Mixed Model Prediction: Part IV

Jiming Jiang and Thuan Nguyen

University of California, Davis, USA

and Oregon Health & Science University, USA

Classified Mixed Model Prediction

- Nowadays, new and challenging problems have emerged from fields as business and health sciences, in addition to the traditional fields.

  Some of these problems occur when interest is at subject level (e.g., individual customer), or (small) sub-population level (e.g., small community), rather than at large population level.

  Examples: online shopping, personalized medicine.

  In such cases, it is possible to make substantial gains in prediction accuracy by identifying a class that a new subject belongs to. Once the subject class is identified, method of MMP can be used to obtain optimal prediction about the subject.

- *Prediction of Mixed Effects*

Suppose that we have a set of taining data, $y_{ij}, i = 1, \dots, m, j = 1, \dots, n_i$ in the sense that their classifications are known, that is, one knows which group, i, that y_{ij} belongs to.
The assumed model for the training data is

$$\begin{aligned} y_i &= \mathrm{E}(y_i|\alpha) + \epsilon_i \\ &= X_i\beta + Z_i\alpha_i + \epsilon_i, \end{aligned} \tag{1}$$

where $y_i = (y_{ij})_{1\le j\le n_i}$, $X_i = (x'_{ij})_{1\le j\le n_i}$ is a matrix of known covariates, β is a vector of unknown regression coefficients (the fixed effects), Z_i is a known $n_i \times q$ matrix, α_i is a $q \times 1$ vector of group-specific random effects, ϵ_i is an $n_i \times 1$ vector of errors, and $\alpha = (\alpha_i)_{1\le i\le m}$.

- It is assumed that the α_i's and ϵ_i's are independent, with $\alpha_i \sim N(0, G)$ and $\epsilon_i \sim N(0, R_i)$, where the covariance matrices G and R_i depend on a vector ψ of dispersion parameters, or variance components.

  The first line of (1) is a true model, where E is the true conditonal expectation. The second line of (1), that is, the LMM, which is potentially misspecified.

  Our goal is to make a classified prediction for a mixed effect, that is, a linear combination of fixed and random effects, associated with a new observation, $y_{\rm n}$ (the subscript n referrs to "new").

- Suppose that

$$
\begin{aligned}
y_{\rm n} &= {\rm E}(y_{\rm n}|\alpha) + \epsilon_{\rm n} \\
&= x_{\rm n}'\beta + z_{\rm n}'\alpha_{\rm I} + \epsilon_{\rm n}, \qquad (2)
\end{aligned}
$$

where $x_{\rm n}, z_{\rm n}$ are known vectors, $I \in \{1, \dots, m\}$ but one does not know which element i, $1 \leq i \leq m$, is equal to I.

Furthermore, $\epsilon_{\rm n}$ is a new error that is independent of $y_i, 1 \leq i \leq m$, and has mean zero. Note that, like (1), the second line of (2) may be misspecified.

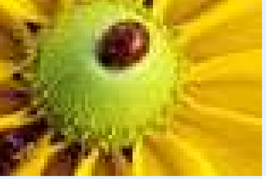
Nevertheless, the true mixed effect that we would like to predict can be expressed as $\theta = {\rm E}(y_{\rm n}|\alpha) = y_{\rm n} - \epsilon_{\rm n}$.

- Suppose that I is known, say, $I = i$. Then, the vectors $y_1, \dots, y_{i-1}, (y_i', \theta)', y_{i+1}, \dots, y_m$ are independent.

  Thus, we have $\mathrm{E}_M(\theta | y_1, \dots, y_m) = \mathrm{E}_M(\theta | y_i)$.

  Here, E_M denotes the conditional expectation under the assumed LMM.

  The consideration here takes into account the potential misspecification of the assumed model, so E and E_M are different.



- Furthermore, by the normal theory, we have

$$
\begin{aligned}
\mathrm{E}_M(\theta|y_i) &= x_{\mathrm{n}}'\beta \\
&\quad + z_{\mathrm{n}}'GZ_i'(R_i + Z_iGZ_i')^{-1}(y_i - X_i\beta). \qquad (3)
\end{aligned}
$$

The right side of (3) is the best predictor (BP) under the assumed LMM, if the true parameters, β and ψ (again, under the assumed model) are known.

Because the latter are unknown, we replace them by $\hat{\beta}$ and $\hat{\psi}$, respectively. The result is what we call empirical best predictor (EBP), denoted by $\tilde{\theta}_{(i)}$.

- By now, we know how to do MMP when the class, I, is known. In practice, however, I is unknown, and thus treated as a parameter.

  In order to identify, or estimate, I, we consider the mean MSPE of θ by the BP when I is classified as i, that is

$$\begin{aligned} \mathrm{MSPE}_i &= \mathrm{E}\{\tilde{\theta}_{(i)} - \theta\}^2 \\ &= \mathrm{E}\{\tilde{\theta}_{(i)}^2\} - 2\mathrm{E}\{\tilde{\theta}_{(i)}\theta\} + \mathrm{E}(\theta^2). \end{aligned}$$

- Using the expression $\theta = y_{\mathrm{n}} - \epsilon_{\mathrm{n}}$, we have $\mathrm{E}\{\tilde{\theta}_{(i)}\theta\} = \mathrm{E}\{\tilde{\theta}_{(i)}y_{\mathrm{n}}\} - \mathrm{E}\{\tilde{\theta}_{(i)}\epsilon_{\mathrm{n}}\} = \mathrm{E}\{\tilde{\theta}_{(i)}y_{\mathrm{n}}\}$.

  Thus, we have the expression:

$$\mathrm{MSPE}_i = \mathrm{E}\{\tilde{\theta}_{(i)}^2 - 2\tilde{\theta}_{(i)}y_{\mathrm{n}} + \theta^2\}. \tag{4}$$

  Note that the E on the right side of (4) is the true expectation. It follows that the observed MSPE corresponding to (4) is the expression inside the expectation.

- Therefore, a natural idea is to identify I as the index i that minimizes the observed MSPE.

Because θ^2 does not depend on i, this is equivalent to

(5) $$I \quad = \quad \mathrm{argmin}_i \left\{ \tilde{\theta}^2_{(i)} - 2\tilde{\theta}_{(i)} y_{\mathrm{n}} \right\}.$$

Denote the I identified by (5) by $\hat{I}$. Then, the classified predictor of θ is given by $\hat{\theta} = \tilde{\theta}_{(\hat{I})}$, in which whatever unknown parameters are estimated by, say, REML estimators based on the training data.

- Empirical Results: Simulation studies were carried out to investigate the finite-sample performance of the proposed CMMP.

  The results are compared with the standard regression prediction (RP). Partial results are presented here.

  1. The matched case

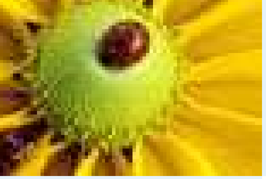
  We consider an NER model that was introduced earlier.

- The value of R is fixed at 1.

  We consider $m = 50$ and $n_i = 5, 1 \leq i \leq m$.

  First consider a mixed effect, $\theta_{\rm n} = x_{\rm n}'\beta + \alpha_I$, corresponding to a new observation, $y_{\rm n}$, where $I \in \{1, \ldots, m\}$.

  The reported results are based on 100 simulation runs.



- Prediction of Mixed Effect: MSPE (SE) and % increase over CMMP

G	0.25	1.00	4.00	9.00
CMMP	.418	.799	.953	.977
	(.006)	(.011)	(.014)	(.015)
RP	.247	1.002	3.909	8.559
	(.003)	(.014)	(.056)	(.121)
% Increase	-41.0%	25.5%	310.3%	775.6

- *2. Comparing matched and unmatched cases*

This simulation study was carried out under the following model:

$$y_{ij} \quad = \quad 3 + 2x_{1,ij} + x_{2,ij} + \alpha_i + \epsilon_{ij},$$

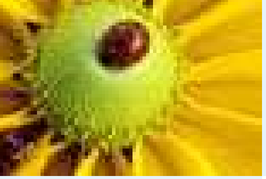
$i = 1, \dots, m$, $j = 1, \dots, n$, with $n = 5$, $\alpha_i \sim N(0, \sigma_\alpha^2$, $\epsilon_{ij} \sim N(0, 1)$, and α_i's, ϵ_{ij}'s are indep.

There are $K = 10$ new observations, generated under two scenarios.

- Scenario I: The new observations have the same $\alpha_i, i = 1, \ldots, K$, but independent ϵ's; that is, they have "matches".

  Scenario II: The new observations have independent α's and ϵ's; that is, they are "unmatched".

  All simulation results were obtained based on $T = 1000$ simulation runs.



- $m = 10$. %MATCH $=$ percentage of times that the new observations were matched, by CMMP, to some of the training-data random effects.

Scenario	σ_α^2	0.1	1	2	3
I	RP	0.16	0.96	1.86	2.79
I	CMMP	0.20	0.63	0.76	0.84
I	%MATCH	98.5	95.0	93.4	92.8
II	RP	0.18	1.17	2.23	3.32
II	CMMP	0.22	0.77	0.96	1.08
II	%MATCH	98.5	94.6	93.3	93.0

- Questions?

  It appears that, regardless of whether the new observations actually have matches or not, the CMMP match them anyway.

  And, more importantly, the results show that even a "fake" match still helps.

  Make sense?

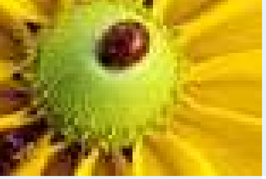
  Think about a business situation. Even if I cannot find a perfect match for a customer, but if I can find a group that is kind of similar, I can still gain in terms of prediction accuracy. This is in fact how business decisions are often made.

- In the simulation study, even if there is no match in terms of the individual random effects, there is at least a “match” in terms of the random effects distribution (i.e., the true random effect is from the same distribution as that for the training data random effects). So, there is still strength that one can borrow.

  Comparing the RP with CMMP, PR says that the mixed effect is $x_i'\beta$ with nothing extra.

  On the other hand, CMMP says that the mixed effect is $x_i'\beta$ plus something extra.

  For the new observation, there is, for sure, something extra (that is, the extra is non-zero).



- So, CMMP is right, at least in that the extra is non-zero;

  it then selects the best extra from a number of choices, some of which are better than the zero extra that PR is using.

  Therefore, it is not surprising that CMMP is doing better, regardless of the actual match (which may or may not exist).

  This interesting feature makes the CMMP method more useful, because in practice an actual match may not happen.

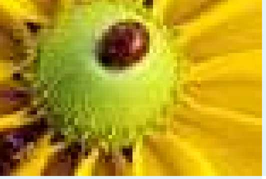
  The empirical results observed above are supported by theory (Jiang *et al.* 2018).

Example

- We use a data set from the Television School and Family Smoking Prevention and Cessation Project (TVSFP; Hedeker *et al.* 1994) to illustrate the OBP method for finite-population sampling.

  The original study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation.

  The subjects were seventh-grade students from Los Angeles (LA) and San Diego in the State of California in the United States.



- The students were pretested in January 1986 in an initial study. The same students completed an immediate postintervention questionnaire in April 1986, a one-year follow-up questionnaire (in April 1987), and a two-year follow-up (in April 1988).

  In this analysis, we consider a subset of the TVSFP data involving students from 28 LA schools, where the schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum (CC); (b) a media (television) intervention (TV); (c) a combination of CC and TV conditions; and (d) a no-treatment control.

- A tobacco and health knowledge scale (THKS) score was one of the primary study outcome variables, and the one used for this analysis.

  The THKS consisted of seven questionnaire items used to assess student tobacco and health knowledge.

  A student's THKS score was defined as the sum of the items that the student answered correctly.

- Only data from the pretest and postintervention are available for the current analysis.

  More specifically, the data only involved subjects who had completed the THKS at both of these time points.

  In all, there were 1,600 students from the 28 schools, with the number of students from each school ranging from 18 to 137.

  Due to the potential model misspecification, the OBP method (Jiang *et al.* 2011) is used instead of the EBLUP. See Part VI of the lecture series.

- We use the data to run a real-data based simulation in which we know the truth, hence can evaluate performance of the CMMP and its comparison with RP.

  We don’t know the population, so we assume that the population is duplications of the training data for the school by 10 times; thus, in particular, the population mean is the same as the sample mean (but we pretend that this is unknown).

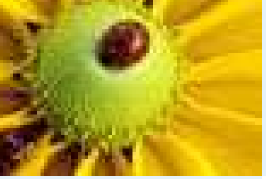
  We also don’t have new observations in this case, so we assume that the new observations are one of the duplicated schools.

- We then use the CMMP to predict (estimate) the population mean for each school and compare it with the regression prediction (RP).

  We did this for each of the 28 LA schools.

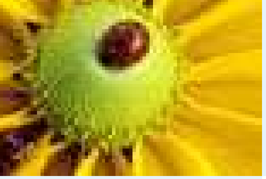
  CMMP has smaller prediction error than RP for each of the 28 schools.

  Partial results are presented in the table below.



- Partial results for the TVSFP data: Presented are absolute values of prediction errors.

School #	CMMP	RP
1	0.680	0.787
2	0.518	0.624
3	0.059	0.143
4	0.003	0.056
5	0.651	0.758
6	0.008	0.092
7	0.562	0.668
8	0.148	0.262

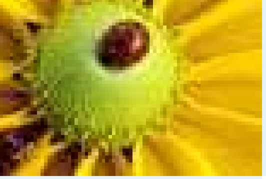


Classified Mixed Logistic Model Prediction

- Many problems in health science are related to prediction, where the main interest is at subject (e.g., precision medicine) or sub-population (e.g., precision public health) level.

  In such cases, it is possible to make substantial gains in prediction accuracy by identifying a class that a new subject belongs to.

  Once such a class is identified for the new subject, existing data with known classification can be used to gain more accuracy in prediction about the new subject.



- This was recently demonstrated by Jiang, Nguyen & Rao (2017), who proposed a method call classified mixed model prediction (CMMP).

  The idea is to create a “match” between the classes, or clusters, of the training data and the potential class of the new data.

  Once such a class is identified, mixed model prediction (MMP) technique can be utilized to improve prediction accuracy.

  However, the CMMP method only applies to linear models for continuous responses.

- Clustered binary data frequently occur in health & medical studies.

  For example, Thromboembolic or hemorrhagic complications (e.g., Glass *et al.* 1997) occur in as many as 60% of patients who underwent extracorporeal membrane oxygenation (ECMO), an invasive technology used to support children during periods of reversible heart or lung failure (e.g., Muntean 2002).

  Over half of pediatric patients on ECMO are currently receiving antithrombin (AT) to maximize heparin sensitivity.

- In a retrospective, multi-center, cohort study of children ($\leq$ 18 years of age) who underwent ECMO between 2003 and 2012, 8,601 subjects participated in 43 free-standing children's hospitals across 27 U.S. states and the District of Columbia known as Pediatric Health Information System (PHIS).

  Many of the outcome variables were binary, such as the bleed_binary variable, which is a main outcome variable indicating hemorrhage complication of the treatment; and the DischargeMortalit1Flag variable, which is associated with mortality. Here the treatment refers to AT.

- The data are also potentially clustered, with the clusters corresponding to the children's hospitals.

  In addition to the treatment indicator, there were 20 other covariate variables, for which information were available. More detail about the data will be provided later.

  Prediction associated with binary outcomes is very important in practice.

- For example, recent studies by researchers from Johns Hopkins University University (Tomasetti & Vogelstein 2015, Tomasetti, Li, & Vogelstein 2017) found that, in a way, cancer is at least partially caused by “bad luck”, and early detection is a cure for cancer.

  This raised the importance of (early) prediction regarding probability of such a bad luck.

  Under a mixed effects model, which are extensively used in medical studies, such probabilities are (nonlinear) mixed effects associated with subject-specific random effects.

Mixed Logistic Model

- In the context of SAE with binary data, Jiang & Lahiri (2001) considered the following mixed logistic model (this is a review of of part of Part II of the lecture series):

  Given the area-specific random effects, $\alpha_1, \dots, \alpha_m$, binary responses $y_{ij}, i = 1, \dots, m, j = 1, \dots, n_i$ are conditionally independent with the conditional probability satisfying

  $$\mathrm{P}(y_{ij} = 1 | \alpha) = p_{ij} \;\text{ with }\; \mathrm{logit}(p_{ij}) = x_{ij}'\beta + \alpha_i,$$

  where $\mathrm{logit}(p) = \log\{p/(1-p)\}$.

  Furthermore, $\alpha_i, 1 \le i \le m$ are independent and distributed as $N(0, \sigma^2)$, where σ^2 is an unknown variance.

Empirical Best Predictor (EBP)

- Here is a review of part of Part III of the lecture series. If the mixed effect of interest is a conditional probability, $\theta = g(x'\beta + \alpha_i)$ for some known function $g(\cdot)$, the best predictor (BP) of θ, in terms of minimum mean squared prediction error (MSPE), is $\tilde{\theta} = \mathrm{E}(\theta|y) =$

  $$\frac{\mathrm{E}[g(x'\beta + \sigma\xi)\exp\{y_{i\cdot}\sigma\xi - \sum_{j=1}^{n_i}\log(1 + e^{x'_{ij}\beta + \sigma\xi})\}]}{\mathrm{E}[\exp\{y_{i\cdot}\sigma\xi - \sum_{j=1}^{n_i}\log(1 + e^{x'_{ij}\beta + \sigma\xi})\}]},$$

  where $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$, and the expectations are taken with respect to $\xi \sim N(0,1)$.

  The empirical best predictor (EBP) is the BP with β, σ replaced by their consistent estimators.

CMLMP

- Our main interest is to predict a mixed effect that is associated with a set of new observations.

  Let the new, binary observations be $y_{\mathrm{n},k}, k = 1, \dots, n_{\mathrm{new}}$, and the corresponding covariates be $x_{\mathrm{n},k}, k = 1, \dots, n_{\mathrm{new}}$.

  Assume that, conditional on α_I that has the same $N(0, \sigma^2)$ distribution, $y_{\mathrm{n},k}, 1 \le k \le n_{\mathrm{new}}$ are independent with

$$\mathrm{P}(y_{\mathrm{n},k} = 1 | \alpha_I) = p_{\mathrm{n},k} \;\text{ and }\; \mathrm{logit}(p_{\mathrm{n},k}) = x'_{\mathrm{n},k}\beta + \alpha_I.$$

- For simplicity of illustration, assume that the covariates are at cluster level.

  The methods can be extended to both cluster-level and unit-level covariates.

  The mixed effect of interest is
  $p_{\mathrm{n}} = \mathrm{P}(y_{\mathrm{n},k} = 1|\alpha_I) = \mathrm{logit}^{-1}(x'_{\mathrm{n}}\beta + \alpha_I) = \mathrm{logit}^{-1}(x'_{\mathrm{n}}\beta + \alpha_i)$.

  A key step is to find a match, $\hat{I}$, among $1 \leq i \leq m$ for the index I corresponding to the random effect associated with the new observations.

1. A CMMP approach

- Let

$$\hat{I} = \text{argmin}_{1 \le i \le m} \{\hat{p}_{\text{n},(i)} - \bar{y}_{\text{n}}\}^2,$$

where $\hat{p}_{\text{n},(i)}$ is the EBP of p_{n} assuming $I = i$, and $\bar{y}_{\text{n}}$ is the sample mean (or proportion) of the new (binary) observations.

This approach is the same as that of Jiang *et al.* (2017).

Once $\hat{I}$ is identified, the CMLMP of p_{n} is $\hat{p}_{\text{n}} = \hat{p}_{\text{n},(\hat{I})}$.

Empirical demonstration

- Training data generated under a mixed logistic model with

$$\text{logit}(p_i) = \text{logit}\{\text{P}(y_{ij} = 1|\alpha_i)\} = 1 + 2x_i + \alpha_i,$$

  $i = 1, \ldots, m, j = 1, \ldots, n_i$, where $n_i = 5$ for all i.
  The covariates, x_i, are generated from the $N(0, 1)$ distribution.
  The random effects, α_i, are then generated independently from the $N(0, \sigma^2)$ distribution.
  We consider two scenarios: A matched case and an unmatched case.
  Matched case (I): $I = 1$. Unmatched case (II): α_I is generated independently with $1 \leq i \leq m$; therefore, there is no match for I.

- Regardless, we carry out CMLMP anyway, assuming that there is a match.

  Note: It can be shown that CMLMP is consistent in predicting the mixed effect of interest regardless of whether there is an actual match or not. The detail is omitted.

  We compare performance of CMLMP with the standard logistic regression prediction (SLRP).

  The results, based on 500 simulation runs, are presented in the following tables. Reported are simulated MSPEs from the simulations.

Table 1

- $\sigma = 1, n_{\text{new}} = 5$, and changing m. I–Matched case; II–Unmatched case:

	m	10	50	100	500	1000
I	CMLMP	.0231	.0167	.0158	.0187	.0197
I	SLRP	.0318	.0247	.0245	.0233	.0211
II	CMLMP	.0255	.0181	.0161	.0177	.0190
II	SLRP	.0332	.0216	.0260	.0249	.0207

Table 2

- $\sigma = 1, m = 50$, and changing n_{ew}. I–Matched case; II–Unmatched case:

	n_{new}	1	5	10	100	1000
I	CMLMP	.0330	.0178	.0109	.0062	.0050
I	SLRP	.0249	.0256	.0232	.0266	.0218
II	CMLMP	.0375	.0182	.0104	.0055	.0047
II	SLRP	.0244	.0291	.0237	.0255	.0219

Table 3

- $m = 50, n_{\text{new}} = 5$, and changing σ. I–Matched case; II–Unmatched case:

	σ	0.25	0.5	1	2	3
I	CMLMP	.0038	.0085	.0168	.0176	.0216
I	SLRP	.0027	.0079	.0238	.0673	.1037
II	CMLMP	.0043	.0101	.0176	.0209	.0193
II	SLRP	.0028	.0091	.0257	.0658	.1119

2. Incorporating cluster-level covariates

- The matching strategy, so far, is similar to Jiang *et al.* (2017) in that no information from the covariates of the new observations is used in identifying the class.

  In practice, covariate information can often help in identifying the class. This is particularly the case when there are covariates at the cluster level.

  For example, much effort has been made in trying to model the functional relationship between the mean response and the covariates, e.g., linear regression, polynomial regression, splines, nonparametric regression.

- Alternatively, random effects are often introduced to “capture the uncaptured”, that is, variation that cannot be explained by the assumed functional relationship with the covariates.

  In such a case, it is reasonable to assume that there is some kind of correspondence between the cluster-level covariates and the cluster-specific random effects, $\alpha_i, 1 \leq i \leq m$.

  Let w_i denote a vector of cluster-level covariates. Our idea is to consider the difference, $w_i - w_{\rm n}$, where $w_{\rm n}$ is the corresponding vector of covariates associated with the new observations.

- Specifically, consider a cluster-specific linear predictor defined as $l_i = w_i' b + \alpha_i$, where b is the vector of fixed effects corresponding to w_i.

  Similarly, we have $l_{\rm n} = w_{\rm n}' b + \alpha_I$ for the new observations.

  Thus, we have

$$\begin{aligned} {\rm E}(l_i - l_{\rm n})^2 &= \{(w_i - w_{\rm n})' b\}^2 + {\rm E}(\alpha_i - \alpha_I)^2 \\ &= {\rm E}\left[\{(w_i - w_{\rm n})' b\}^2 + (\alpha_i - \alpha_I)^2\right]. \end{aligned}$$

- Thus, using the CMMP idea (Jiang *et al.* 2017), to minimize $\mathrm{E}(l_i - l_\mathrm{n})^2$ we need to minimize

$$\{(w_i - w_\mathrm{n})'b\}^2 + (\alpha_i - \alpha_I)^2.$$

This leads to the new class identifier:

$$\hat{I} = \mathrm{argmin}_{1 \le i \le m} \left[\{(w_i - w_\mathrm{n})'\hat{b}\}^2 + (\hat{\alpha}_i - \hat{\alpha}_\mathrm{n})^2 \right].$$

where $\hat{b}$ is the consistent estimator of b, $\hat{\alpha}_i$ is the EBP of α_i, and $\hat{\alpha}_\mathrm{n}$ is obtained similarly by replacing β, σ in the following expression by $\hat{\beta}, \hat{\sigma}$, respectively, the consistent estimators based on the training data:

$$\sigma \frac{\mathrm{E}[\xi \exp\{y_{\mathrm{n},\cdot}\sigma\xi - \sum_{j=1}^{n_\mathrm{new}} \log(1 + e^{x'_{\mathrm{n},j}\beta + \sigma\xi})\}]}{\mathrm{E}[\exp\{y_{\mathrm{n},\cdot}\sigma\xi - \sum_{j=1}^{n_\mathrm{new}} \log(1 + e^{x'_{\mathrm{n},j}\beta + \sigma\xi})\}]}.$$

More simulation study

- We carry out a simulation study under the same model as before, with $w_i = x_i$ as the cluster-level covariate.

  A difference is that now the random effect, α_i, is introduced to capture the uncaptured, as noted above.

  Specifically, $\alpha_i = g(w_i) + v_i$, where $g(w_i)$ corresponds to the uncaptured, an unknown function of the covariate, and v_i is a small noise, generated independently from the $N(0, D)$ distribution.

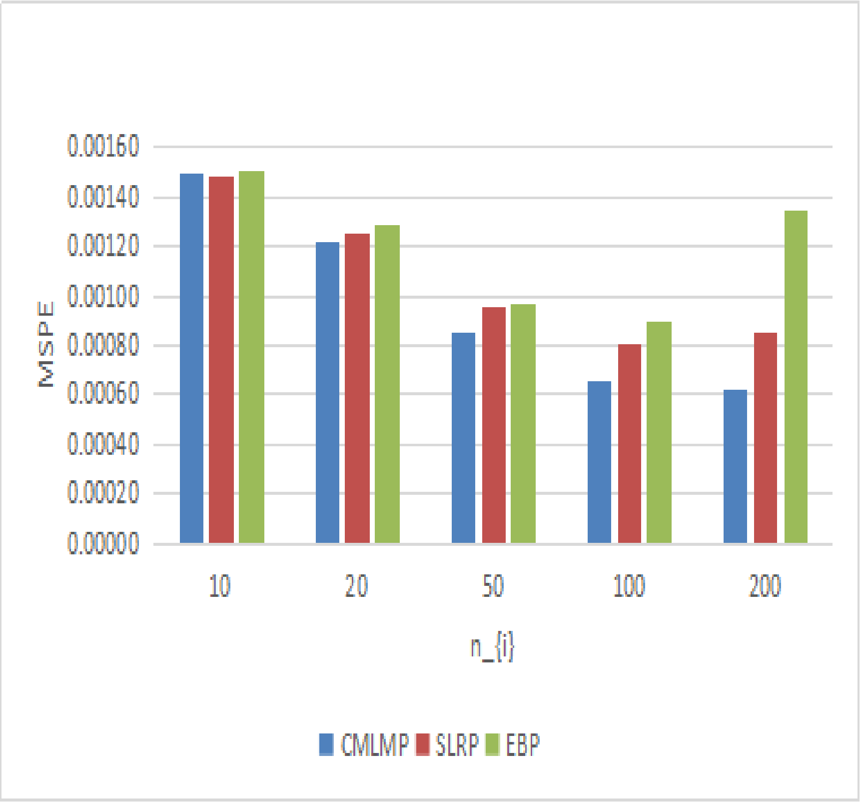
- Two difference functions are considered: $g(w_i) = w_i^3$; and $g(w_i) = w_i^2 - 1$.

  We consider a matched case with $I = 1$.

  We illustrate the results, in terms of the simulated MSPEs (based on 500 simulation runs), using figures.

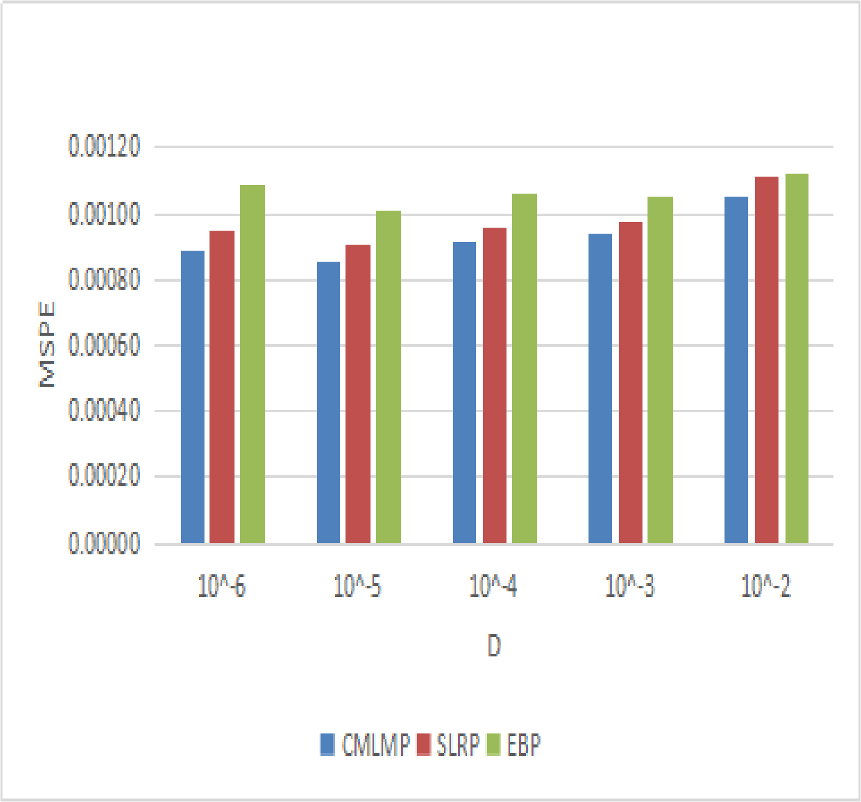
  Figure 1: $g(w_i) = w_i^3$; $m = 50, n_{\rm new} = 10, D = 10^{-4}$ and changing n_i.

  Figure 2: $g(w_i) = w_i^2 - 1$; $m = 50, n_{\rm new} = 10, n_i = 50$ and changing D.



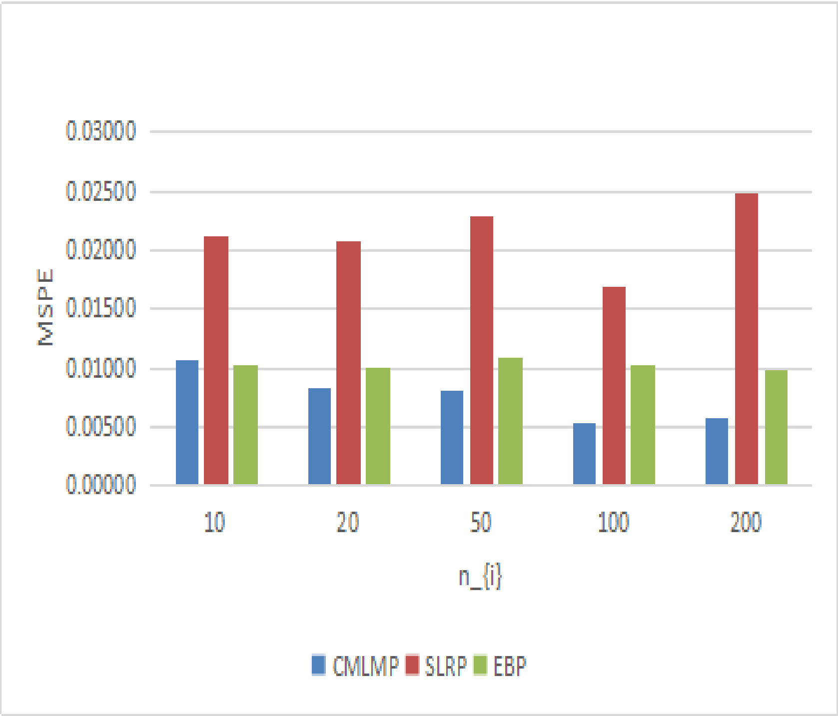
0.00160
0.00140
0.00120
0.00100
0.00080
0.00060
0.00040
0.00020
0.00000
MSPE
10
20
50
100
200
n_{i}
CMLMP SLRP EBP

(a)

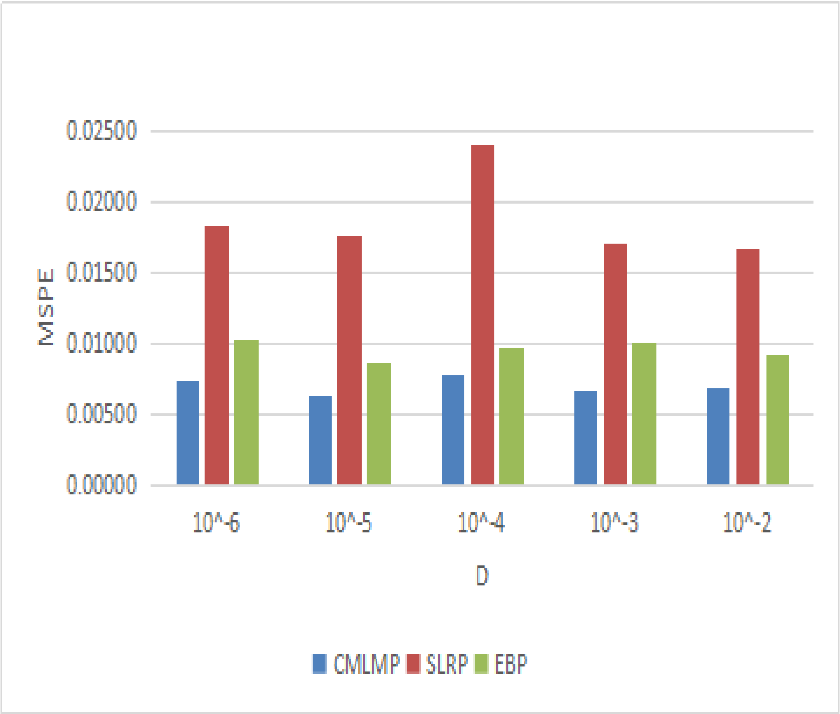


0.00120
0.00100
0.00080
0.00060
0.00040
0.00020
0.00000
MSPE
10^-6
10^-5
10^-4
10^-3
10^-2
D
CMLMP SLRP EBP

(b)


0.03000
0.02500
0.02000
0.01500
0.01000
0.00500
0.00000
MSPE
10
20
50
100
200
n_{i}
CMLMP SLRP EBP

(a)


0.02500
0.02000
0.01500
0.01000
0.00500
0.00000
MSPE
10^-6
10^-5
10^-4
10^-3
10^-2
D
CMLMP SLRP EBP

(b)

Application: ECMO data revisited

- We focus on the two outcomes of interest, bleed_binary variable and DischargeMortalit1Flag variable, that were mentioned.

  The data includes 8601 patients data from 42 hospitals. The numbers of patients in different hospitals range from 3 to 487.

  We first use a forward-backward (F-B) BIC procedure (e.g., Broman & Speed 2002) to build a mixed logistic model.

  The F-B BIC procedure leads to a subset of 12 patient-level covariates (two are continuous; the rest binary), out of a total of more than 20 covariates.

- In addition, there are 2 hospital-level covariates, yesat and total.

  The former refers to the total number of patients, in each hospital, during the 10 year study who did get ATC.

  The latter is the total number of patients in each hospital that were included in the 10-year study.

  Both hospital-level covariates are continuous.

- The mixed effects of interest are probabilities of hemorrhage complication corresponding to bleed_binary, and mortality probabilities associated with DischargeMortalit1Flag, for new observations.

  In order to test the CMLMP method, we randomly select 5 patients from a given hospital and treat these as the new observations.

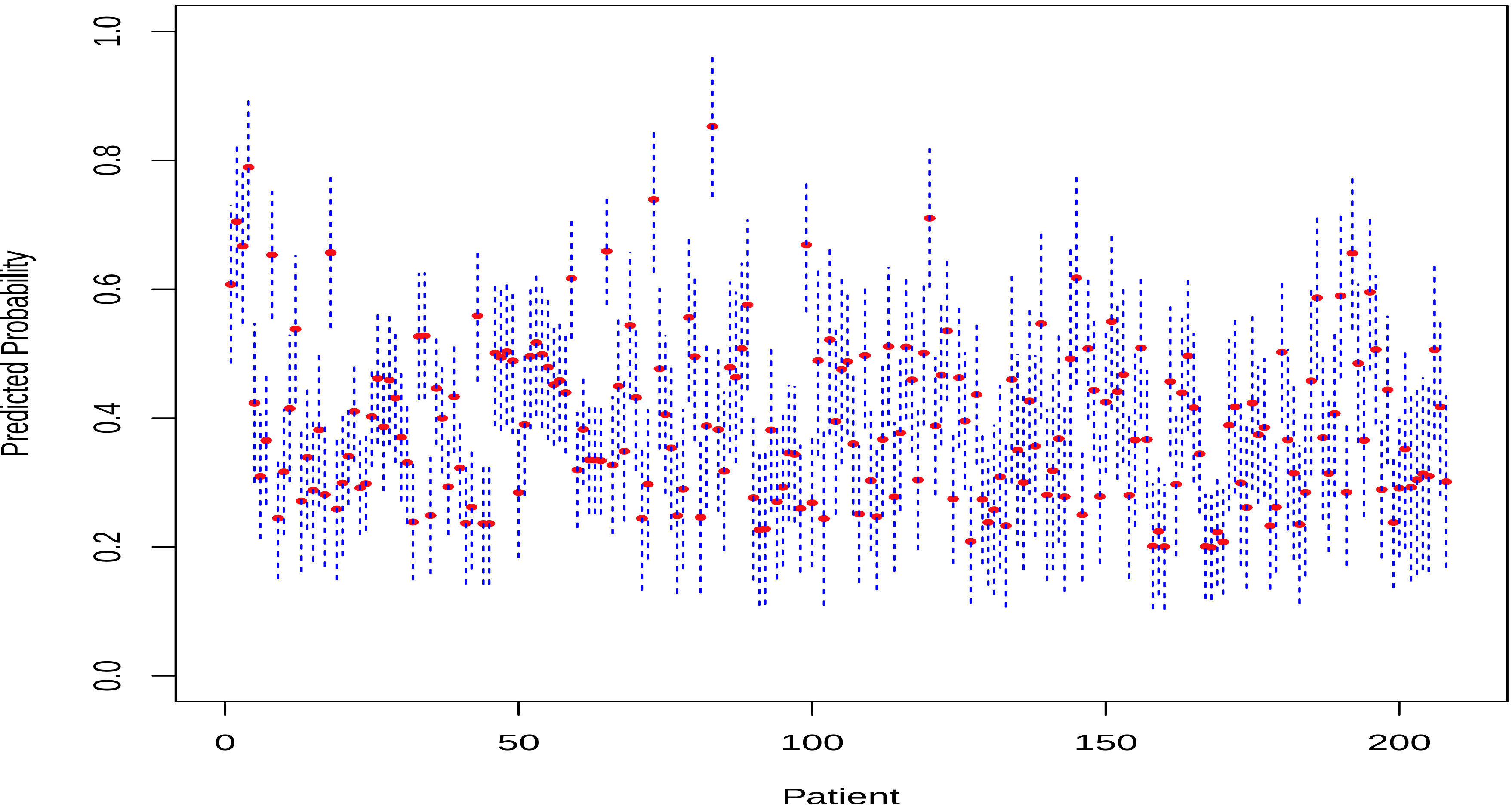
  The rest of the hospitals, and rest of the patients from the same hospital (if any), correspond to the training data.

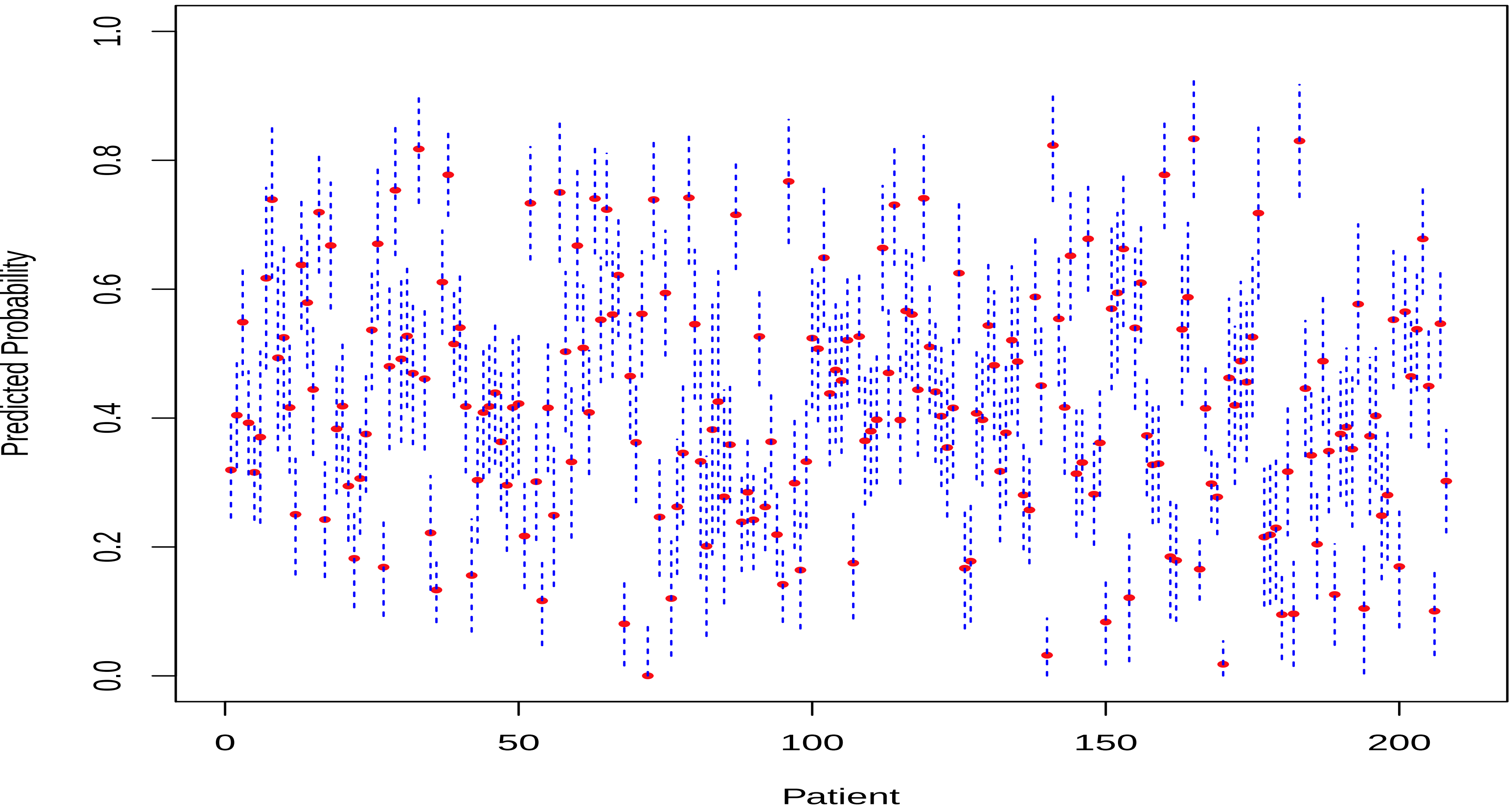
  One hospital (#2033) has only three patients available. This, a total of 208 patients were selected for CMLMP prediction.

- The CMLMP method that incorporates cluster-level covariates is applied to each group of 5 selected patients (and the group of 3 for #2033).

  The results are presented in Figure 3 and Figure 4.

  Dash lines indicate margins of errors, which are obtained using a SAE method (see part V of the lecture series in the sequel).


Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Patient


Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Patient