Advances in Applied Mathematics 应用数学进展, 2022, 11(6), 3826-3838
Published Online June 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.116410

Hans汉斯

基于SPSRs准则下的分类混合模型预测方法研究

肖新海

东华理工大学理学院，江西 南昌

收稿日期：2022年5月23日；录用日期：2022年6月15日；发布日期：2022年6月27日

摘 要

分类混合模型预测(CMMP)方法是近年来小区域估计领域中提出的一种新方法，该方法是在待预测效应识别后的基础上形成的方法，较传统的混合效应预测方法有更高的预测精度，得到许多统计学者的关注。最早的分类混合模型预测方法是基于均方预测误差(MSPE)准则进行分类识别构造最佳预测。MSPE准则虽然是一个具有较好数学性质(对称性和平滑性)的不确定性度量准则，但是其不是一个严格适当的评分准则(SPSRs)。因此，提出了基于SPSRs准则(即对数评分)进行分类识别，构造最佳预测的方法。首先，在最佳预测的基础上构造了SPSRs分类器，并进行识别预测；其次分析了该预测的渐近性质，并通过数值模拟证明了该方法较经典的回归预测方法具有更高的准确度；最后，给出实例进一步论证了我们的理论结果。

关键词

分类混合模型预测方法，均方预测误差，严格适当的评分准则

The Study of Classified Mixed Model Prediction Method Based on SPSRs Rules

Xinhai Xiao

School of Science, East China University of Technology, Nanchang Jiangxi

Received: May 23rd, 2022; accepted: Jun. 15th, 2022; published: Jun. 27th, 2022

Abstract

The Classified Mixed Model Prediction (CMMP) method is a newly proposed method in the field of small area estimation in recent years. The prediction accuracy of CMMP has attracted the attention of many statisticians. The earliest Classified Mixed Model Prediction is based on the Mean

文章引用: 肖新海. 基于 SPSRs 准则下的分类混合模型预测方法研究[J]. 应用数学进展, 2022, 11(6): 3826-3838.
DOI: 10.12677/aam.2022.116410

Square Prediction Error (MSPE) criterion for classification, identification and construction of the best prediction. Although the MSPE criterion is an uncertainty measurement criterion with good mathematical properties (symmetry and smoothness), it is not a Strictly Proper Scoring Rules (SPSRs). Therefore, we propose a method for classification and identification based on SPSRs criterion (*i.e.* logarithmic score) to construct the best prediction. Firstly, on the basis of the best prediction, the SPSRs classifier is constructed, identified and predicted. Secondly, the asymptotic properties of the prediction are analyzed, and the numerical simulation proves that this method has higher accuracy than the classical regression prediction method. Finally, an example is given to further demonstrate our theoretical results.

Keywords

Classified Mixed Model Prediction, Mean Square Prediction Error, Strictly Proper Scoring Rules

1. 引言

在现实生活中许多领域都涉及对相关特征预测的问题，而且其问题通常在个体层次或亚群体层次发生。在对特征预测问题分析时，可将新观测数据(含有特征信息的数据)在训练集(已有)数据中，找到一个与其相匹配的群组，即借用这个群组中训练集数据的相关信息来提高预测的准确性。基于这思想，Jiang等[1]提出了分类混合模型预测(Classified Mixed Model Prediction, CMMP)，通过数值模拟显示，在预测准确性方面上，CMMP 方法显著优于回归预测(Regression Prediction, RP)方法。CMMP 方法得到统计学者的认同，此外该方法在小区域估计领域存在广泛的应用，如刘育孜等[2]通过用 CMMP 方法对农作物面积的估算；王婕雯等[3]通过用 CMMP 方法对小麦面积估算。同时，基于 CMMP 思想，有许多文献对该方法进行了改进和推广，如 Sun 等[4]将适用于连续响应变量的线性混合模型的 CMMP 方法拓展到集群二分类数据(离散响应变量)，从而提出了分类混合逻辑模型预测方法。此外，Sun 等[5]又对该方法进行改进，结合协变量数据中信息进行分类匹配，提出新的分类混合模型预测方法，其预测准确性也优于(Mixed Model Prediction, MMP)方法[6]。

CMMP 方法的匹配判别策略度量是均方预测误差(Mean Squared Prediction Error, MSPE)，研究发现，在这种准则下，即使训练集数据中的某组数据和新观测数据的确来自同一群组，但是在判别匹配关系即I值时，仍有可能存在不正确的情况，由于这种匹配的错误率，从而影响 CMMP 方法预测的准确性。因此通过改善 CMMP 方法的匹配准则，提高匹配的正确率，就有可能提高 CMMP 方法预测的正确性。Gneiting等[7]提出严格适当评分准则(Strictly Proper Scoring Rules, SPSRs)。SPSRs 优点在于分布预测，比点预测更加实用，同时分布预测能提供更多有用的信息。SPSRs 准则也得到各学者的相应研究，比如 Merkle 和 Steyvers [8]，Landes [9]，Du, H.L. [10]等。根据 SPSRs 的定义，MSPE 准则满足“适当的(Proper)”的条件，但不满足“严格适当的(Strictly Proper)”条件，然而“严格适当的(Strictly Proper)”的评分规则优于“适当的(Proper)”的评分规则。从这个角度看，将 CMMP 方法中的匹配准则(MSPE)换成 SPSRs 准则，则匹配的正确率可能会提高，从而提高预测的准确性，为此我们提出 SPSRs 准则下分类混合效应预测记为 CMMP-SPSRs。在 SPSRs 中，关于分类变量的评分规则提出多种方法，如 Brier 评分(Brier score)，对数评分(Logarithmic score)和 0~1 评分(Zero-one score)等，本章选择其中的对数评分(Logs)进行相关分析。

本文内容安排如下：第二节将在嵌套误差回归模型中，提出 CMMP-SPSRs 方法的混合效应预测，并验证该方法预测的优良性。在第三节通过数值模拟分析，将 CMMP-SPSRs 方法与 RP 方法比较。第四节用一个真实数据对 CMMP-SPSRs 方法进行验证。

2. 基于 SPSRs 准则下分类混合模型预测

假设已知一组训练集数据为，$y_{ij}, i=1,\cdots,m;\ j=1,\cdots,n_i$ 并且知其分组情况，即 y_{ij} 属于第 i 组的第 j 个数据。假定训练集数据可采用嵌套误差回归(Nested-Error Regression, NER)模型来建模，如下所示

$$y_{ij}=x_{ij}^{\mathrm{T}}\beta+\alpha_i+\epsilon_{ij}, \tag{1}$$

其中 x_{ij} 是已知协变量向量，β 为未知的回归系数向量，即固定效应，α_i 是随机效应，ϵ_{ij} 为误差项。同时也假设随机效应 α_i 和误差项 ϵ_{ij} 相互独立，且有 $\alpha_i \sim N\left(0,\sigma_\alpha^2\right)$ 和 $\epsilon_{ij} \sim N\left(0,\sigma_\epsilon^2\right)$。

假设现有一组新观测数据为 $y_{n,1},\cdots,y_{n,j}, 1\le j\le n_{new}$ 其下标 n 与上文含义相同，只做符号区分并没有其他含义。假定新观测数据也可以嵌套误差回归模型来建模

$$y_{n,j}=x_n^{\mathrm{T}}\beta+\alpha_I+\epsilon_{n,j}, 1\le j\le n_{new}, \tag{2}$$

其中 x_n 为不依赖 j 的协变量；同时参数 β 和(1)式中的 β 相同；α_I 可能为：第一种情况(匹配 $I\in\{1,\cdots,m\}$)，与 $\alpha_i, 1\le i\le m$ 中的某一个相同，但并不知道具体哪一个，需要在训练集数据中寻找与新观测数据相匹配的群组；第二种情况(不匹配 $I\notin\{1,\cdots,m\}$)，一个全新的随机效应。假设 $E\left(\alpha_I\right)=0, var\left(\alpha_I\right)=\sigma_{n,\alpha}^2$ 有界。$\epsilon_{n,j}, 1\le j\le n_{new}$ 是新观测误差项且相互独立，并假设 $E\left(\epsilon_{n,j}\right)=0, var\left(\epsilon_{n,j}\right)=\sigma_{n,\epsilon}^2$ 有界，且和训练集数据中的 α_i，ϵ_{ij} 也相互独立。此外，α_I 和 $\epsilon_{n,j}$ 并不要求服从正态分布，同时也不要求 $\sigma_{n,\alpha}^2=\sigma_\alpha^2$，$\sigma_{n,\epsilon}^2=\sigma_\epsilon^2$。则需要预测的混合效应为

$$\theta=E\left(y_{n,j}\mid\alpha_I\right)=x_n^{\mathrm{T}}\beta+\alpha_I, \tag{3}$$

其中此时的参数 $\beta,\sigma_\alpha^2,\sigma_\epsilon^2$ 表示已知的；同时记 $\hat\beta,\hat\sigma_\alpha^2,\hat\sigma_\epsilon^2$ 分别表示其对应的一致估计，可以利用训练集数据得到 $\hat\beta,\hat\sigma_\alpha^2,\hat\sigma_\epsilon^2$。理由如下：由于新观测数据中的样本量通常不是很大，如果仅仅通过新观测数据提供的信息来对参数进行估计是不够的。虽然新观测数据不足，但是训练集数据的样本量通常是很多的。因此，通过训练集数据来对相关参数进行估计远优于新观测数据。很明显，如果参数估计值更准确，此时的经验最佳预测和最佳预测两者的预测也会更接近，同时得到的预测结果也更好。本章参数 $\beta,\sigma_\alpha^2,\sigma_\epsilon^2$ 统一采用极大似然估计。

2.1. 新观测数据与训练集存在匹配关系

假设新观测数据与训练集数据中的某一群组来自同一组群，即 $I=i\in\{1,\cdots,m\}$，但并不知道属于哪一组，故需要寻找具体的参数 I，可根据已知的训练集数据提供的相关信息来估计出这个具体的 I 值，即 $\hat I\in\{1,\cdots,m\}$，则新观测数据的混合效应(3)式可写成 $\theta=x_n^{\mathrm{T}}\beta+\alpha_i$，通过此表达式，可以看出此时新观测数据与训练集数据中的第 i 群组相匹配，故 y_{ij} 与(3)式混合效应的关系为：$y_1,\cdots,y_{i-1},\left(y_i^{\mathrm{T}},\theta\right)^{\mathrm{T}},y_{i+1},\cdots,y_m$。混合效应 $\theta=x_n^{\mathrm{T}}\beta+\alpha_i$ 的最佳预测，即 $\theta\mid y$，给定训练集数据后的条件期望，

$$\tilde\theta_{(i)}=E\left(\theta\mid y\right)=E\left(\theta\mid y_1,\cdots,y_m\right)=E\left(\theta\mid y_i\right).$$

根据模型的正态性假定，则有，

$$\begin{pmatrix}y_i\\ \alpha_i\end{pmatrix}\sim N\left(\begin{pmatrix}x_i^{\mathrm{T}}\beta\\ 0\end{pmatrix}\ \begin{pmatrix}\sigma_\alpha^2J_{n_i}+\sigma_\epsilon^2I_{n_i} & \sigma_\alpha^2 1_{n_i}\\ \sigma_\alpha^2 1_{n_i}^{\mathrm{T}} & \sigma_\alpha^2\end{pmatrix}\right),$$

其中训练集数据集为 $y_i=\left(y_{ij}\right)_{1\le j\le n_i}$ 并且 1_n 记为 n 维全 1 列向量， I_n 记为 n 阶单位矩阵， J_n 表示元素全为 1 的 n 阶矩阵。那么待预测新观测数据的混合效应最佳预测为

$$\begin{aligned}\mu_i=\tilde{\theta}_{(i)}&=E\left(x_n^{\mathrm{T}}\beta+\alpha_i\mid y_i\right)\\&=x_n^{\mathrm{T}}\beta+E\left(\alpha_i\mid y_i\right)\\&=x_n^{\mathrm{T}}\beta+\frac{n_i\sigma_\alpha^2}{n_i\sigma_\alpha^2+\sigma_\epsilon^2}\left(\overline{y}_{i\cdot}-\overline{x}_{i\cdot}^{\mathrm{T}}\beta\right).\end{aligned}\tag{4}$$

同时，其 $\theta\mid y$，给定训练集数据后的条件的方差为，

$$\begin{aligned}\sigma_i^2=Var\left(\theta\mid y\right)&=Var\left(\theta\mid y_1,\cdots,y_m\right)=Var\left(\theta\mid y_i\right)\\&=Var\left(x_n^{\mathrm{T}}\beta+\alpha_i\mid y_i\right)\\&=Var\left(x_n^{\mathrm{T}}\beta\right)+Var\left(\alpha_i\mid y_i\right)+2Cov\left(x_n^{\mathrm{T}}\beta,\alpha_i\mid y_i\right)\\&=\frac{\sigma_\alpha^2\sigma_\epsilon^2}{n_i\sigma_\alpha^2+\sigma_\epsilon^2}.\end{aligned}\tag{5}$$

将(4)式中的参数 $\beta,\sigma_\alpha^2,\sigma_\epsilon^2$ 分别用它们的一致估计 $\hat{\beta},\hat{\sigma}_\alpha^2,\hat{\sigma}_\epsilon^2$ 代替，则可得到经验最佳预测，

$$\hat{\mu}_i=x_n^{\mathrm{T}}\hat{\beta}+\frac{n_i\hat{\sigma}_\alpha^2}{\hat{\sigma}_\epsilon^2+n_i\hat{\sigma}_\alpha^2}\left(\overline{y}_{i\cdot}-\overline{x}_{i\cdot}^{\mathrm{T}}\hat{\beta}\right),\tag{6}$$

同样，将(5)式中的参数换成其对应的估计，则 $\theta\mid y$，给定训练集数据后的条件方差估计表达式为，

$$\hat{\sigma}_i^2=\frac{\hat{\sigma}_\alpha^2\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\epsilon^2+n_i\hat{\sigma}_\alpha^2},\tag{7}$$

综上所述，既有 $\theta\mid y$ 的期望也有方差，故其表达形式为： $\theta\mid y\sim N\left(\mu_i,\sigma_i^2\right)$。从而 $\theta\mid y$ 的密度函数为，

$$f\left(\theta\mid y\right)=f_i\left(\theta\mid y\right)=\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left\{-\frac{\left(\theta-\mu_i\right)^2}{2\sigma_i^2}\right\}.\tag{8}$$

根据 $\theta\mid y$ 的密度函数(8)式，又根据 SPSRs 的定义，可得对数评分(Logs)表达式为，

$$\log\left\{f_i\left(\theta\mid y\right)\right\}\propto-\left\{\log\sigma_i^2+\frac{\left(\mu_i-\theta\right)^2}{\sigma_i^2}\right\}.\tag{9}$$

接下来对参数 I 进行估计。采用 SPSRs 匹配准则来估计参数 I 进行估计，根据其定义有

$$\begin{aligned}\mathrm{SPSR}_i&=E\left\{\log\hat{\sigma}_i^2+\frac{\left(\hat{\mu}_i-\theta\right)^2}{\hat{\sigma}_i^2}\right\}\\&=E\left\{\log\hat{\sigma}_i^2+\frac{\left(\hat{\mu}_i-\overline{y}_n+\overline{\epsilon}_n\right)^2}{\hat{\sigma}_i^2}\right\}\\&=E\left\{\log\hat{\sigma}_i^2+\frac{\left(\hat{\mu}_i-\overline{y}_n\right)^2}{\hat{\sigma}_i^2}+\frac{2\left(\hat{\mu}_i-\overline{y}_n\right)\overline{\epsilon}_n}{\hat{\sigma}_i^2}+\frac{\left(\overline{\epsilon}_n\right)^2}{\hat{\sigma}_i^2}\right\}\\&=E\left\{\log\hat{\sigma}_i^2+\frac{\left(\hat{\mu}_i-\overline{y}_n\right)^2}{\hat{\sigma}_i^2}+\frac{\sigma_\epsilon^2}{\left(n_{new}\hat{\sigma}_i^2\right)}\right\}.\end{aligned}\tag{10}$$

那么与(10)式相对应的 SPSRs 即为期望符号内的表达式，从而提出基于 SPSRs 的匹配准则

$$\hat{I}_{LogS}=\arg\min_{1\le i\le m}\left\{\log\hat{\sigma}_i^2+\frac{\left(\hat{\mu}_i-\overline{y}_n\right)^2}{\hat{\sigma}_i^2}+\frac{\hat{\sigma}_\epsilon^2}{\left(n_{new}\hat{\sigma}_i^2\right)}\right\},\tag{11}$$

其中 $\theta=\overline{y}_n-\overline{\epsilon}_n$，$\overline{y}_n=n_{new}^{-1}\sum_{j=1}^{n_{new}}y_{n,j}$，$\overline{\epsilon}_n=n_{new}^{-1}\sum_{j=1}^{n_{new}}\epsilon_{n,j}$。

最终，通过将 $\hat{I}_{LogS}$ 替换(6)式中的 i，则 θ 的分类混合效应预测(Classified Mixed-Effect Predictor, CMEP)为 $\hat{\theta}=\hat{\mu}_{\hat{I}}$。

下面的定理阐述了当给定一些合理条件下，则可保持分类混合效应预测的渐进性质。

记 $N=\left(m,n_i,1\le i\le m,n_{new}\right)$ 是本文中所涉及的全部样本容量。假设方差参数 ψ 的参数空间为 Ψ，而参数 β 的参数空间是 p 维的欧几里得空间 R^p。$\|A\|=\sqrt{\lambda_{\max}\left(A'A\right)}$ 表示矩阵 A 的范数。让 $\mu_{(0i')}$ 表示(4)式等号右边的部分，但其中 i 被 i' 替换，并且参数 β,ψ 为真参数向量。当参数 β,ψ 换成其一致估计 $\hat{\beta},\hat{\psi}$ 时，则对应其经验最佳预测，记为 $\hat{\mu}_{(i')}$。做如下假设：

A1. 当 $I=i\in\{1,\cdots,m\}$ 情形下，将 i 换成 i' 时，模型(1)依然成立；下标由 I 换成 i 时，模型(2)也成立。

A2. 参数空间 Ψ 的内点为参数 ψ 的真值 ψ_0。

A3. $x_{ijk},1\le i\le m,1\le j\le n_i$ 是有界的，其中 x_{ijk} 为 x_{ij} 的第 k 个元素。

A4. 记 $R_{\max}=\max_{1\le i\le m}n_i$，$R_{\min}=\min_{1\le i\le m}n_i$，若 $n_{new}\to\infty,R_{\min}\to\infty$ 时，$R_{\max}-R_{\min}\to 0$，且 $\hat{\beta}-\beta_0=O_p\left(a_N\right)$，$\hat{\psi}-\psi_0=O_p\left(b_N\right)$，$\max_{1\le i\le m}\left|\alpha_i\right|=O_p\left(c_N\right)$，而且当 $\psi=\psi_0$ 时，$\max_{1\le i\le m}\left|I_{n_i}\epsilon_i\right|=O_p\left(d_N\right)$ 其中 a_N,b_N 等为正常数，且使得 $d_N\left(a_N+b_N+c_Nb_N\right)\to 0$，$d_N^2\to 0$。为了不失一般性，假定 $c_N\ge 1$。

定理 1. 如果假设 A1~A4 都成立，那么有 $\hat{\mu}_{\hat{I}}\xrightarrow{p}\theta$，即 SPSRs 准则下的分类混合效应预测满足渐进性。

证明：首先考虑当 $I=i\in\{1,\cdots,m\}$ 存在匹配关系时的情形。通过本文上述理论的推导与证明，从而可知新观测数据的分类混合效应预测(CMEP)可表示为 $\hat{\mu}_{\hat{I}}=x_n^{\mathrm{T}}\hat{\beta}+n_{\hat{I}}\hat{\sigma}_\alpha^2/\hat{\sigma}_\epsilon^2+n_{\hat{I}}\hat{\sigma}_\alpha^2\left(\overline{y}_{\hat{I}\cdot}-\overline{x}_{\hat{I}\cdot}^{\mathrm{T}}\hat{\beta}\right)$。接下来证明 $\hat{\mu}_{\hat{I}}\xrightarrow{p}\theta$。

记 $b=\sigma_\epsilon^2/\sigma_\alpha^2$，则有 $\sigma_i^2=\sigma_\epsilon^2\cdot\sigma_\alpha^2/\sigma_\epsilon^2+n_i\sigma_\alpha^2=\sigma_\epsilon^2/b+n_i$，同时也有 $B_i=n_i\hat{\sigma}_\alpha^2/\hat{\sigma}_\epsilon^2+n_i\hat{\sigma}_\alpha^2=n_i/b+n_i\le 1$。设 $B_{0i'}$ 表示 $B_{i'}$ 中 $\sigma_\alpha^2=\sigma_{\alpha0}^2$，$\sigma_\epsilon^2=\sigma_{\epsilon0}^2$，而有

$$\begin{aligned}\mu_{(0i')}&=x_n^{\mathrm{T}}\beta_0+B_{0i'}\left(\overline{y}_{i'\cdot}-\overline{x}_{i'\cdot}^{\mathrm{T}}\beta_0\right)\\&=x_n^{\mathrm{T}}\beta_0+\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\left(y_{i'}-x_{i'}\beta_0\right)\\&=x_n^{\mathrm{T}}\beta_0+\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\left(1_{n_{i'}}\alpha_{i'}+\epsilon_{i'}\right),\\\theta&=x_n^{\mathrm{T}}\beta+\alpha_I=x_n^{\mathrm{T}}\beta_0+\alpha_i,\end{aligned}$$

因此，有

$$\begin{aligned}\mu_{(0i')}-\theta&=x_n^{\mathrm{T}}\beta_0+\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\left(1_{n_{i'}}\alpha_{i'}+\epsilon_{i'}\right)-x_n^{\mathrm{T}}\beta_0-\alpha_i\\&=\alpha_{i'}-\alpha_i+\left(\frac{n_{i'}}{n_{i'}+b_0}-I_q\right)\alpha_{i'}+\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\epsilon_{i'}\\&=\alpha_{i'}-\alpha_i-\frac{b_0}{n_{i'}+b_0}\alpha_{i'}+\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\epsilon_{i'},\end{aligned}\tag{12}$$

另外，通过泰勒公式展开有

$$\begin{aligned}\hat{\mu}_{(i')}-\mu_{(0i')}&=x_n^{\mathrm{T}}\hat{\beta}+\frac{1}{n_{i'}+\hat{b}}1_{n_{i'}}^{\mathrm{T}}\left(y_{i'}-x_{i'}\hat{\beta}\right)-x_n^{\mathrm{T}}\beta_0-\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\left(y_{i'}-x_{i'}\beta_0\right)\\&=\frac{\partial\hat{\mu}_{(i')}}{\partial\beta'}\left(\hat{\beta}-\beta_0\right)+\frac{\partial\hat{\mu}_{(i')}}{\partial\psi'}\left(\hat{\psi}-\psi_0\right),\end{aligned}\tag{13}$$

其中 $\partial\hat{\mu}_{(i)}/\partial\beta'$ 表示 $\partial\hat{\mu}_{(i)}/\partial\beta$ 在点 $\left(\beta',\psi'\right)'$ 上的微分，其中 $\left(\beta',\psi'\right)'$ 介于 $\left(\beta_0',\psi_0'\right)'$ 与 $\left(\hat{\beta}',\hat{\psi}'\right)'$ 之间。

通过结合上述的(12)式和(13)式，同时又根据上面理论可知 $\bar{y}_n=\theta+\bar{\varepsilon}_n$。从而有

$$\begin{aligned}\hat{\mu}_{(i')}-\bar{y}_n&=\hat{\mu}_{(i')}-\theta-\bar{\epsilon}_n=\hat{\mu}_{(i')}-\mu_{(0i')}+\mu_{(0i')}-\theta-\bar{\epsilon}_n\\&=\alpha_{i'}-\alpha_i-\bar{\epsilon}_n-\frac{b_0}{n_{i'}+b_0}\alpha_{i'}+\frac{1}{n_{i'}+b_0}1_{n_{i'}}^{\mathrm{T}}\epsilon_{i'}+\frac{\partial\hat{\mu}_{(i')}}{\partial\beta'}\left(\hat{\beta}-\beta_0\right)+\frac{\partial\hat{\mu}_{(i')}}{\partial\psi'}\left(\hat{\psi}-\psi_0\right)\\&=\alpha_{i'}-\alpha_i-\bar{\epsilon}_n+\xi_{i'},\end{aligned}\tag{14}$$

其中根据上述的假设有 $\xi_{i'}\le O_p\left(a_N\right)+O_p\left[\left(c_N+1\right)\cdot b_N\right]+O_p\left(d_N\right)$。

另一方面 $\hat{I}$ 使 $\log\hat{\sigma}_{i'}^2+\left(\hat{\mu}_{i'}-\bar{y}_n\right)^2/\left(\hat{\sigma}_{i'}^2+\hat{\sigma}_\epsilon^2/\left(n_{new}\hat{\sigma}_{i'}^2\right)\right.$ 在 $1\le i'\le m$ 中达到最小，即可表示为

$$\begin{aligned}\hat{I}&=\arg\min_{1\le i'\le m}\left\{\log\hat{\sigma}_{i'}^2+\frac{\left(\hat{\mu}_{i'}-\bar{y}_n\right)^2}{\hat{\sigma}_{i'}^2}+\frac{\hat{\sigma}_\epsilon^2}{n_{new}\hat{\sigma}_{i'}^2}\right\}\\&=\arg\min_{1\le i'\le m}\left\{\log\hat{\sigma}_{i'}^2+\frac{\left(\hat{\mu}_{i'}-\bar{y}_n\right)^2}{\hat{\sigma}_{i'}^2}+\frac{\hat{b}+n_{i'}}{n_{new}}\right\}\\&=\arg\min_{1\le i'\le m}\left\{-\log\left(\hat{b}+n_{i'}\right)+\frac{\left(\hat{\mu}_{i'}-\bar{y}_n\right)^2\left(\hat{b}+n_{i'}\right)}{\hat{\sigma}_\epsilon^2}+\frac{\hat{b}+n_{i'}}{n_{new}}\right\},\end{aligned}$$

因此，有

$$\begin{aligned}&-\log\left(\hat{b}+n_{\hat{I}}\right)+\frac{\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\left(\hat{b}+n_{\hat{I}}\right)}{\hat{\sigma}_\varepsilon^2}+\frac{\hat{b}+n_{\hat{I}}}{n_{new}}\\&\le-\log\left(\hat{b}+n_I\right)+\frac{\left(\hat{\mu}_I-\bar{y}_n\right)^2\left(\hat{b}+n_I\right)}{\hat{\sigma}_\epsilon^2}+\frac{\hat{b}+n_I}{n_{new}},\end{aligned}\tag{15}$$

整理得

$$\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\left(\hat{b}+n_{\hat{I}}\right)-\left(\hat{\mu}_I-\bar{y}_n\right)^2\left(\hat{b}+n_I\right)\le\hat{\sigma}_\epsilon^2\left[\log\frac{\hat{b}+n_{\hat{I}}}{\hat{b}+n_I}+\frac{n_I-n_{\hat{I}}}{n_{new}}\right],$$

当 $n_i=n_j\left(i=j\right)$ (表示 n_i 是一个固定值，只与参数 i 相关)时，显然有 $\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\le\left(\hat{\mu}_I-\bar{y}_n\right)^2$；如果当 $n_i\ne n_j\left(i\ne j\right)$ 时，记 $R_{\max}=\max_{1\le i\le m}n_i$，$R_{\min}=\min_{1\le i\le m}n_i$，如果有 $n_{new}\to\infty$，$\min_{1\le i\le m}n_i\to\infty$，$R_{\max}-R_{\min}\to0$，则有 $\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\le\left(\hat{\mu}_I-\bar{y}_n\right)^2$。故，综上所述有，

$$\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\le\left(\hat{\mu}_I-\bar{y}_n\right)^2.\tag{16}$$

令 $\xi_{\hat{I}}$ 和 ξ_I 分别表示 ξ_i 项中参数 i 换成其对应参数 $\hat{I}$ 和 I；同时为了方便区分，故将 $\xi_{i'}$ 用 s_N 来表示。则一方面通过(14)式有

$$\begin{aligned}\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2&=\left(\alpha_{\hat{I}}-\alpha_I-\bar{\epsilon}_n+\xi_{\hat{I}}\right)^2\\&=\left(\alpha_{\hat{I}}-\alpha_I-\bar{\epsilon}_n\right)^2+\xi_{\hat{I}}^2+2\left(\alpha_{\hat{I}}-\alpha_I-\bar{\epsilon}_n\right)\cdot\xi_{\hat{I}}\\&=\left(\alpha_{\hat{I}}-\alpha_I-\bar{\epsilon}_n\right)^2-2\xi_{\hat{I}}\cdot\bar{\epsilon}_n+\xi_{\hat{I}}^2+2\left(\alpha_{\hat{I}}-\alpha_I\right)\xi_{\hat{I}}\\&\geq\left(\alpha_{\hat{I}}-\alpha_I-\bar{\epsilon}_n\right)^2-2\xi_{\hat{I}}\cdot\bar{\epsilon}_n-O_p\left(c_N\right)s_N,\end{aligned}\tag{17}$$

另一方面，同理，通过(14)式有

$$\begin{aligned}\left(\hat{\mu}_I-\bar{y}_n\right)^2&=\left(\alpha_I-\alpha_I-\bar{\epsilon}_n+\xi_{\hat{I}}\right)^2\\&=\left(-\bar{\epsilon}_n+\xi_I\right)^2\\&=\bar{\epsilon}_n^2-2\xi_I\cdot\bar{\epsilon}_n+\xi_I^2\\&\leq\bar{\epsilon}_n^2-2\xi_I\cdot\bar{\epsilon}_n+s_N^2,\end{aligned}\tag{18}$$

结合(16)，(17)和(18)式，有

$$\begin{aligned}&\left(\alpha_{\hat{I}}-\alpha_I-\bar{\varepsilon}_n\right)^2-2\xi_{\hat{I}}\cdot\bar{\epsilon}_n-O_p\left(c_N\right)s_N\leq\bar{\epsilon}_n^2-2\xi_I\cdot\bar{\epsilon}_n+s_N^2\\&\left(\alpha_{\hat{I}}-\alpha_I\right)^2-2\left(\alpha_{\hat{I}}-\alpha_I\right)\cdot\bar{\epsilon}_n+\bar{\epsilon}_n^2\leq2\xi_{\hat{I}}\cdot\bar{\epsilon}_n-2\xi_I\cdot\bar{\varepsilon}_n+\bar{\epsilon}_n^2+O_p\left(c_N\right)s_N+s_N^2\\&\left(\alpha_{\hat{I}}-\alpha_I\right)^2\leq2\left(\alpha_{\hat{I}}-\alpha_I+\xi_{\hat{I}}-\xi_I\right)\cdot\bar{\epsilon}_n+O_p\left(c_N\right)s_N+s_N^2,\end{aligned}\tag{19}$$

根据 $\hat{\mu}_{\hat{I}}-\theta=\hat{\mu}_{\hat{I}}-\bar{y}_n+\bar{y}_n-\theta=\left(\alpha_{\hat{I}}-\alpha_I\right)-\bar{\epsilon}_n+\xi_{\hat{I}}+\bar{\epsilon}_n=\left(\alpha_{\hat{I}}-\alpha_I\right)+\xi_{\hat{I}}$，再结合(19)式可得

$$\begin{aligned}\left(\hat{\mu}_{\hat{I}}-\theta\right)^2&=\left(\left(\alpha_{\hat{I}}-\alpha_I\right)+\xi_{\hat{I}}\right)^2\\&=\left(\alpha_{\hat{I}}-\alpha_I\right)^2+2\left(\alpha_{\hat{I}}-\alpha_I\right)\cdot\xi_{\hat{I}}+\xi_{\hat{I}}^2\\&\leq2\left(\alpha_{\hat{I}}-\alpha_I+\xi_{\hat{I}}-\xi_I\right)\cdot\bar{\epsilon}_n+O_p\left(c_N\right)s_N+s_N^2+2\left(\alpha_{\hat{I}}-\alpha_I\right)\cdot\xi_{\hat{I}}+\xi_{\hat{I}}^2\\&\leq2\left(\alpha_{\hat{I}}-\alpha_I+\xi_{\hat{I}}-\xi_I\right)\cdot\bar{\epsilon}_n+O_p\left(c_N\right)s_N+2s_N^2.\end{aligned}\tag{20}$$

设 $\hat{\eta}=\xi_{\hat{I}}-\bar{\epsilon}_n$ 和 $\eta_1=\xi_{i'}-\bar{\epsilon}_n$，则有 $|\hat{\eta}|\vee|\eta_1|\leq s_N+|\bar{\epsilon}_n|=o_p(1)$。又假设存在 $\mathcal{A}=\left\{\left|\left(\alpha_{\hat{I}}-\alpha_I\right)\right|\geq1,|\hat{\eta}|\leq1/4\right\}$，则一方面有

$$\left\{\hat{\mu}_{\hat{I}}-\bar{y}_n\right\}^2=\left\{\left(\alpha_{\hat{I}}-\alpha_I\right)+\hat{\eta}\right\}^2\geq(1/2)\left\{\left(\alpha_{\hat{I}}-\alpha_I\right)\right\}^2.$$

另一方面有

$$\left\{\hat{\mu}_{i'}-\bar{y}_n\right\}^2=\left\{\left(\alpha_{i'}-\alpha_I\right)+\eta_1\right\}^2\leq2\left[\left(\alpha_{i'}-\alpha_I\right)^2+\eta_1^2\right].$$

由此结合(16)式得到

$$\begin{aligned}&\left\{\hat{\mu}_{\hat{I}}-\bar{y}_n\right\}^2\leq\left\{\hat{\mu}_{i'}-\bar{y}_n\right\}^2\\&\frac{1}{2}\left\{\left(\alpha_{\hat{I}}-\alpha_I\right)\right\}^2\leq2\left[\left(\alpha_{i'}-\alpha_I\right)^2+\eta_1^2\right]\\&\left\{\left(\alpha_{\hat{I}}-\alpha_I\right)\right\}^2\leq4\left[\left(\alpha_{i'}-\alpha_I\right)^2+\eta_1^2\right]\\&\left|\alpha_{\hat{I}}-\alpha_I\right|\leq2\sqrt{\left(\alpha_{i'}-\alpha_I\right)^2+\eta_1^2},\end{aligned}$$

由此可见，当 $|\hat{\eta}|\leq1/4$ 情况成立时，则有 $\left|\left(\alpha_{\hat{I}}-\alpha_I\right)\right|=O_p(1)$，$\left|\alpha_{\hat{I}}-\alpha_I\right|\leq1\vee\left[2\sqrt{\left(\alpha_{i'}-\alpha_I\right)^2+\eta_1^2}\right]=O_p(1)$。

因此，回到最初的问题中，即由(20)式可知 $\left(\hat{\mu}_{\hat{I}}-\theta\right)^2 \le 2\left[O_p(1)+o_p(1)\right]o_p(1)+O_p(c_N)s_N+2s_N^2=o_p(1)$。从而，定理得证。

2.2. 新观测数据与训练集数据不匹配

在不假设匹配关系存在的情况下，则匹配关系可能存在也有可能不存在，即不存在新观测数据与训练集数据中的某一组群相匹配的前提，这比较符合实际。如果匹配关系存在则可按 2.1 节推导；当匹配关系不存在时，即 $I \ne i \in \{1,\cdots,m\}$，则新观测数据与训练集数据是相互独立的。接下来讨论匹配关系不存在的情况。

如果不匹配，则此时的混合效应为(3)式，从而有 (θ,ϵ_n) 与训练集数据相互独立，故可得的最佳预测为，

$$\tilde{\theta}=E(\theta\mid y)=E(\theta\mid y_1,\cdots,y_m)=E(\theta),$$

同理，与匹配情况一样，BP 为，

$$\tilde{\mu}=\tilde{\theta}=x_n^{\mathrm{T}}\beta, \tag{21}$$

其 $\theta\mid y$，方差为

$$\begin{aligned}\sigma^2 &= Var(\theta\mid y)=Var(\theta\mid y_1,\cdots,y_m)=Var(\theta)\\ &= Var\left(x_n^{\mathrm{T}}\beta+\alpha_I\right)\\ &= Var\left(x_n^{\mathrm{T}}\beta\right)+Var(\alpha_I)+2Cov\left(x_n^{\mathrm{T}}\beta,\alpha_I\right)\\ &= \sigma_{n,\alpha}^2.\end{aligned} \tag{22}$$

将(21)的参数 β 用它的一致估计 $\hat{\beta}$ 代替，则可 θ 得到相应的经验最佳预测 $\hat{\mu}=x_n^{\mathrm{T}}\hat{\beta}$。

同样，将(22)式中的参数 $\sigma_{n,\alpha}^2$ 换成其对应的一致估计 $\hat{\sigma}_{n,\alpha}^2$，则有 $\theta\mid y$，的方差估计为：$\hat{\sigma}^2=\hat{\sigma}_{n,\alpha}^2$。

综上所述，可得 $\theta\mid y$ 表达形式为：$\theta\mid y\sim N\left(\mu,\sigma^2\right)$。从而，$\theta\mid y$ 的密度函数可以表示为

$$f(\theta\mid y)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\theta-\mu)^2}{2\sigma^2}\right\}. \tag{23}$$

同理，这种情况下，则对数评分(Logs)表达式为

$$\log\left\{f(\theta\mid y)\right\}\propto-\left\{\log\sigma^2+\frac{(\mu-\theta)^2}{\sigma^2}\right\}. \tag{24}$$

与匹配情况中的(10)式的推导过程相同，可得出此时的 SPSRs 为：

$$\begin{aligned}\mathrm{SPSRs} &= E\left\{\log\hat{\sigma}^2+\frac{(\hat{\mu}-\theta)^2}{\hat{\sigma}^2}\right\}\\ &= E\left\{\log\hat{\sigma}^2+\frac{\left(\hat{\mu}-\bar{y}_n+\bar{\epsilon}_n\right)^2}{\hat{\sigma}^2}\right\}\\ &= E\left\{\log\hat{\sigma}^2+\frac{(\hat{\mu}-\bar{y}_n)^2}{\hat{\sigma}^2}+\frac{2(\hat{\mu}-\bar{y}_n)\bar{\epsilon}_n}{\hat{\sigma}^2}+\frac{(\bar{\epsilon}_n)^2}{\hat{\sigma}^2}\right\}\\ &= E\left\{\log\hat{\sigma}^2+\frac{(\hat{\mu}-\bar{y}_n)^2}{\hat{\sigma}^2}+\frac{\sigma_\epsilon^2}{\left(n_{new}\hat{\sigma}^2\right)}\right\}.\end{aligned} \tag{25}$$

将(10)式和(16)式比较，区别是：将 $\hat{\mu}_i$ 换成 $\hat{\mu}$；$\hat{\sigma}_i^2$ 换成 $\hat{\sigma}^2$。故将之前的方法做如下延伸：令 $\hat{I}$ 由(3.11)式给出，比较 $\log\hat{\sigma}_{\hat{I}}^2+\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\Big/\hat{\sigma}_{\hat{I}}^2+\sigma_\epsilon^2\Big/\left(n_{new}\hat{\sigma}_{\hat{I}}^2\right)$ 和 $\log\hat{\sigma}^2+\left(\hat{\mu}-\bar{y}_n\right)^2\Big/\hat{\sigma}^2+\sigma_\epsilon^2\Big/\left(n_{new}\hat{\sigma}^2\right)$ 的大小。如果前者更小，则 θ 的 CMEP 为 $\hat{\mu}_{\hat{I}}$；否则，θ 的 CMEP 是 $\hat{\mu}$。

3. 数值模拟分析

本节主要对 SPSRs 准则下 CMEP 的预测效果进行分析，主要将 CMEP-SPSRs 方法与 PR 方法的 MSPE 结果进行比较，讨论其预测效果。针对新观测数据与训练数据是否匹配两种情形分别讨论，具体步骤如下：

1) 新观测数据与训练集数据存在匹配关系

步骤 1：训练集数据 y_{ij} 按照(1)式生成，其中 $n_i=5,1\le i\le m$。给定 $\beta=(5,1)^{\mathrm{T}}$，$x_{ij}=\left(1,x_{ij2}\right)^{\mathrm{T}}$，协变量 x_{ij2} 服从 $N(0,1)$ 的随机数；群组特定的随机效应 α_i 服从 $N\left(0,\sigma_\alpha^2\right)$ 的随机数，其中给定 σ_α^2；误差项 ϵ_{ij} 服从 $N\left(0,\sigma_\epsilon^2\right)$ 的随机数，其中给定 σ_ϵ^2。

步骤 2：新观测数据 $y_{n,j}$ 按照(2)式生成，其中给定 $\beta=(5,1)^{\mathrm{T}}$，$x_n=\left(1,x_{n2}\right)^{\mathrm{T}}$，协变量 x_{n2} 服从 $N(0,1)$ 的随机数；将(1)式中 α_i，ϵ_{ij} 分别替换为 α_I，$\epsilon_{n,j}$。故需要预测的混合效应为(3)式。

步骤 3：先基于(4)式可获得 θ 的最佳预测 $\tilde{\theta}_{(i)}$，同时又根据(5)式可得到 $\theta\mid y$ 的方差为 σ_i^2；然后得到参数的最大似然估计，$\hat{\beta},\hat{\sigma}_\alpha^2,\hat{\sigma}_\epsilon^2$。最后将参数估计代回(4)与(5)式，则分别可获得 θ 的经验最佳预测 $\hat{\mu}_i$ 和 $\theta\mid y$ 方差估计 $\hat{\sigma}_i^2$。

步骤 4：通过(11)式获得 $\hat{I}_{LogS}$。接着将 $\hat{I}_{LogS}$ 替换(6)式中的 i，则得到 θ 的分类混合效应预测(CMEP)：$\hat{\theta}=\hat{\mu}_{\hat{I}}$。

步骤 5：记 $\hat{\theta}_{n,r}=x_n^{\mathrm{T}}\hat{\beta}_{LS}$ 表示 θ_n 的标准回归预测(RP)，其中 $\hat{\beta}_{LS}$ 表示为 β 的最小二乘(Least-Square, LS)估计。因此要计算出 $\hat{\beta}_{LS}$，从而得到 RP。

步骤 6：分别计算 CMMP-SPSRs 方法和 RP 方法的 MSPE：$\mathrm{MSPE}_1=\left(\hat{\theta}-\theta\right)^2$，$\mathrm{MSPE}_2=\left(\hat{\theta}_{n,r}-\theta\right)^2$。

通过 Matlab 软件按以上步骤进行编程，除步骤 2 以外，所有步骤重复 100 次，然后获得 CMMP-SPSRs 方法和 RP 方法预测混合效应的 MSPE 平均值，这样有利于降低误差，最后将 MSPE 的平均值作为评价预测准确性的指标。结果如表 1~3 中所示。%Improve 为 CMMP-SPSRs 方法和 RP 方法 MSPE 的相对大小。

Table 1. Comparison of two MSPE methods with matching relationship, fixed at $m=50,n_{new}=5,\sigma_\epsilon^2=1$

表 1. 存在匹配关系两种方法 MSPE 的比较，固定 $m=50,n_{new}=5,\sigma_\epsilon^2=1$

σ_α^2	0.25	0.5	1	2	4
CMMP-SPSRs	0.0420	0.0705	0.2416	0.6053	0.7052
RP	0.1588	0.3057	1.2803	3.8752	5.1652
%Improve	277.87	333.71	429.93	540.25	632.41

Table 2. Comparison of two MSPE methods with matching relationship, fixed at $m=50,n_{new}=5,\sigma_\alpha^2=1$

表 2. 存在匹配关系两种方法 MSPE 的比较，固定 $m=50,n_{new}=5,\sigma_\alpha^2=1$

σ_ϵ^2	0.25	0.5	1	2	4
CMMP-SPSRs	0.0243	0.0586	0.2451	0.3074	0.5982
RP	0.3101	0.4462	1.7147	1.7386	2.6358
%Improve	1176.13	661.92	599.49	465.58	340.62

Table 3. Comparison of two MSPE methods with matching relationship, fixed at $m=50,\sigma_\epsilon^2=1,\sigma_\alpha^2=1$
表 3. 存在匹配关系两种方法 MSPE 的比较，固定 $m=50,\sigma_\epsilon^2=1,\sigma_\alpha^2=1$

n_{new}	1	5	10	50	100
CMMP-SPSRs	0.721	0.2327	0.1799	0.1141	0.0924
RP	1.7852	1.5729	1.5089	1.9622	1.7549
%Improve	147.60	157.29	738.74	1619.72	1799.24

$$\%\text{Improve}=\left(\frac{\text{MSPE of RP}-\text{MSPE of CMMP}}{\text{MSPE of CMMP}}\right)\times 100\%.$$

从表 1~3 中数据可以发现 RP 方法的 MSPE 都大于 CMMP-SPSRs 方法的 MSPE。具体来说，表 1：当只有随机效应方差 σ_α^2 改变时，随着 σ_α^2 值增加，两者方法的 MSPE 都有增大，但通过%Improve 可以发现 RP 方法的 MSPE 增大的速度高于 CMMP-SPSRs 方法；同时在 MSPE 方面上，发现 CMMP-SPSRs 方法小于 RP 方法，故 CMMP-SPSRs 方法优于 RP 方法。表 2：当只有误差项方差 σ_ϵ^2 改变时，随着 σ_ϵ^2 值增加，CMMP-SPSRs 方法和 RP 方法的 MSPE 也增加，但 CMMP-SPSRs 方法始终小于 RP 方法，即 CMMP-SPSRs 方法更优。表 3：当只有 n_{new} 改变时，从数据中可以看出，CMMP-SPSRs 方法混合效应预测的准确性随着 n_{new} 值增加而提高，这符合实际，当新观测数据增多，从而预测的准确性也会增加。RP 方法并没有因为 n_{new} 的改变而发生明显的变化，且 MSPE 值一直比 CMMP-SPSRs 方法大，因此 CMMP-SPSRs 方法优于 RP 方法。

2) 新观测数据与训练集数据不匹配

步骤 1： 训练集数据 y_{ij} 可以按照模型(1)式生成，其中 $n_i=5,1\le i\le m$。同时给定 $\beta=(1,2,3)^{\mathrm{T}}$，$x_{ij}=\left(1,x_{1,ij},x_{2,ij}\right)^{\mathrm{T}}$，协变量 $x_{1,ij},x_{2,ij}$ 由 $N(0,1)$ 分布产生的随机数；群组特定的随机效应 α_i 服从 $N\left(0,\sigma_\alpha^2\right)$ 的随机数，其中给定 σ_α^2；误差项 ϵ_{ij} 服从 $N\left(0,\sigma_\epsilon^2\right)$ 的随机数，其中给定 σ_ϵ^2。

步骤 2： 新观测数据 $y_{n,j}$ 可以按照模型(2)式生成，其中同时给定 $\beta=(1,2,3)^{\mathrm{T}}$，$x_n=\left(1,x_{1,n},x_{2,n}\right)^{\mathrm{T}}$，协变量 $x_{1,n},x_{2,n}$ 由 $N(0,1)$ 分布产生的随机数；将(1)式中 α_i，ϵ_{ij} 分别替换为 α_I，$\epsilon_{n,j}$。故需要预测的混合效应为(3)式。

步骤 3： 首先基于(21)式获得 θ 的最佳预测 $\tilde{\theta}$，同时通过(22)式得到 σ^2。然后得到参数的 MLE。最后可获得 θ 的经验最佳预测 $\hat{\mu}$ 和 $\hat{\sigma}^2$。

步骤 4： 与匹配关系中的步骤 5 相同，得到 RP。

步骤 5： 基于(1)匹配情况下得到 $\hat{I}$ 和 θ 的分类混合效应预测：$\hat{\theta}=\hat{\mu}_{\hat{I}}$。同时计算 $\log\hat{\sigma}_{\hat{I}}^2+\left(\hat{\mu}_{\hat{I}}-\bar{y}_n\right)^2\Big/\hat{\sigma}_{\hat{I}}^2+\sigma_\epsilon^2\Big/n_{new}\hat{\sigma}_{\hat{I}}^2$ 和 $\log\hat{\sigma}^2+\left(\hat{\mu}-\bar{y}_n\right)^2\Big/\hat{\sigma}^2+\sigma_\epsilon^2\Big/n_{new}\hat{\sigma}^2$ 的大小。如果前者较小，则 θ 的分类混合效应预测为 $\hat{\mu}_{\hat{I}}$；否则，θ 的分类混合效应预测是 $\hat{\mu}$。

步骤 6： 与(1)匹配中的步骤 6 一样。

通过表 4 和表 5 中的数据比较可以发现，在不匹配情况下，两者的 MSPE 的变化趋势和匹配情况相类似，即 CMMP-SPSRs 方法相对 RP 方法依旧存在优越，同时%Improve 的变化趋势也是相似的。由此，可以得出不管新观测数据和训练集数据中是否存在匹配关系，CMMP-SPSRs 方法相对于 RP 方法预测都更具有优越性。

4. 实例应用分析

本节以电视学校和家庭预防与戒烟项目(Television School and Family Smoking Prevention and Cessa-

tion Project, TVSFP) [11]的数据来对 CMMP-SPSRs 方法验证，其中研究对象是来自美国加利福尼亚州洛杉矶和圣地亚哥学校的七年级学生。最初该项目主要以学校为基础的社会抵抗课程和以电视为基础的预防和戒烟方面的独立与联合效果研究。为了对嵌套误差回归模型的演示，选择 TVSFP 数据中的一个子集，即取洛杉矶中的 28 所学校，这些学校被随机分配为四种研究条件中的一种：1) 抵制社会的课程(social-resistance classroom curriculum, CC)；2) 电视干预(television intervention, TV)；3) 结合 CC 和 TV；4) 不做任何处理。烟草与健康知识量表(Tobacco and Health knowledge scale, THKS)得分是主要研究结果变量之一，作为本次实验的响应变量。THKS 是由七个问卷项目组成，用于评估学生的烟草和健康知识。目前数据只涉及干预前和干预后时间点完成的 THKS。该数据是来自 28 所学校的 1600 名学生，每所学校有 1 至 13 间教室，每个教室有 2 至 28 名学生。该数据可在 http://www.hsph.harvard.edu/fitzmaur/ala/tvsfp.txt 上查找。

Table 4. Comparison of two MSPE methods without matching relationship, fixed at $m=50, n_{new}=5, \sigma_{\varepsilon}^2=1$
表 4. 不存在匹配关系两种方法 MSPE 的比较，固定 $m=50, n_{new}=5, \sigma_{\varepsilon}^2=1$

σ_{α}^2	0.25	0.5	1	2	4
CMMP-SPSRs	0.0921	0.1205	0.3383	0.6160	0.6863
RP	0.1849	0.5785	1.7598	3.8079	5.1960
%Improve	100.74	379.90	420.21	518.22	657.05

Table 5. Comparison of two MSPE methods without matching relationship, fixed at $m=50, n_{new}=5, \sigma_{\alpha}^2=1$
表 5. 不存在匹配关系两种方法 MSPE 的比较，固定 $m=50, n_{new}=5, \sigma_{\alpha}^2=1$

σ_{ε}^2	0.25	0.5	1	2	4
CMMP-SPSRs	0.0210	0.0915	0.2727	0.3861	0.6433
RP	0.2228	0.7791	1.3481	1.7807	2.6642
%Improve	962.21	751.63	394.35	361.15	314.15

将这 28 所学校的数据作为训练集数据的一个子集。因为并不知道训练集(总体)数据，所以假设总体数据为相应学校数据的 10 倍(重复)，故而，相应学校的总体均值与样本均值相同的。在 28 所学校中选择其中一所学校作为新观测数据，但并不知道这新观测数据来自具体哪一个学校，故而需要对其进行估计。接下来分别采用 CMMP-SPSRs 方法和 RP 方法对新观测的混合效应(均值)进行预测。最后为 28 所学校中的每所都重复以上操作。

假设训练集数据符合嵌套误差回归模型为

$$y_{ij}=\beta_0+\beta_1 x_{ij1}+\beta_2 x_{ij2}+\beta_3 x_{ij3}+\beta_4 x_{ij4}+\alpha_i+\varepsilon_{ij}, \tag{26}$$

$i=1,\cdots,m$， $j=1,\cdots,n_i$， $m=28$， n_i 为 2 至 28 不等。其中 y_{ij} 是干预后的 THKS 得分， x_{ij1} 为干预前的 THKS 得分，即不做任何处理的数据； x_{ij2} 为 CC 的数据； x_{ij3} 为 TV 的数据； x_{ij4} 为结合 CC 和 TV 的数据。 $\beta=(\beta_0,\beta_1,\beta_2,\beta_3,\beta_4)^{\mathrm{T}}$ 是固定效应且为未知参数，其中 $\alpha_i,\varepsilon_{ij}$ 和(1)式中的假设相同。参数 $\psi=(\beta,\sigma_{\alpha}^2,\sigma_{\varepsilon}^2)^{\mathrm{T}}$ 的最大似然估计可由(26)式得到，记为 $\hat{\psi}=(\hat{\beta},\hat{\sigma}_{\alpha}^2,\hat{\sigma}_{\varepsilon}^2)^{\mathrm{T}}$，同时通过计算也可获得参数 β 的最小二乘估计记为 β_{LS}。依次选择一所学校的模拟样本作为新观测数据，28 所学校则作为训练集数据。假设新观测数据也符合嵌套误差回归模型(NER)，则需要预测的混合效应为

$\theta=\beta_0+\beta_1 x_{ij1}+\beta_2 x_{ij2}+\beta_3 x_{ij3}+\beta_4 x_{ij4}+\alpha_1$。

接下来分别用 CMMP-SPSRs 方法和 RP 方法对新观测数据的混合效应进行预测，在预测准确性方面

比较了这两种方法。结果如表 6 所示，其中 CMMP-SPSRs 和 RP 列表示为绝对预测误差，即，

$$\left|\hat{\theta}-n_i^{-1}\sum_{j=i}^{n_i}y_{ij}\right|$$

$$\left|\hat{\theta}_{n,r}-n_i^{-1}\sum_{j=i}^{n_i}y_{ij}\right|$$

其中 $\hat{\theta}$ 表示 CMMP-SPSRs 方法的混合效应预测总体均值，$\hat{\theta}_{n,r}$ 表示 RP 方法下的混合效应预测总体均值，$\bar{y}_i=n_i^{-1}\sum_{j=i}^{n_i}y_{ij}$ 表示为总体均值。

Table 6. CMMP vs RP for TVSFP data
表 6. TVSFP 数据的 CMMP vs RP

学校	CMMP-SPSRs	RP	%Improve	学校	CMMP-SPSRs	RP	%Improve
1	0.1723	0.2999	74.0	15	0.3740	0.6292	68.2
2	0.0036	0.1008	267.2	16	0.2821	0.4242	50.3
3	0.0658	0.1384	110.1	17	0.0720	0.1124	56.0
4	0.0303	0.0848	179.3	18	0.0791	0.1689	135.0
5	0.0284	0.0840	195.8	19	0.2784	0.4360	56.5
6	0.0320	0.1015	217.3	20	0.0932	0.2457	163.6
7	0.1344	0.2898	115.7	21	0.2375	0.4506	89.7
8	0.1085	0.2094	92.9	22	0.1417	0.2241	58.1
9	0.1403	0.2412	71.8	23	0.2300	0.3533	53.6
10	0.0050	0.0156	209.3	24	0.0594	0.1414	138.0
11	0.0582	0.0979	68.0	25	0.0116	0.0498	329.8
12	0.2345	0.4328	84.5	26	0.2455	0.4033	64.2
13	0.0570	0.0596	4.6	27	0.0593	0.1343	126.4
14	0.0709	0.1271	79.3	28	0.0318	0.0815	156.1

通过表 6 中的数据可以看出，在这 28 所学校中，CMMP-SPSRs 方法的预测误差都比 RP 方法小，换句话讲，CMMP-SPSRs 方法的预测准确性比 RP 方法更优。改进的百分比由 4.6%到 329.8%不等。新观测数据和训练集数据间的匹配关系可能存在也有可能并不存在，本文通过匹配关系判别策略，利用训练集数据中尽量相似的群组中信息来提高预测的准确性。

5. 总结

本文主要讨论了嵌套误差模型中分类混合效应预测问题，主要对 CMMP 方法在识别准则的预测理论和方法进行改进，总结如下：对参数 I 的匹配准则方法的改进，在已有的分类混合模型 CMMP 方法，其匹配准则采用的方法为均方预测误差。为了优化 CMMP 方法，故提出了新的匹配准则方法：SPSRs。本文主要对 CMMP-SPSRs 方法和 RP 方法进行了比较，通过大量数值模拟可看出，当匹配关系存在或者不存在，CMMP-SPSRs 方法对于回归预测方法依然保持着很好的预测效果。

下一步的工作是在对待预测效应进行识别时，只讨论了两种可能，即新的样本属于某个小域或者不属于任何小域。结合模型结构误定和 SPSRs 准则，我们可以进一步考虑待预测样本按一定概率(或者模糊隶属度)落在小域中的相应预测方法。

参考文献

[1] Jiang, J.M., Rao, J.S., Fan, J. and Nguyen, T. (2018) Classified Mixed Model Prediction. *Journal of the American Statistical Association*, **113**, 269-279. https://doi.org/10.1080/01621459.2016.1246367

[2] 刘育孜, 曲维荣, 崔珍, 刘小惠, 徐文婧, 蒋继明. 基于分类混合效应模型预测方法和卫星遥感数据的农作物面积估算[J]. 数理统计与管理, 2021, 40(6): 1-15.

[3] 王婕雯. 基于卫星遥感数据和混合效应模型的小麦面积估算问题研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2021.

[4] Sun, H.M., Nguyen, T., Luan, Y.H. and Jiang, J.M. (2018) Classified Mixed Logistic Model Prediction. *Journal of Multivariate Analysis*, **168**, 63-74. https://doi.org/10.1016/j.jmva.2018.06.004

[5] Sun, H.M., Luan, Y.H. and Jiang, J.M. (2020) A New Classified Mixed Model Predictor. *Journal of Statistical Planning and Inference*, **207**, 45-54. https://doi.org/10.1016/j.jspi.2019.11.001

[6] Jiang, J.M. and Lahiri, P. (2006) Mixed Model Prediction and Small Area Estimation. *TEST*, **15**, Article No. 1. https://doi.org/10.1007/BF02595419

[7] Gneiting, T. and Raftery, A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American statistical Association*, **102**, 359-378. https://doi.org/10.1198/016214506000001437

[8] Merkle, E.C. and Steyvers, M. (2013) Choosing a Strictly Proper Scoring Rule. *Decision Analysis*, **10**, 292-304. https://doi.org/10.1287/deca.2013.0280

[9] Landes, J. (2015) Probabilism, Entropies and Strictly Proper Scoring Rules. *International Journal of Approximate Reasoning*, **63**, 1-21. https://doi.org/10.1016/j.ijar.2015.05.007

[10] Du, H.L. (2021) Beyond Strictly Proper Scoring Rules: The Importance of Being Local. *Weather and Forecasting*, **36**, 457-468. https://doi.org/10.1175/WAF-D-19-0205.1

[11] Hedeker, D., Gibbons, R.D. and Flay, B.R. (1994) Random Effects Regression Models for Clustered Data with an Example from Smoking Prevention Research. *Journal of Consulting and Clinical Psychology*, **62**, 757-765. https://doi.org/10.1037/0022-006X.62.4.757