal. Define $\mathcal{L}(Z_{i1}, Z_{i2}, X_{i1}, X_{i2}) = \{(Z_{i1} - Z_{i2})^T, L_1^T(X_{i1}, X_{i2}), L_2^T(X_{i1}, X_{i2})\}^T$. Define $\hat{\mathcal{B}}(\alpha)$ to be the penalized least-square estimator that maximizes the penalized log likelihood

$$\begin{aligned} \log l(\mathcal{B}) &= \sum_{i=1}^n \log [H\{(Z_{i1} - Z_{i2})^T \beta_0 + L_1^T(X_{i1}, X_{i2}) \beta_1 \\ &\quad + L_2^T(X_{i1}, X_{i2}) \beta_2\}] - \frac{\alpha}{2} \mathcal{B}^T D \mathcal{B} \\ &= \sum_{i=1}^n \log [H\{\mathcal{L}^T(Z_{i1}, Z_{i2}, X_{i1}, X_{i2}) \mathcal{B}\}] - \frac{\alpha}{2} \mathcal{B}^T D \mathcal{B} \\ &= \sum_{i=1}^n l_i(\mathcal{B}) - \frac{\alpha}{2} \mathcal{B}^T D \mathcal{B}. \end{aligned} \tag{4}$$

The form of the penalty term is standard for the truncated power series basis, see Ruppert and Carroll (2000) and Ruppert, Wand, and Carroll (2003).

Newton-Raphson was used to estimate $\hat{\mathcal{B}}(\alpha)$ for each α . To estimate the smoothing parameter, one possible method is cross-validation. Let $\hat{\mathcal{B}}_{(i)}(\alpha)$ maximize the corresponding version of (4), but with the i th observation deleted. Cross-validation would maximize $\sum_{i=1}^n l_i\{\hat{\mathcal{B}}_{(i)}(\alpha)\}$ in α . The difficulty with this is, of course, computational: n different analyses for each value of α .

To avoid such lengthy computation, we developed an approximate cross-validation method in which only the complete data estimate, $\hat{\mathcal{B}}(\alpha)$, need be computed. Specifically, we find an approximation $\hat{\mathcal{B}}_{(i)}^*\{\alpha, \hat{\mathcal{B}}(\alpha)\} = \hat{\mathcal{B}}_{(i)}^*(\alpha)$ that differs from $\hat{\mathcal{B}}_{(i)}$ by term of order $O_p(n^{-2})$. We then maximize in α the approximate penalized log likelihood $\sum_{i=1}^n l_i\{\hat{\mathcal{B}}_{(i)}^*(\alpha)\}$. The actual formulae are fairly lengthy, although routine to program. They are given in Appendix A.1, along with the technical derivation of the method. We have already commented upon the idea that, unlike O’Sullivan (1988), we argue that the Hessian for leave-one-out estimator cannot be replaced by the Hessian of estimator without considering convergence rate, which then results in a more complicated form (A.2).

We also explain in Section A.1 how our approximate cross-validation methodology can be generalized to 1- m matched case-control studies.

3.3 Monte Carlo EM

Recall that k is the number of knots and p is the order of the regression spline. As in Coull et al. (2001), to smooth the regression spline, we will use the mixed model formulation with $\beta_2 \sim N(0, \sigma^2 I_k)$, the standard prior for the truncated power series basis.

Let $\theta = (\beta_0, \beta_1, \sigma^2)$ be the parameter vector, $Y_{\text{mis}} = \beta_2$ be the missing data, Y_{obs} be the observed data, and $Y = (Y_{\text{mis}}, Y_{\text{obs}})$ be the complete data. We let $f(\beta_2 | \sigma^2)$ be the probability density function of the multivariate normal distribution of β_2 , with mean 0 and covariance matrix $\sigma^2 I_k$.

In this case, the likelihood of the complete data is

$$\begin{aligned} &f\{(Y_{i1}, Y_{i2})_{i=1}^n | \beta_2, (Z_{i1}, Z_{i2})_{i=1}^n, \\ &\quad (X_{i1}, X_{i2})_{i=1}^n, (Y_{i1} + Y_{i2} = 1)_{i=1}^n\} f(\beta_2 | \sigma^2) \\ &= f(\beta_2 | \sigma^2) \prod_{i=1}^n H\{(Z_{i1} - Z_{i2})^T \beta_0 \\ &\quad + L_1^T(X_{i1}, X_{i2}) \beta_1 + L_2^T(X_{i1}, X_{i2}) \beta_2\}. \end{aligned}$$

Except for a constant, the log likelihood is

$$\begin{aligned} l(Y | \beta_0, \beta_1, \beta_2, \sigma^2) &= \sum_{i=1}^n \log [H\{(Z_{i1} - Z_{i2})^T \beta_0 \\ &\quad + L_1^T(X_{i1}, X_{i2}) \beta_1 + L_2^T(X_{i1}, X_{i2}) \beta_2\}] - \frac{\kappa \log \sigma^2}{2} - \frac{\beta_2^T \beta_2}{2\sigma^2}. \end{aligned}$$

The E-Step requires computation of $Q(\theta) = E_{\theta_{\text{old}}}\{l(Y | \theta) | Y_{\text{obs}}\}$, while the M-step involves an update of parameter estimates via maximization: $\theta_{\text{new}} = \text{argmax}_{\theta} Q(\theta | \theta_{\text{old}})$. However, computation of the expectation in the E-step involves intractable integrals. Hence, Monte Carlo methods are employed to handle the integrations (McCulloch, 1997). First, we choose initial values of $\beta_{0,\text{old}}$, $\beta_{1,\text{old}}$ and σ_{old}^2 as values from the quadratic regression spline. Then we conduct the following steps until convergence:

Step 1. We generate β_2 ’s from $[\beta_2 | (Z_{i1}, Z_{i2})_{i=1}^n, (X_{i1}, X_{i2})_{i=1}^n, \beta_{0,\text{old}}, \beta_{1,\text{old}}, \sigma_{\text{old}}^2, \{Y_{i1} + Y_{i2} = 1\}_{i=1}^n]$ using the Metropolis-Hastings algorithm. Since the posterior distribution $[\beta_2 | \text{rest}]$ is proportional to

$$\begin{aligned} &\prod_{i=1}^n H\{(Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} + L_1^T(X_{i1}, X_{i2}) \beta_{1,\text{old}} \\ &\quad + L_2^T(X_{i1}, X_{i2}) \beta_2\} f(\beta_2 | \sigma_{\text{old}}^2), \end{aligned} \tag{5}$$

and (5) is a unimodal and log-concave function in β_2 , we choose the multivariate normal distribution with mean $\beta_{2,\text{old}}$ and covariance $\Sigma_{\beta_2} = \sigma_{\text{old}}^2 I_k$ as a proposal distribution. The proof of log-concavity of (5) is given in Section A.2. Choosing this proposal distribution is based on work of Chib and Greenberg (1995).

Step 2. The E-step can then be easily computed as

$$\begin{aligned} &\hat{Q}(\beta_0, \beta_1, \sigma^2 | \beta_{0,\text{old}}, \beta_{1,\text{old}}, \sigma_{\text{old}}^2) \\ &= -\frac{k \log \sigma^2}{2} - M^{-1} \sum_{j=1}^M \frac{\beta_2^{(j)T} \beta_2^{(j)}}{2\sigma^2} \\ &\quad + M^{-1} \sum_{j=1}^M \sum_{i=1}^n \log [H\{(Z_{i1} - Z_{i2})^T \beta_0 \\ &\quad + L_1^T(X_{i1}, X_{i2}) \beta_1 \\ &\quad + L_2^T(X_{i1}, X_{i2}) \beta_2^{(j)}\}]. \end{aligned}$$

Step 3. The M-step consists of estimating

$$\sigma_{\text{new}}^2 = M^{-1} k^{-1} \sum_{j=1}^M \beta_2^{(j)T} \beta_2^{(j)},$$

and estimating $(\beta_{0,\text{new}}, \beta_{1,\text{new}})$ by maximizing

$$M^{-1} \sum_{j=1}^M \sum_{i=1}^n \log \left[H \{ (Z_{i1} - Z_{i2})^T \beta_0 + L_1^T(X_{i1}, X_{i2}) \beta_1 + L_2^T(X_{i1}, X_{i2}) \beta_2^{(j)} \} \right]$$

using the Newton-Raphson method.

Unlike McCulloch (1997), or Booth and Hobert (1998, 1999), who recommend increasing the Monte-Carlo sample size M as a function of the EM iteration, we instead simply set $M = 12,000$. Our convergence criterion was to stop when the parameter estimates changed by less than 0.001, but we then continued the algorithm to check that convergence was achieved.

3.4 A Bayesian Approach

As for MCEM, we also start with the mixed model formulation by treating β_2 as a random effect, $\beta_2 \sim N(0, \sigma^2 I_k)$. The parameters have prior distributions $(\beta_0, \beta_1) \sim N(0, \sigma_1^2 I_{1+p})$ and $\sigma^2 \sim IG(A, B)$, the inverse-gamma distribution with density

$$\frac{1}{\Gamma(A)B^A} (\sigma^2)^{-(A+1)} \exp\left(\frac{-1}{\sigma^2 B}\right).$$

The prior parameters were chosen to be proper but vague, with $\sigma_1^2 = 100$, $A = 1$, and $B = 1$. We varied (A, B) somewhat, without appreciable change in the results.

The complete conditional distributions are proportional to

$$[\mathcal{B} | \text{rest}] \propto \prod_{i=1}^n H \{ \mathcal{L}^T(Z_{i1}, Z_{i2}, X_{i1}, X_{i2}) \mathcal{B} \} f(\mathcal{B} | \Sigma);$$

$$[\sigma^2 | \text{rest}] \propto \text{IG}\left(A + \frac{\# \text{ of knots}}{2}, \frac{1}{B} + \frac{\beta_2^T \beta_2}{2}\right).$$

We sample $[\mathcal{B} | \text{rest}]$ from the Metropolis-Hastings algorithm, with proposal density being the multivariate normal distribution with mean \mathcal{B}_{old} and covariance matrix $\Sigma_{\text{old}} = \text{diag}(\sigma_1^2 I_{1+p}, \sigma_{\text{old}}^2 I_k)$; this is because the complete conditional distribution of \mathcal{B} is a log-concave function (see Section A.3).

4. Simulation Study

We performed a small simulation study to compare the following methods:

- CLR(L): conditional logistic linear regression model;
- CLR(Q): conditional logistic quadratic regression model, with a linear and a quadratic term;
- RSWO: quadratic regression splines model with 8 knots, without considering a smoothing parameter ($\alpha = 0$);
- PSPLINE: quadratic regression splines with 8 knots, with a smoothing parameter using our new approximate cross-validation;
- MCEM: quadratic regression splines with 8 knots, using MCEM;
- BMCMC: Bayesian quadratic regression splines with 8 knots.

We choose 8 knots for convenience, although spot checks were performed with 20 knots and the results were essentially the same as reported here.

In this simulation study, X was generated from Normal(0,1), Uniform[-2, 2] and Skewed Normal with index $\lambda = 5$. The skewed normal distribution (Azzalini, 1985) has density proportional to $2\phi(x)\Phi(\lambda x)$, where ϕ and Φ represent the standard normal density and distribution. This density is reasonably skewed for any value of $\lambda \geq 5$. For each distribution of X , 100, 250, and 500, matched pairs were generated. The stratum effect $q(S)$ was generated from Normal(0, 1).

For each of the 9 combinations of distributions of X and the number of matched pairs, we simulated 250 data sets. In MCEM and Bayesian methods, a burn-in time of 10,000 observations is followed by 2000 observations from the posterior. While convergence can always be an issue, for the MCEM method, we assessed convergence as described above, while for the MCMC method, we examined trace plots in a number of examples.

We used two regression functions: (a) $m(X) = \sin(\pi X/2)/1 + 2X^2\{\text{sign}(X) + 1\}$, and (b) $m(X) = X^2/2$. Function (a) has significant nonlinearity, and we would expect to see that our methods dominate the other three: CLR(L), CLR(Q), and RWSO. Function (b) was chosen to investigate the loss of efficiency when compared to the correctly parameterized model CLR(Q).

We computed mean square errors on a grid of 101 points in the interval [-2, 2]. The mean square error values are given in Tables 1–2. Figures 1–2 contain the average of the estimated functions over the simulation with 500 matched pairs and X having the Normal(0,1) distribution; the other cases give a similar graphical impression. In Figure 1, we see some bias, especially for MCEM, but as expected, the bias is greatest at the endpoints of the functions.

For the nonlinear function, we expected our methods to completely dominate the three other methods. For the quadratic function, our methods are not much less efficient than the properly parameterized model, suggesting not too great a loss of efficiency compared to fully parametric methods. However, for $n = 500$ with skew-normal X 's, small amounts of bias in the spline methods lead to increases in mean squared errors.

It is striking in Tables 1–2 that our methods seem roughly comparable to one another. We have done a few other simulations and repeated this observation. This suggests that the approximate cross-validation method can be used to give a quick impression of the function, with the MCMC method yielding Bayesian credible intervals.

In addition, we compared our approximate cross-validation approach to GCV, using the form of O'Sullivan (1988) in the following case: $X = \text{Normal}(0, 1)$ and $n = 500$. When $m(X) = \sin(\pi X/2)/1 + 2X^2\{\text{sign}(X) + 1\}$, the mean square of error of GCV was 0.0810, while when $m(X) = X^2/2$, the MSE of GCV was 0.0399. In both cases, these mean squared errors are significantly larger than the values displayed in Tables 1 and 2.

5. Example: Equine Epidemiology

Because of the statistically significant quadratic effect reported in the introduction, we fit a semiparametric quadratic ($p = 2$) regression spline with, $Z =$, a change of diet, $X =$ horse

Table 1

Mean square error value of each method. In this simulation, the true regression spline function is $m(X) = \sin(\pi X/2)/1 + 2X^2\{\text{sign}(X) + 1\}$.

Distribution of X	Number of matched pairs	CLR(L) ^a MSE	CLR(Q) ^b MSE	RSWO ^c MSE	PSPLINE ^d MSE	MCEM ^e MSE	BMCMC ^f MSE
Normal	100	0.1994	0.3081	0.5002	0.0719	0.0722	0.0751
	250	0.1794	0.2046	0.4411	0.0672	0.0699	0.0710
	500	0.1715	0.2025	0.4398	0.0624	0.0687	0.0602
Uniform	100	0.1871	0.2440	0.1722	0.0899	0.0879	0.0858
	250	0.1769	0.1865	0.1578	0.0831	0.0824	0.0723
	500	0.1703	0.1755	0.1513	0.0792	0.0722	0.0680
Skewed Normal	100	0.2287	0.3824	0.2721	0.1711	0.1829	0.1677
	250	0.1975	0.2577	0.2159	0.1102	0.1236	0.1081
	500	0.1944	0.2479	0.2091	0.1028	0.1109	0.0945

^a CLR(L) = method fitting conditional logistic linear regression model.
^b CLR(Q) = method fitting conditional logistic quadratic regression model.
^c RSWO = method fitting quadratic regression splines model without considering smoothing parameter.
^d PSPLINE = method fitting quadratic regression splines model with considering smoothing parameter.
^e MCEM = method fitting quadratic regression splines model via Monte Carlo EM algorithm.
^f BMCMC = method fitting quadratic regression splines model via MCMC.

age and $k = 20$ knots selected at the quantiles of the distribution of age. The same priors, computational settings, etc., were used as described previously.

The fitted regression spline functions via penalized regression splines, MCEM, and the Bayesian method are given in Figure 3, where we also display a 95% posterior credible interval. Figure 4 gives a comparison with the quadratic fit. As expected from our simulations, Figure 3 suggests that all of our methods give essentially the same fits. Figure 4 shows that the fits are also similar to the quadratic regression case. We confirmed this using various starting values in the algorithms. The basic conclusion here epidemiologically is that the risk of colic rises rapidly with horse age in younger horses, but appears to level off, or even decrease, as the horse becomes 20+ years old.

From this result, our example seems not to show the advantage of using splines, since a quadratic fit does just

as well as the spline. We, of course, have no control over the results of the example. However, simulations such as those described in Table 1 do indicate that, if the model were in fact nonquadratic, then our method would have the chance to capture this; yet, as shown in this example and in Table 2, it would incur little cost were the function to be quadratic.

Results of the association of colic with age have varied among equine epidemiologic studies (Reeves et al., 1989; Uhlinger, 1990; Proudman, 1991; Cohen et al., 1995; Reeves et al., 1996; Kaneene et al., 1997; Cohen et al., 1999; Traub-Dargatz et al., 2001). These conflicting results may be attributable to differences in study designs and study populations. It is also possible that the difference is attributable to differences among studies in the methods used to characterize the outcome of age. Age was often—but not exclusively—considered in analyses as a categorical factor, and age

Table 2

Mean square error value of each method. In this simulation, the true regression spline function is $m(X) = X^2/2$.

Distribution of X	Number of matched pairs	CLR(L) ^a MSE	CLR(Q) ^b MSE	RSWO ^c MSE	PSPLINE ^d MSE	MCEM ^e MSE	BMCMC ^f MSE
N(0,1)	100	0.8658	0.1734	0.2263	0.1869	0.1899	0.1872
	250	0.8430	0.0577	0.0938	0.0641	0.0677	0.0611
	500	0.8373	0.0252	0.1281	0.0299	0.0311	0.0328
U(-2,2)	100	0.8586	0.1214	0.1767	0.1329	0.1389	0.1373
	250	0.8412	0.0452	0.0912	0.0522	0.0504	0.0552
	500	0.8361	0.0193	0.1152	0.0213	0.0206	0.0283
Skewed Normal	100	0.9343	0.1809	0.4184	0.1909	0.1930	0.1912
with $\lambda = 5$	250	0.9199	0.0746	0.3259	0.0801	0.0799	0.0810
	500	0.9023	0.0350	0.3404	0.0499	0.0484	0.0502

^a CLR(L) = method fitting conditional logistic linear regression model.
^b CLR(Q) = method fitting conditional logistic quadratic regression model.
^c RSWO = method fitting quadratic regression splines model without considering smoothing parameter.
^d PSPLINE = method fitting quadratic regression splines model with considering smoothing parameter.
^e MCEM = method fitting quadratic regression splines model via Monte Carlo EM algorithm.
^f BMCMC = method fitting quadratic regression splines model via MCMC.

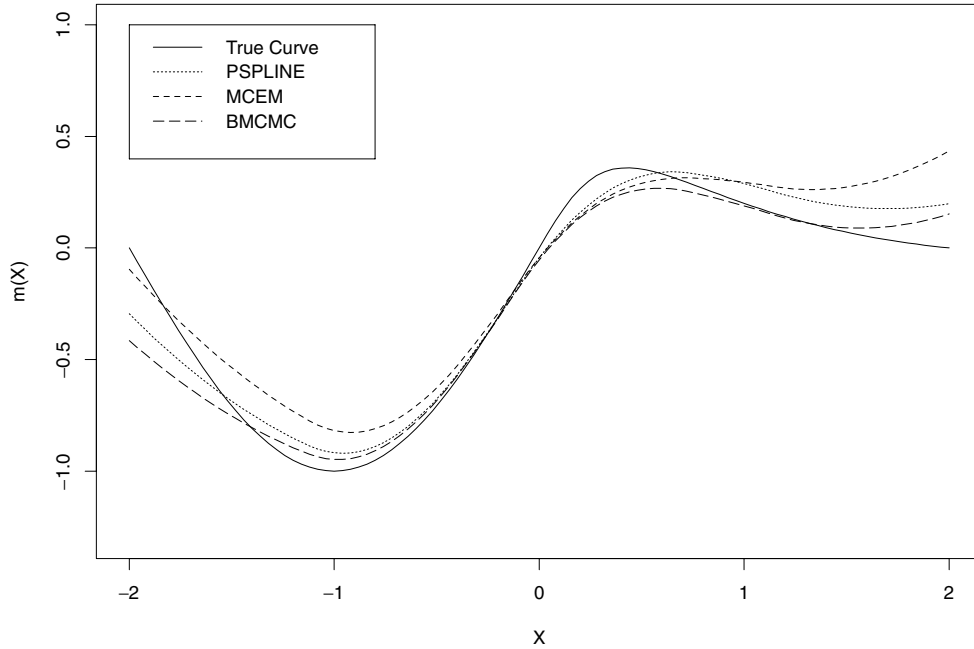


Figure 1. 250 data sets were simulated with $X \sim \text{Normal}(0,1)$ and 500 matched pairs which were generated from $m(X) = \sin(\pi X/2)/1 + 2X^2\{\text{sign}(X) + 1\}$. The fitted regression spline functions are created by averaging estimators from 250 simulated data sets.

categories differed among studies. Results of our study indicate that the estimated risk of colic increases rapidly with age among younger horses, but tended to level off or even decrease after age 20. Efforts to characterize age using linear methods of analysis cannot, of course, be expected to capture such a pattern, while the use of age categories can tend to obscure rather than illuminate.

6 Discussion

In the article, we have developed semiparametric regression spline methods in matched case-control studies. We used penalized regression splines and developed a new cross-validation method using an expansion for leave-one-out estimation. We also developed an MCEM algorithm and a Bayesian MCMC approach.

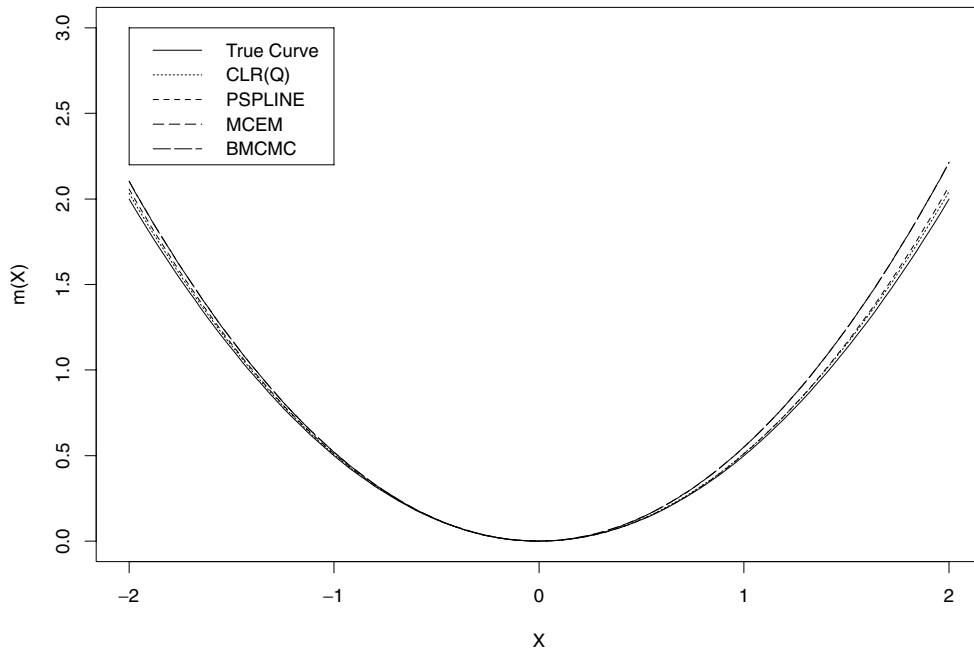


Figure 2. 250 data sets were simulated with $X \sim \text{Normal}(0,1)$ and 500 matched pairs which were generated from $m(X) = X^2/2$. The fitted regression spline functions are created by averaging estimators from 250 simulated data sets.

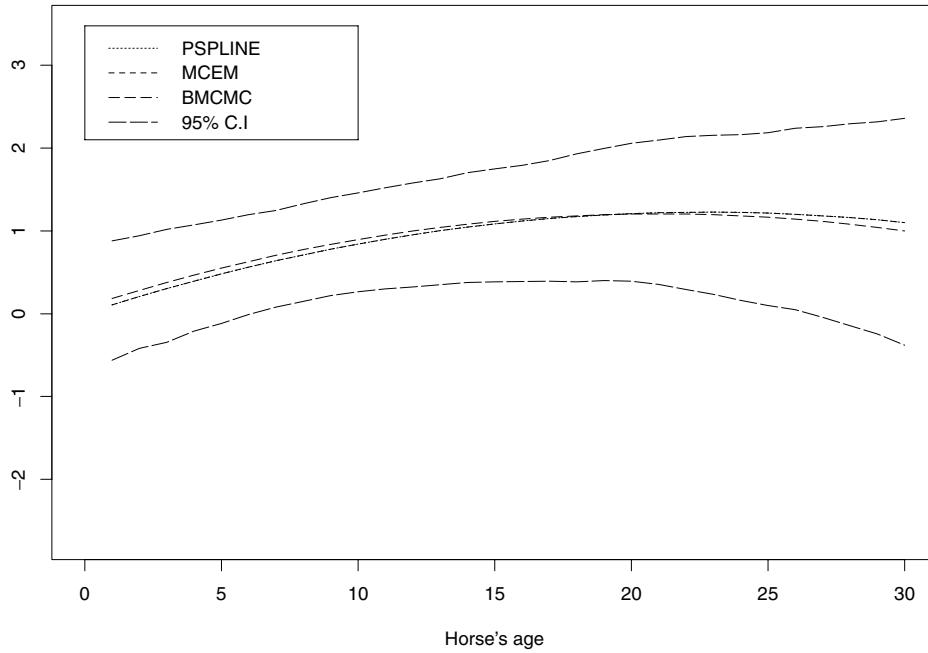


Figure 3. Fitted regression splines functions and 95% credible interval in the equine colic example.

Our simulation study suggests that the approximate cross-validation method can be as good as the MCEM and Bayesian approaches, and it is computationally fast. We make no claim to global efficiency of the approximate cross-validation method, but in all of our examples, it has performed well. At the very least, it can serve to define starting values for the MCEM and Bayesian approaches. Infer-

ence for splines is complicated by the fact (Ruppert et al., 2003, pp. 141–142) that undercoverage can be expected in regions where the function is varying rapidly. We tend to favor the use of pointwise Bayesian credible intervals, since the Bayesian machinery accounts naturally for the estimation of the smoothing parameters. In our simulations, we evaluated 95% credible regions, pointwise at $-2.0, -1.5, \dots, 2.0$,

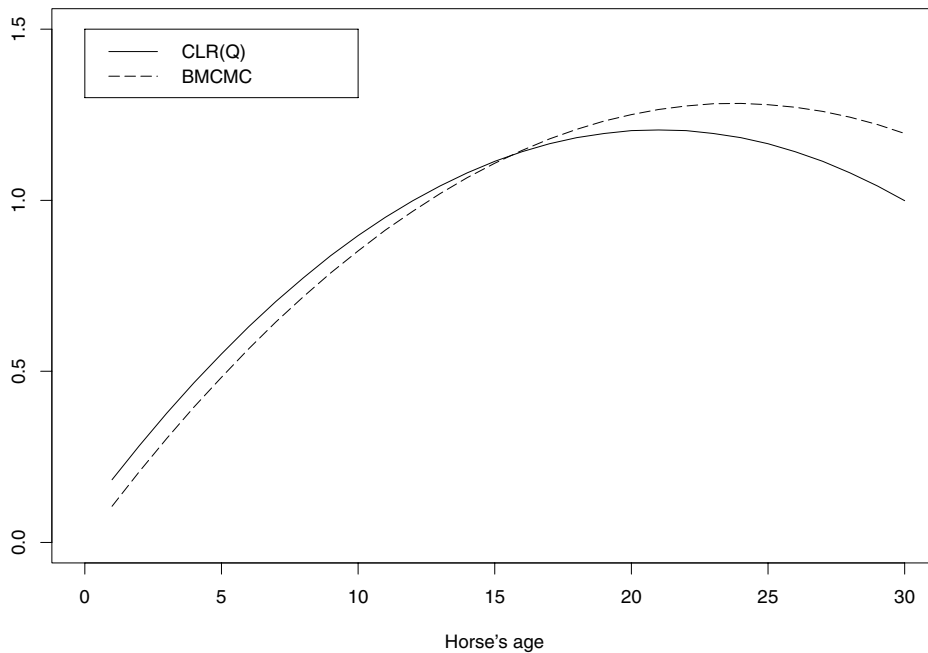


Figure 4. Fitted quadratic regression model along with the Bayes regression spline function in the equine colic example.

with coverage probabilities mostly near nominal, and all above 92%.

We can extend our methods to semiparametric regression spline models in any type of matched case-controls studies. In Section A.1.3, we explain how our approximate cross-validation can be generalized to 1- m matching.

Finally, we note a subtlety with the definition of the MCEM and MCMC approaches. Specifically, we have assumed that the random effect β_2 is normally distributed in the matched study, i.e., after conditioning. An alternative would have been to assume that β_2 is normally distributed prospectively. The difficulty with this approach is that unconditionally, the nuisance function $q(\bullet)$ does not disappear upon conditioning. The same sort of phenomenon is known to happen with likelihood approaches to matched case-control studies with missing data. In principle, one could model the nuisance function, although this may not be easy or practical. An alternative is to model the distribution of β_2 among the controls only, wherein the nuisance function disappears; see Satten and Kupper (1993), Paik and Sacco (2000), and Satten and Carroll (2000) for calculations in the missing data context. However, the cost of this is that the model must include the stratification variables, if they have any effect, and in the context of regression spline smoothing it is not clear whether such an approach makes any sense.

ACKNOWLEDGEMENTS

Our research was supported by a grant from the National Cancer Institute (CA57030), by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106), and by the Link Equine Research Fund at Texas A&M University. We thank the associate editor and two referees for their many helpful comments.

RÉSUMÉ

Nous développons des méthodes semi-paramétriques pour des études cas-témoin appariés, en utilisant les splines de régression. Trois méthodes sont développées : un schéma approximatif de validation croisée pour estimer le paramètre de lissage inhérent aux splines de régression, et également des méthodes Monte Carlo de type EM (MCEM), et bayésienne pour ajuster le modèle de spline de régression. Nous comparons l'approche approximative de validation croisée, les approches MCEM et bayésienne par simulation, et nous montrons qu'elles apparaissent approximativement de même efficacité, la méthode approximative par validation croisée semblant la plus intéressante sur le plan des calculs. Un exemple d'épidémiologie équine ayant motivé ce travail, est utilisé pour expliciter nos approches.

REFERENCES

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178.
- Booth, J. G. and Hobert, J. H. (1998). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of Royal Statistical Society, Series B* **62**, 265–285.
- Booth, J. G. and Hobert, J. H. (1999). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association* **93**, 262–272.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician* **49**, 327–335.
- Cohen, N. D. (1997). Epidemiology of colic. *Veterinary Clinician North America Equine Practice* **13**, 91–201.
- Cohen, N. D., Matejka, P. L., and Honnas, C. M. (1995). Case-control study of the association between various management factors and development of colic in horses. *Journal of American Veterinary Medical Association* **206**, 667–673.
- Cohen, N. D., Gibbs, P. G., and Woods, A. M. (1999). Dietary and other management factors associated with colic in horses. *Journal of American Veterinary Medical Association* **215**, 53–60.
- Coull, B. A., Ruppert, D., and Wand, M. P. (2001). Simple incorporation of interactions into additive models. *Biometrics* **57**, 539–545.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science* **11**, 89–121.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, 205–208. Monographs on Statistics and Applied Probability, 43. New York: Chapman and Hall.
- Hillyer, M. H., Taylor, F. G. R., and French, N. P. (2001). A cross-sectional study of colic in horses on thoroughbred training premises in the British Isles in 1997. *Equine Veterinary Journal* **33**, 380–385.
- Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley, New York.
- Hudson, J. M., Cohen, N. D., and Gibbs, P. (2001). Feeding practices associated with colic in horses. *Journal of the American Veterinary Medical Association* **219**, 1419–1425.
- Joly, P., Commenges, D., and Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics* **54**(1), 185–194.
- Kaneene, J. B., Miller, R., and Ross W. A. (1997). Risk factors for colic in the Michigan (U.S.A.) equine population. *Preventive Veterinary Medicine* **30**, 23–36.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.
- O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal of Scientific and Statistical Computing* **9**, 363–379.
- Paik, M. C. and Sacco, R. L. (2000). Matched case-control data analyses with missing covariates. *Applied Statistics* **49**, 145–156.
- Proudman, C. J. (1991). A two year, prospective survey of equine colic in general practice. *Equine Veterinary Journal* **24**, 90–93.
- Proudman, C. J. and Edwards, G. B. (1993). Are tapeworms associated with equine colic? A case control study. *Equine Veterinary Journal* **25**, 224–226.
- Reeves, M. J., Gay, J. M., and Hilbert, B. J. (1989). Association of age, sex and breed factors in acute equine colic: A retrospective study of 320 cases admitted to a

veterinary teaching hospital in the U.S.A. *Preventive Veterinary Medicine* **7**, 149–160.

Reeves, M. J., Salman, M. D., and Smith, G. (1996). Risk factors for equine acute abdominal disease (colic): results from a multi-center case-control study. *Preventive Veterinary Medicine* **26**, 285–301.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**, 735–757.

Ruppert, D. and Carroll, R. J. (2000). Specially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics* **42**, 205–223.

Ruppert D., Wand, M. P., and Carroll, R. J. (2003). *Semi-parametric Regression*. Cambridge: Cambridge University Press.

Satten, G. A. and Carroll, R. J. (2000). Conditional and unconditional categorical regression models with missing covariates. *Biometrics* **56**, 384–388.

Satten, G. A. and Kupper, L. L. (1993). Inferences about exposure-disease association using probability of exposure information. *Journal of the American Statistical Association* **88**, 200–208.

Tinker, M. K., White, N. A., and Lessard P. (1997a). Prospective study of equine colic incidence and mortality. *Equine Veterinary Journal* **29**, 448–453.

Tinker, M. K., White, N. A., and Lessard P. (1997b). Prospective study of equine colic risk factors. *Equine Veterinary Journal* **29**, 454–458.

Traub-Dargatz, J. L., Koprak, C. A., and Seitzinger, A. H. (2001). Estimate of the national incidence of and operation-level risk factors for colic among horses in the United States, spring 1998 to spring 1999. *Journal of the American Veterinary Medical Association* **219**, 67–71.

Uhlinger, C. (1990). Effects of three anthelmintic schedules on the incidence of colic in horses. *Equine Veterinary Journal* **22**, 251–254.

Received October 2002. Revised June 2003.
Accepted June 2003.

APPENDIX

A.1 Approximate Cross-Validation

A.1.1 Definition of the method. The approximate cross-validation method is defined as follows. Let C_n^m be the component of m th row and n th column of matrix C and let \hat{B}^n be the n th row of vector \hat{B} . We will suppose dependence on the smoothing parameter α , unless specified otherwise.

Further define

$$H_B(\bullet) = \frac{\partial}{\partial B} H \{ \mathcal{L}^T(Z_{i1}, Z_{i2}, X_{i1}, X_{i2})B \} = \frac{\partial}{\partial B} H(\bullet);$$

$$F_i(B) = \frac{H_B(\bullet)}{H(\bullet)};$$

$$\mathcal{F}(B) = \sum_i F_i(B);$$

$$c_i(B) = \frac{H_B(\bullet)}{[H(\bullet)\{1 - H(\bullet)\}]^{1/2}};$$

$$C_i(B) = c_i(B)c_i(B)^T;$$

$$\mathcal{C}(B) = \sum_i C_i(B);$$

$$\Lambda_j = \left(\frac{\partial C_j(B)_n^m}{\partial B} \Big|_{B=\hat{B}} (\hat{B}_{(i)} - \hat{B}) \right) = (\Lambda_j^m);$$

$$\Lambda = \sum \Lambda_j.$$

We let $\mathcal{F}_{(i)}$, $\mathcal{C}_{(i)}$, and $\Lambda_{(i)}$ be \mathcal{F} , \mathcal{C} , and Λ , but with the i th observation deleted, respectively. The penalized estimator $\hat{B}(\alpha)$ maximizes (4). The gradient is exactly $\mathcal{F}(B) - \alpha DB$, while the method of scoring Hessian is $-\{\mathcal{C}(B) + \alpha D\}$.

Using the method of scoring, we see that the solution \hat{B} must satisfy

$$\hat{B} = \{\mathcal{C}(\hat{B}) + \alpha D\}^{-1} \{\mathcal{F}(\hat{B}) + \mathcal{C}(\hat{B})\hat{B}\}.$$

Then,

$$\hat{B}_{(i)} = \{\mathcal{C}_{(i)}(\hat{B}_{(i)}) + \alpha D\}^{-1} \{\mathcal{F}_{(i)}(\hat{B}_{(i)}) + \mathcal{C}_{(i)}(\hat{B}_{(i)})\hat{B}_{(i)}\}. \quad (A.1)$$

Make the further definitions.

$$\mathcal{E}(\hat{B}) = \left\{ \sum_{j \neq i} C_j(\hat{B}) + \alpha D \right\}^{-1} \sum_{j \neq i} F_j(\hat{B});$$

$$\mathcal{G}(\hat{B}) = \left\{ \sum_{j \neq i} C_j(\hat{B}) + \alpha D \right\};$$

$$\mathcal{K}(\hat{B}) = \sum_{j \neq i} C_j(\hat{B});$$

$$\mathcal{M}(\hat{B}) = \mathcal{G}(\hat{B})^{-1} \mathcal{K}(\hat{B});$$

$$\mathcal{N}(\hat{B}) = \left(\sum_{j \neq i} \sum_k \frac{\partial C_j(B)_k^n}{\partial B^l} \Big|_{B=\hat{B}} \hat{B}^k \right)$$

$$= \left(\sum_k \frac{\partial \mathcal{C}_{(i)}(B)_k^n}{\partial B^l} \Big|_{B=\hat{B}} \hat{B}^k \right) = (\mathcal{N}_l^n);$$

$$\mathcal{P}(\hat{B}) = \left\{ \sum_{j \neq i} C_j(\hat{B}) + \alpha D \right\}^{-1} \left\{ \sum_{j \neq i} F_j(\hat{B}) + \sum_{j \neq i} C_j(\hat{B})\hat{B} \right\};$$

$$\mathcal{Q}(\hat{B}) = \left(\sum_{j \neq i} \sum_k \frac{\partial C_j(B)_k^n}{\partial B^l} \Big|_{B=\hat{B}} \hat{P}(\hat{B})^k \right)$$

$$= \left(\sum_k \frac{\partial \mathcal{C}_{(i)}(B)_k^n}{\partial B^l} \Big|_{B=\hat{B}} \mathcal{P}(\hat{B})^k \right) = (\mathcal{Q}_l^n);$$

$$\mathcal{R}(\hat{B}) = \left(\sum_{j \neq i} \sum_l \frac{\partial C_j(B)_k^n}{\partial B^l} \Big|_{B=\hat{B}} \hat{B}^l \right)$$

$$= \left(\sum_l \frac{\partial \mathcal{C}_{(i)}(B)_k^n}{\partial B^l} \Big|_{B=\hat{B}} \hat{B}^l \right) = (\mathcal{R}_k^n).$$

In what follows, we will write \mathcal{F} for $\mathcal{F}\{\hat{B}(\alpha)\}$ and similarly for other terms.

Our approximation is

$$\widehat{\mathcal{B}}_{(i)} \approx \widehat{\mathcal{B}}_{(i)}^* = (I - 2\mathcal{M} - \mathcal{G}^{-1}\mathcal{N} + \mathcal{G}^{-1}\mathcal{Q})^{-1} \times (\mathcal{E} - \mathcal{M}\widehat{\mathcal{B}} - \mathcal{G}^{-1}\mathcal{R}\widehat{\mathcal{B}} + \mathcal{G}^{-1}\mathcal{R}\mathcal{P}). \tag{A.2}$$

A.1.2 *Justification of (A.2).* Using first-order approximations of \mathcal{C} and \mathcal{F} , and using the facts that $\mathcal{C} = O_p(n)$, $\mathcal{F} = O_p(n)$ and $\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}} = O_p(n^{-1})$, we can obtain

$$\begin{aligned} \mathcal{C}_{(i)}(\widehat{\mathcal{B}}_{(i)}) &= \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \Lambda_{(i)} + O_p(n^{-1}); \\ \mathcal{F}_{(i)}(\widehat{\mathcal{B}}_{(i)}) &= \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) + O_p(n^{-1}). \end{aligned} \tag{A.3}$$

Plugging in (A.3) into (A.1), we find that

$$\begin{aligned} \{\mathcal{C}_{(i)}(\widehat{\mathcal{B}}_{(i)}) + \alpha D\}^{-1} &= \left\{ \sum_{j \neq i} C_j + \alpha D + \Lambda_{(i)} + O_p(n^{-1}) \right\}^{-1} \\ &= \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} - \{\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}\} + \mathcal{G}^{-1}O_p(n^{-1})\mathcal{G}^{-1} \\ &= \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} - \{\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}\} + O_p(n^{-3}). \end{aligned} \tag{A.4}$$

Using (A.3), we find that

$$\begin{aligned} \mathcal{F}_{(i)}(\widehat{\mathcal{B}}_{(i)}) + \mathcal{C}_{(i)}(\widehat{\mathcal{B}}_{(i)})\widehat{\mathcal{B}}_{(i)} &= \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \\ &\quad + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}}_{(i)} + \Lambda_{(i)}\widehat{\mathcal{B}}_{(i)} + O_p(n^{-1}). \end{aligned} \tag{A.5}$$

Now, we are going to multiply (A.4) by (A.5), keeping terms up to $O_p(n^{-1})$ with an error of $O_p(n^{-2})$. Then we have that (A.1) is

$$\begin{aligned} &\left[\left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) + O_p(n^{-3}) \right] \\ &\quad \times \left\{ \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \right. \\ &\quad \left. + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}}_{(i)} + \Lambda_{(i)}\widehat{\mathcal{B}}_{(i)} + O_p(n^{-1}) \right\} \end{aligned}$$

$$\begin{aligned} &= \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} \left\{ \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}}_{(i)} \right\} \\ &\quad + \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \right\} \\ &\quad + \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} (\Lambda_{(i)}\widehat{\mathcal{B}}_{(i)}) \\ &\quad - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) \left\{ \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}}_{(i)} \right\} \\ &\quad - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \right\} \\ &\quad - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1})(\Lambda_{(i)}\widehat{\mathcal{B}}_{(i)}) + O_p(n^{-2}). \end{aligned}$$

The following step is the key step in our method. If we compute terms up to order $O_p(n^{-1})$ with an error $O_p(n^{-2})$, we can replace $\widehat{\mathcal{B}}_{(i)}$ by $\widehat{\mathcal{B}}$ whenever $\widehat{\mathcal{B}}_{(i)}$ is multiplied by a term of order $O_p(n^{-1})$; for example,

$$\begin{aligned} O_p(n^{-1})\widehat{\mathcal{B}}_{(i)} &= O_p(n^{-1})\widehat{\mathcal{B}} + O_p(n^{-1})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \\ &= O_p(n^{-1})\widehat{\mathcal{B}} + O_p(n^{-2}). \end{aligned}$$

Hence, we can replace only these two terms:

$$\begin{aligned} &\left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} (\Lambda_{(i)}\widehat{\mathcal{B}}_{(i)}) \\ &= \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} (\Lambda_{(i)}\widehat{\mathcal{B}}) + O_p(n^{-2}); \\ &(\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}}_{(i)} \right\} \\ &= (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}} \right\} + O_p(n^{-2}). \end{aligned}$$

Therefore, (A.1) is given by

$$\begin{aligned} &\left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} \left\{ \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}}_{(i)} \right\} \\ &\quad + \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \right\} \\ &\quad + \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}}) + \alpha D \right\}^{-1} (\Lambda_{(i)}\widehat{\mathcal{B}}) \end{aligned}$$

$$\begin{aligned}
 & - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) \left\{ \sum_{j \neq i} F_j(\widehat{\mathcal{B}}) + \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})\widehat{\mathcal{B}} \right\} \right. \\
 & - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1}) \left\{ \sum_{j \neq i} C_j(\widehat{\mathcal{B}})(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}) \right\} \\
 & - (\mathcal{G}^{-1}\Lambda_{(i)}\mathcal{G}^{-1})(\Lambda_{(i)}\widehat{\mathcal{B}}_{(i)}) + O_p(n^{-2}) \\
 & = A_1 + A_2 + A_3 - A_4 - A_5 - A_6 + O_p(n^{-2}).
 \end{aligned}$$

Since A_5 and A_6 have order $O_p(n^{-2})$, we obtain

$$\begin{aligned}
 \widehat{\mathcal{B}}_{(i)} & = \mathcal{E} + \mathcal{G}^{-1}\mathcal{K}\widehat{\mathcal{B}}_{(i)} + \mathcal{G}^{-1}\{\mathcal{K}(\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}})\} \\
 & + \mathcal{G}^{-1}\Lambda_{(i)}\widehat{\mathcal{B}} - \mathcal{G}^{-1}\Lambda_{(i)}\mathcal{P} + O_p(n^{-2}), \tag{A.6}
 \end{aligned}$$

To obtain the solution $\widehat{\mathcal{B}}_{(i)}$ from (A.6), we let $\mathcal{C}(\mathcal{B})_{nl}^m$ be the derivative of the component of the m th row and n th column of $\mathcal{C}(\mathcal{B})$ with respect to the l th component of \mathcal{B} , i.e., $\mathcal{C}_{nl}^m = \partial \mathcal{C}(\mathcal{B})_n^m / \partial \mathcal{B}^l$. For simplicity, we are going to use simplified summation notation stating that any repeated upper and lower indexes mean a summation in that index, i.e., $A_j^i B_k^j = \sum_j A_j^i B_k^j$ or $A_j^i B^j = \sum_j A_j^i B^j$.

We express (A.6) as the following with error $O_p(n^{-2})$:

$$\begin{aligned}
 \widehat{\mathcal{B}}_{(i)}^m & = \mathcal{E}^m + (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}_{(i)}^n + (\mathcal{G}^{-1}\mathcal{K})_n^m \{\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}\}^n \\
 & + (\mathcal{G}^{-1})_n^m \Lambda_{(i)k}^n \widehat{\mathcal{B}}^k - (\mathcal{G}^{-1})_n^m \Lambda_{(i)k}^n \mathcal{P}^k.
 \end{aligned}$$

Define $I_n^m = 1$ if $n = m$ and $= 0$ otherwise, so that $\widehat{\mathcal{B}}_{(i)}^m = I_l^m \widehat{\mathcal{B}}_{(i)}^l$. Then we have that since

$$\begin{aligned}
 \Lambda_{(i)k}^n & = \sum_{j \neq i} \sum_l \frac{\partial C_j(\mathcal{B})_k^n}{\partial \mathcal{B}^l} \Big|_{\mathcal{B}=\widehat{\mathcal{B}}} (\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}})^l \\
 & = \sum_l \frac{\partial \mathcal{C}_{(i)k}(\mathcal{B})_k^n}{\partial \mathcal{B}^l} \Big|_{\mathcal{B}=\widehat{\mathcal{B}}} (\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}})^l = \mathcal{C}_{(i)kl}(\widehat{\mathcal{B}})_{kl}^n (\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}})^l,
 \end{aligned}$$

it follows that

$$\begin{aligned}
 \widehat{\mathcal{B}}_{(i)}^m & = \mathcal{E}^m + (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}_{(i)}^n + (\mathcal{G}^{-1}\mathcal{K})_n^m \{\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}\}^n \\
 & + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \{\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}\}^l \widehat{\mathcal{B}}^k \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \{\widehat{\mathcal{B}}_{(i)} - \widehat{\mathcal{B}}\}^l \mathcal{P}^k; \\
 \widehat{\mathcal{B}}_{(i)}^m & = (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}_{(i)}^n + (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}_{(i)}^n + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^k \widehat{\mathcal{B}}_{(i)}^l \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \mathcal{P}^k \widehat{\mathcal{B}}_{(i)}^l + \mathcal{E}^m - (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}^n \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \widehat{\mathcal{B}}^k + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \mathcal{P}^k; \\
 I_n^m \widehat{\mathcal{B}}_{(i)}^n & = (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}_{(i)}^n + (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}_{(i)}^n + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^k \widehat{\mathcal{B}}_{(i)}^l \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \mathcal{P}^k \widehat{\mathcal{B}}_{(i)}^l + \mathcal{E}^m - (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}^n \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \widehat{\mathcal{B}}^k + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \mathcal{P}^k, \\
 I_l^m \widehat{\mathcal{B}}_{(i)}^l & = (\mathcal{G}^{-1}\mathcal{K})_l^m \widehat{\mathcal{B}}_{(i)}^l + (\mathcal{G}^{-1}\mathcal{K})_l^m \widehat{\mathcal{B}}_{(i)}^l + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^k \widehat{\mathcal{B}}_{(i)}^l \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \mathcal{P}^k \widehat{\mathcal{B}}_{(i)}^l + \mathcal{E}^m - (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}^n \\
 & - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \widehat{\mathcal{B}}^k + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \mathcal{P}^k.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 & \{I_l^m - 2(\mathcal{G}^{-1}\mathcal{K})_l^m - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^k + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \mathcal{P}^k\} \widehat{\mathcal{B}}_{(i)}^l \\
 & = \mathcal{E}^m - (\mathcal{G}^{-1}\mathcal{K})_n^m \widehat{\mathcal{B}}^n - (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \widehat{\mathcal{B}}^k + (\mathcal{G}^{-1})_n^m \mathcal{C}_{(i)kl}^n \widehat{\mathcal{B}}^l \mathcal{P}^k
 \end{aligned}$$

In the other words,

$$(I - 2\mathcal{M} - \mathcal{G}^{-1}\mathcal{N} + \mathcal{G}^{-1}\mathcal{Q})\widehat{\mathcal{B}}_{(i)} = (\mathcal{E} - \mathcal{M}\widehat{\mathcal{B}} - \mathcal{G}^{-1}\mathcal{R}\widehat{\mathcal{B}} + \mathcal{G}^{-1}\mathcal{R}\mathcal{P}).$$

Hence, the solution $\widehat{\mathcal{B}}_{(i)}$ has the following form:

$$\begin{aligned}
 \widehat{\mathcal{B}}_{(i)} & = (I - 2\mathcal{M} - \mathcal{G}^{-1}\mathcal{N} + \mathcal{G}^{-1}\mathcal{Q})^{-1} \\
 & \times (\mathcal{E} - \mathcal{M}\widehat{\mathcal{B}} - \mathcal{G}^{-1}\mathcal{R}\widehat{\mathcal{B}} + \mathcal{G}^{-1}\mathcal{R}\mathcal{P}) + O_p(n^{-2}) \\
 & = \widehat{\mathcal{B}}_{(i)}^* + O_p(n^{-2}).
 \end{aligned}$$

This yields the desired result.

A.1.3 *Generalization of the approximate cross-validation to 1-m matched case-control studies.* The approximate cross-validation method can be generalized to 1-m matched case-control studies. For this generalization, we make the further definitions

$$\begin{aligned}
 \Omega(Z_{i1}, Z_{ij}, X_{i1}, X_{ij}) & = \left[1 + \sum_{j=2}^m \exp\left[-\{(Z_{i1} - Z_{ij})^T \beta_0\right. \right. \\
 & \left. \left. + \mathcal{L}_1(X_{i1}, X_{ij})^T \beta_1 + \mathcal{L}_2(X_{i1}, X_{ij})^T \beta_2\} \right] \right]^{-1} \\
 & = \left[1 + \sum_{j=2}^m \exp\left[-\{\mathcal{L}_j^T(Z_{i1}, Z_{ij}, X_{i1}, X_{ij})\mathcal{B}\} \right] \right]^{-1}; \\
 \Omega_{\mathcal{B}}(\bullet) & = \frac{\partial}{\partial \mathcal{B}} \Omega(\bullet).
 \end{aligned}$$

In what follows, we will write \mathcal{L}_j for $\mathcal{L}_j(\bullet)$.

$$F_i(\mathcal{B}) = \frac{\Omega_{\mathcal{B}}(\bullet)}{\Omega(\bullet)} = \frac{\sum_{j=2}^m \mathcal{L}_j \exp\{-\mathcal{L}_j^T \mathcal{B}\}}{1 + \sum_{j=2}^m \exp\{-\mathcal{L}_j^T \mathcal{B}\}}; \tag{A.7}$$

$$\begin{aligned}
 C_i(\mathcal{B}) & = \frac{\sum_{j=2}^m \mathcal{L}_j \mathcal{L}_j^T \exp(-\mathcal{L}_j^T \mathcal{B}) + \sum_{j=2}^m \mathcal{L}_j \mathcal{L}_j^T \exp(-\mathcal{L}_j^T \mathcal{B}) \sum_{j=2}^m \exp\{-\mathcal{L}_j^T \mathcal{B}\}}{\left[1 + \sum_{j=2}^m \exp\{-\mathcal{L}_j^T \mathcal{B}\} \right]^2} \\
 & \times \frac{\left\{ \sum_{j=2}^m \mathcal{L}_j \exp(\mathcal{L}_j^T \mathcal{B}) \right\} \left\{ \sum_{j=2}^m \mathcal{L}_j \exp(\mathcal{L}_j^T \mathcal{B}) \right\}^T}{\left[1 + \sum_{j=2}^m \exp\{-\mathcal{L}_j^T \mathcal{B}\} \right]^2}. \tag{A.8}
 \end{aligned}$$

By arguments similar to those of Sections A.1.1–A.1.2, our approximate leave-one-out estimator is (A.2) with $F_i(\mathcal{B})$ in (A.8) and $C_i(\mathcal{B})$ in (A.8).

A.2 Log-Concavity of the Posterior Distribution $[\beta_2 \mid \text{rest}]$

If we take the logarithm of the function

$$\prod_{i=1}^n H \left\{ (Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} + L_1^T(X_{i1}, X_{i2}) \beta_{1,\text{old}} + L_2^T(X_{i1}, X_{i2}) \beta_2 \right\} f(\beta_2 \mid \sigma^2),$$

we obtain

$$\begin{aligned} h(\beta_2) = & -\beta_2^T \beta_2 / (2\sigma^2) + \sum_i \left[(Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} \right. \\ & + L_1^T(X_{i1}, X_{i2}) \beta_{1,\text{old}} + L_2^T(X_{i1}, X_{i2}) \beta_2 \\ & - \log \left\{ 1 + \exp \left((Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} \right. \right. \\ & \left. \left. + L_1^T(X_{i1}, X_{i2}) \beta_{1,\text{old}} + L_2^T(X_{i1}, X_{i2}) \beta_2 \right) \right\} \left. \right]. \end{aligned}$$

Since

$$\begin{aligned} \frac{\partial h(\beta_2)}{\partial \beta_2^T} = & \\ \times \sum_i & \left[L_2^T(\bullet) - \frac{L_2^T(\bullet) \exp \left\{ (Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} + L_1^T(\bullet) \beta_{1,\text{old}} + L_2^T(\bullet) \beta_2 \right\}}{1 + \exp \left\{ (Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} + L_1^T(\bullet) \beta_{1,\text{old}} + L_2^T(\bullet) \beta_2 \right\}} \right] - \frac{\beta_2^T}{\sigma^2}; \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 h(\beta_2)}{\partial \beta_2 \partial \beta_2^T} = & \\ - \sum_i & \left[\frac{L_2(\bullet) \exp \left\{ (Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} + L_1^T(\bullet) \beta_{1,\text{old}} + L_2^T(\bullet) \beta_2 \right\} L_2^T(\bullet)}{\left[1 + \exp \left\{ (Z_{i1} - Z_{i2})^T \beta_{0,\text{old}} + L_1^T(\bullet) \beta_{1,\text{old}} + L_2^T(\bullet) \beta_2 \right\} \right]^2} \right] - \frac{1}{\sigma^2} I_k, \end{aligned}$$

$h(\beta_2)$ is log-concave function.

A.3 Log-Concavity of the Posterior Distribution $[\mathcal{B} \mid \text{rest}]$

If we take the logarithm of the function

$$\prod_{i=1}^n H \left\{ \mathcal{L}^T(Z_{i1}, Z_{i2}, X_{i1}, X_{i2}) \mathcal{B} \right\} f(\mathcal{B} \mid \Sigma), \tag{A.9}$$

we obtain

$$\begin{aligned} h(\mathcal{B}) = & -\mathcal{B}^T \mathcal{B} / (2\sigma^2) + \sum_i \left[\mathcal{L}^T(Z_{i1}, Z_{i2}, X_{i1}, X_{i2}) \mathcal{B} \right. \\ & \left. - \log \left\{ 1 + \exp \left[\mathcal{L}^T(Z_{i1}, Z_{i2}, X_{i1}, X_{i2}) \mathcal{B} \right] \right\} \right]. \end{aligned}$$

We find that

$$\begin{aligned} \frac{\partial h(\mathcal{B})}{\partial \mathcal{B}^T} = & \sum_i \left[\mathcal{L}^T(\bullet) - \frac{\mathcal{L}^T(\bullet) \exp \{ \mathcal{L}^T(\bullet) \mathcal{B} \}}{1 + \exp \{ \mathcal{L}^T(\bullet) \mathcal{B} \}} \right] - \mathcal{B}^T \Sigma^{-1}; \\ \frac{\partial^2 h(\mathcal{B})}{\partial \mathcal{B} \partial \mathcal{B}^T} = & - \sum_i \left[\frac{\mathcal{L}(\bullet) \exp \{ \mathcal{L}^T(\bullet) \mathcal{B} \} \mathcal{L}^T(\bullet)}{\left[1 + \exp \{ \mathcal{L}^T(\bullet) \mathcal{B} \} \right]^2} \right] - \Sigma^{-1}. \end{aligned}$$

Hence, $h(\mathcal{B})$ is a log-concave function.