STA 232B - Applied Statistics

Final Project

Analyses of the Iowa Crops Data

Authors:
Alphabetical ordering

Rui Hu
Heqiao Ruan
Tesi Xiao
Zitong Zhang
Yejiong Zhu

Instructor:
Dr. Jiming Jiang

April 9, 2019

Contents

1 Reading

1.1 Paper Summary

This paper [1] considers the problem of transforming satellite information into good estimates of crop areas at the individual pixel and segment levels. The authors analyzed the data about corn and soybeans from both farm-level survey in 1978 June and land observatory satellites (LANDSAT) during the 1978 growing season of 12 Iowa counties. A linear regression model is specified for the relationship between the reported hectares of corn and soybeans in the survey and the corresponding satellite determination of them. The correlation structure within the counties is given by a nested-error model, i.e. the mean hectares of the crop per segment in a county is defined as the conditional mean of reported hectares given the satellite data and the realized (random) county effect. Based on the proposed model, the authors defined and estimated the variance-component and obtained the generalized least-squares estimators. Predictions of mean hectares of corn and soybeans and their standard errors were obtained as well.

The components-of-variance model considered in the paper is

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_{ij}$$

where $i = 1, \cdots, 12$ denotes the county; $j = 1, \cdots, n_i$ denotes the segment, n_i is the number of segments in the ith county; y_{ij} is the number of hectares of crops in the jth segment of the ith county as reported in the June Survey; x_{1ij} and x_{2ij} are the number of pixels classified as corn and soybeans, respectively, in the j-th segment of the i-th county. The random error u_{ij} can be written as

$$u_{ij} = v_i + e_{ij}$$

where $v_i \sim N\left(0, \sigma_v^2\right)$ is the i-th county effect and $e_{ij} \sim N\left(0, \sigma_e^2\right)$ is the random effect associated with the j-th sample segment within the i-th county. Thus, the model expressed in matrix notation is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$.

Bisides the above model, the authors also considered several other circumstances: (i) other correlation structures considering geographical distance between segments; (ii) a multivariate framework that considers the correlation between reported areas of crop; (iii) the model including quadratic terms of the numbers of pixels of the crops. But none of these circumstances improve the precision of the estimation or have statistically significant results.

Firstly, the authors defined the sample mean and the population mean of the reported hectares of the two crops. The sample mean of the reported hectares of the two crops $\overline{y}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ can be expressed as $\overline{y}_{i\cdot} = \beta_0 + \beta_1 \overline{x}_{1i\cdot} + \beta_2 \overline{x}_{2i\cdot} + v_i + \overline{e}_{i\cdot}$, where $\overline{x}_{1i\cdot} \equiv n_i^{-1} \sum_{j=1}^{n_i} x_{1ij}$ and $\overline{x}_{2i\cdot} \equiv n_i^{-1} \sum_{j=1}^{n_i} x_{2ij}$ are the sample mean numbers of pixels of two crops, respectively, within county i, and $\overline{e}_{i\cdot} \equiv n_i^{-1} \sum_{j=1}^{n_i} e_{ij}$ is the sample mean of the within county effects in the i-th county. The population mean hectares of two crops in the i-th county (y_i) can be defined as the conditional mean given the realized county effect v_i and the values of the satellite data $y_i \equiv \beta_0 + \beta_1 \overline{x}_{1i(p)} + \beta_2 \overline{x}_{2i(p)} + v_i$, where $\overline{x}_{1i(p)} \equiv N_i^{-1} \sum_{j=1}^{N_i} x_{1ij}$ and $\overline{x}_{2i(p)} \equiv N_i^{-1} \sum_{j=1}^{N_i} x_{2ij}$ are the population mean numbers of pixels classified as two crops, respectively, in the i-th county, which are known. The focus of this paper is to predict the mean crop hectares per segment.

The best predictor of v_i is the conditional expectation of v_i given the sample mean $\overline{u}_{i\cdot}$. Suppose the variance σ_v^2 and σ_e^2 are known, then the generalized least-squares estimator of β is $\tilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$, where $\mathbf{V} = E(\mathbf{u}\mathbf{u}')$. Then a possible predictor of the i-th county effect v_i is

$$\tilde{v}_i = \tilde{u}_{i\cdot} g_i$$

where $\tilde{u}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} \tilde{u}_{ij}, \tilde{u}_{ij} = y_{ij} - \mathbf{x}_{ij}\tilde{\boldsymbol{\beta}}$, and $g_i = \frac{\sigma_v^2}{\sigma_v^2 + n_i^{-1}\sigma_e^2}$. Then the corresponding predictor $\tilde{y}_i$ is

$$\tilde{y}_i = \overline{\mathbf{x}}_{i(p)}\tilde{\boldsymbol{\beta}} + \tilde{v}_i$$

which is the best linear unbiased predictor (BLUP) of y_i and have the variance of the error as

$$\mathrm{E}\left\{\left(\tilde{y}_i - y_i\right)^2\right\} = \sigma_v^2\left(1 - g_i\right) + \mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})\mathbf{c}_i'$$

where $\mathbf{V}(\tilde{\boldsymbol{\beta}}) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$, $\mathbf{c}_i = \overline{\mathbf{x}}_{i(p)} - g_i\overline{\mathbf{x}}_i$ and $\overline{\mathbf{x}}_\mathbf{i} = n_i^{-1}\sum_{j=1}^{n_i}\mathbf{x}_{ij}$

Then, the authors considered a class of predictors of the county mean crop area y_i as

$$N_i^{-1}\left[\sum_{j=1}^{n_i} y_{ij} + \sum_{j=n_i+1}^{N_i}\left(\mathbf{x}_{ij}\tilde{\boldsymbol{\beta}} + \tilde{v}_i\right)\right]$$

When $\delta_i = g_i$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$, this is the BLUP above. When $\delta_i = 0$, this predictor is called the regression synthetic predictor. When $\delta_i = 1$, this predictor is called the survey regression predictor.

Since the variances σ_v^2 and σ_e^2 in the nested-error model are unknown, they are estimated by the residual mean square for the regression model. One choice for $\hat{\sigma_e^2}$ is

$$\hat{\sigma}_e^2 = \hat{\mathbf{e}}'\hat{\mathbf{e}}\left[\sum_{i=1}^{T}\left(n_i - 1\right) - 2\right]^{-1}$$

where $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ is the residual sum of squares for the regression of the y deviations, $y_{ij} - \overline{y}_{i\cdot}$, on the x deviations, $\mathbf{x}_{ij} - \overline{\mathbf{x}}_i$ for the counties with more than one samples. Then

$$d_e\frac{\hat{\sigma}_e^2}{\sigma_e^2} \sim \chi^2(d_e)$$

where $d_e \equiv \sum_{i=1}^{T}\left(n_i - 1\right) - 2$. In order to get an estimator of σ_v, consider the average of the ordinary least-squares residuals for county i:

$$\overset{\Delta}{u}_{i\cdot} = \overline{y}_{i\cdot} - \overline{\mathbf{x}}_i\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

and the weighted sum of squares of the average residuals for the counties

$$\hat{m}_{\cdot\cdot} \equiv \left(\sum_{i=1}^{T} n_i b_i\right)^{-1}\left(\sum_{i=1}^{T} n_i \overset{\Delta}{u}_{i\cdot}^2\right)$$

then the estimator of σ_v^2 is

$$\hat{\sigma}_v^2 = \max\left\{\hat{m}_\cdot - c\hat{\sigma}_e^2, 0\right\}$$

Thus, an accessable predictor for the mean crop area in county i is

$$\hat{y}_i \equiv \overline{\mathbf{x}}_{i(p)}\hat{\boldsymbol{\beta}} + \hat{u}_i\hat{g}_i$$

where

$$\hat{g}_i = \left(\hat{\sigma}_v^2 + n_i^{-1}\hat{\sigma}_e^2\right)^{-1}\hat{\sigma}_v^2$$

In addition, the authors also tried to modify the nested-error model of SUPER CARP, by which the variance components are first estimated and generalized least-squares estimators for the β parameters are obtained. The results showed that all coefficients for corn model is significant but only the coefficient of soybeans pixels is significantly different from 0. Moreover, the among-county variance is more significant for soybeans than for corn. The mean estimator retains desirable properties for non-normal assumption but the variance estimator can be seriously biased without normal assumption. The hypothesis testing of normality based on this data showed no reason to reject the normality assumption.

Denote the multiple regression estimators and the generalized least-squares estimators for β_1 and β_2 as $\hat{\boldsymbol{\beta}}_{\boldsymbol{W}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}$, respectively. Correspondingly, the estimated covariance matrix are $\hat{\sum}_{\boldsymbol{W}}$ and $\hat{\sum}_{\boldsymbol{G}}$, respectively. Then the approximate distribution of the statistic

$$F = 2^{-1}\left(\hat{\boldsymbol{\beta}}_W - \hat{\boldsymbol{\beta}}_G\right)'\left(\hat{\boldsymbol{\Sigma}}_W - \hat{\boldsymbol{\Sigma}}_G\right)^{-1}\left(\hat{\boldsymbol{\beta}}_W - \hat{\boldsymbol{\beta}}_G\right)$$

Under the null hypothesis that the slope parameters are the same within and among counties, $F \sim F(2, 22)$ and it can not be rejected based on this dataset.

Based on the predictor $\overline{y}_i$, the predictions for the mean crop hectares per segment along with the estimated standard errors can be computed, as well as the standard errors for the survey regression predictor and the sample mean of the survey data. From the result, we can see that as the number of sample segments increases, the differences between the predicted hectares of two crops and the corresponding sample means decrease. The standard errors of the sample mean are greater than those of the survey regression predictor, while the ratio of the standard error of the best predictor to that of the survey regression predictor increases as the sample segments increases. The improvement of the precision is modest while the sample segments increases from 3 to 5.

The survey regression predictor is unbiased and has relatively small variance. Based on this, the renewable predictor is defined by

$$\overset{*}{y}_i = \hat{y}_i + a_i\left[\sum_{j=1}^{T} W_j\left(\overline{y}_i - \overline{\mathbf{x}}_{i\cdot}\hat{\boldsymbol{\beta}}\right)(1 - \hat{g}_i)\right]$$

where $a_i = \left[\sum_{j=1}^{T} W_j^2 \hat{V}\left(\hat{y}_j\right)\right]^{-1} W_i^2 \hat{V}\left(\hat{y}_i\right)$. This adjustment produces a very small increase in the variance.

In all, the nested-error regression model in the paper provides a promising approach to predicting crop areas in small domains, and the USDA allows people to use supplementary information including estimates of variances form other areas and other years.

1.2 Derivation of formula (3.1) and (3.2)

Proof. Based on the assumptions above, we have

$$\begin{pmatrix} v_i \\ \overline{u_{i\cdot}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & \rho\sigma_v^2 \\ \rho\sigma_v^2 & \sigma_v^2 + n_i^{-1}\sigma_e^2 \end{pmatrix}\right)$$

Then the expectation of v_i given $\overline{u}_{i\cdot}$ is

$$E\left(v_i|\overline{u}_i\right) = 0 + \sigma_v^2 * (\sigma_v^2 + n_i^{-1}\sigma_e^2)^{-1}(\overline{u}_{i\cdot}) = \overline{u}_i g_i$$

where $g_i = m_i^{-1}\sigma_v^2$ and $m_i = \left(\sigma_v^2 + n_i^{-1}\sigma_e^2\right)$. Consider matrix $A = (1, -g_i)$, then $A(v_i, \overline{u}_{i\cdot}) \sim N(0, A\begin{pmatrix} \sigma_v^2 & \rho\sigma_v^2 \\ \rho\sigma_v^2 & \sigma_v^2 + n_i^{-1}\sigma_e^2 \end{pmatrix}A^T) = N(0, \sigma_v^2(1 - g_i) - g_i(\sigma_v^2 - g_i m_i)) = N(0, \sigma_v^2(1 - g_i))$. So the error variance in this best predictor is

$$\mathrm{E}\left\{\left(v_i - \overline{u}_i g_i\right)^2\right\} = \sigma_v^2\left(1 - g_i\right) = n_i^{-1}\sigma_e^2 - n_i^{-2}\sigma_e^2 m_i^{-1}\sigma_e^2$$

□

1.3 Derivation of formula (3.6)

Proof. The predictor $\tilde{y}_i$ is

$$\tilde{y}_i = \overline{\mathbf{x}}_{i(p)}\tilde{\boldsymbol{\beta}} + \tilde{v}_i$$

which is the best linear unbiased predictor (BLUP) of y_i and have the variance of the error as

$$\mathrm{E}\left\{(\tilde{y}_i - y_i)^2\right\}$$

Consider the model by matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} + \boldsymbol{\epsilon} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

where $\mathbf{V} = \mathrm{diag}(\mathbf{V_1}, \mathbf{V_2}, \cdots, \mathbf{V_{12}}), \mathbf{V_i} = \sigma_v^2 \mathbf{1_{n_i}} \mathbf{1_{n_i}^T} + \sigma_e^2 \mathbf{1_{n_i}}$

Since the generalized least-squares estimator of β is $\tilde{\boldsymbol{\beta}} = \left(\mathbf{X'V^{-1}X}\right)^{-1} \mathbf{X'V^{-1}Y}$, we have

$$\tilde{\boldsymbol{u}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$$

and

$$\tilde{v_i} = g_i(\overline{y}_{i.} - \overline{x}_{i.}\tilde{\boldsymbol{\beta}})$$

Then we have

$$\begin{aligned}
\mathrm{E}\left\{(\tilde{y}_i - y_i)^2\right\} &= \mathrm{Var}\left\{(\tilde{y}_i - y_i)^2\right\} \\
&= \mathrm{Var}(\overline{\mathbf{x}}_{i(p)}\tilde{\boldsymbol{\beta}} + \tilde{v}_i - \overline{\mathbf{x}}_{1(p)}\boldsymbol{\beta} - v_i) \\
&= \mathrm{Var}[\overline{\mathbf{x}}_{i(p)}\tilde{\boldsymbol{\beta}} + g_i(\overline{y}_{i.} - \overline{\mathbf{x}}_i) - v_i] \\
&= \mathrm{Var}[(\overline{\mathbf{x}}_{i(p)} - g_i\overline{x}_i)\tilde{\boldsymbol{\beta}} + g_i\overline{y}_{i.} - v_i]
\end{aligned}$$

Now let

$$Z_i = (0, \cdots, 0, \frac{1}{n_i}, \cdots, \frac{1}{n_i}, 0, \cdots, 0)^T$$

where only the elements for the ith county is $\frac{1}{n_i}$ and others are zero. Also let $\mathbf{c}_i = \overline{\mathbf{x}}_{i(p)} - g_i\overline{\mathbf{x}}_i$ and also notice that $\mathbf{V}(\tilde{\boldsymbol{\beta}}) = (X^TV^{-1}X)^{-1}$, so

$$\begin{aligned}
\mathrm{E}\left\{(\tilde{y}_i - y_i)^2\right\} &= \mathrm{Var}(\mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})X^TV^{-1}y + g_iZ_i^Ty - Z_i^Tv) \\
&= [\mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})X^TV^{-1} + g_iZ_i^T]V[V^{-1}XV(V\tilde{\boldsymbol{\beta}})\mathbf{c}_i^T + g_iZ_i] + \sigma_v^2 - 2\mathrm{Cov}((\mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})X^TV^{-1} + g_iZ_i^T)y, Z_i^Tv) \\
&= \mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})\mathbf{c}_i^T + 2g_iZ_i^TX\mathbf{V}(\tilde{\boldsymbol{\beta}})\mathbf{c}_i^T + g_i^2Z_i^TVZ_i + \sigma_v^2 - 2\sigma_v^2(\mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})X^TV^{-1} + g_iZ_i^T)n_iZ_i \\
&= \mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})\mathbf{c}_i^T + 2g_i\overline{x}_{i.}V(\tilde{\boldsymbol{\beta}})\mathbf{c}_i^T + g_i^2\frac{\sigma_v^2}{g_i} + \sigma_v^2 - 2\sigma_v^2(\mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})\frac{g_i}{\sigma_v^2}\overline{x}_{i.}^T + g_i) \\
&= \mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})c_i^T + \sigma_v^2 - \sigma_v^2g_i \\
&= \sigma_v^2\left(1 - g_i\right) + \mathbf{c}_i\mathbf{V}(\tilde{\boldsymbol{\beta}})\mathbf{c}_i'
\end{aligned}$$

where $\mathbf{V}(\tilde{\boldsymbol{\beta}}) = \left(\mathbf{X'V^{-1}X}\right)^{-1}, \mathbf{c}_i = \overline{\mathbf{x}}_{i(p)} - g_i\overline{\mathbf{x}}_i$ and $\overline{\mathbf{x}}_\mathbf{i} = n_i^{-1}\sum_{j=1}^{n_i}\mathbf{x}_{ij}$ □

1.4 Derivation of formula (3.10)

Proof. Considering the ordinary least-squares for county i, the average of the residuals is

$$\overset{\triangle}{u}_{i.} = \overline{y}_{i.} - \overline{x}_{i.}(\mathbf{X'X})^{-1}\mathbf{X'Y} = \frac{1}{n_i}\mathbf{1}'_{n_i}\mathbf{Y}_i - \overline{\mathbf{x}}_{i.}(\mathbf{X'X})^{-1}\mathbf{X'Y}$$

Since

$$\mathrm{E}(\mathbf{Y}_i) = \mathbf{0}, \mathrm{E}(\mathbf{Y}) = \mathbf{0}, Var(\mathbf{Y}) = \mathrm{diag}(\mathrm{Var}(\mathbf{Y}_i)) = \mathrm{diag}(\mathbf{J}_i\sigma_v^2 + \mathbf{I}_i\sigma_e^2)_{i=1,\cdots,T},$$

we have

$$\begin{aligned}
\mathrm{E}(\overset{\triangle}{u}_{i\cdot}^2) &= \mathrm{Var}(\overset{\triangle}{u}_{i\cdot}) + \{\mathrm{E}(\overset{\triangle}{u}_{i\cdot})\}^2 \\
&= \frac{1}{n_i^2}\mathbf{1}'_{n_i}\mathrm{Var}(\mathbf{Y}_i)\mathbf{1}_{n_i} + \overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}'_{i\cdot} \\
&\quad - \frac{2}{n_i}\mathbf{1}'_{n_i}\mathrm{Cov}(\mathbf{Y}_i, \mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}'_{i\cdot} \\
&= \frac{1}{n_i^2}\mathbf{1}'_{n_i}(\mathbf{J}_i\sigma_v^2 + \mathbf{I}_i\sigma_e^2)\mathbf{1}_{n_i} + \overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{diag}(\mathbf{J}_i\sigma_v^2 + \mathbf{I}_i\sigma_e^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}'_{i\cdot} \\
&\quad - \frac{2}{n_i}\mathbf{1}'_{n_i}(\mathbf{J}_i\sigma_v^2 + \mathbf{I}_i\sigma_e^2)\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}'_{i\cdot} \\
&= b_i\sigma_v^2 + d_i\sigma_e^2,
\end{aligned}$$

where

$$b_i = 1 - 2n_i\overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}_{i\cdot} + \overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\{\sum_{j=1}^{T} n_j^2\overline{\mathbf{x}}'_{j\cdot}\overline{\mathbf{x}}_{j\cdot}\}(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}'_{i\cdot},$$

and $d_i = n_i^{-1}\{1 - n_i\overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\overline{\mathbf{x}}'_{i\cdot}\}$ □

1.5 Proof of the statement after (3.12)

Under the assumptions of the model (2.1)-(2.2), the weighted sum of squares $\hat{m}_{\cdot\cdot}$ is independent of $\hat{\sigma}_e^2$.

Proof.

$$\hat{m}_{\cdot\cdot} \equiv \left(\sum_{i=1}^{T} n_i b_i\right)^{-1}\left(\sum_{i=1}^{T} n_i \overset{\Delta}{u}_i^2\right)$$

where

$$\overset{\Delta}{u}_{i\cdot} = \overline{y}_{i\cdot} - \overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

So

$$\begin{aligned}
\hat{m}_{\cdot\cdot} &= \text{constant} * \left(\sum_{i=1}^{T} n_i \overset{\Delta}{u}_i^2\right) \\
&= \text{constant} * \left(\sum_{i=1}^{T} n_i(\overline{y}_{i\cdot} - \overline{\mathbf{x}}_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})^2\right) \\
&= \text{constant} * (\bar{\mathbf{Y}} - \hat{\mathbf{Y}})^T(\bar{\mathbf{Y}} - \hat{\mathbf{Y}}) \\
&= \text{constant} * \mathbf{Y}^T(\mathbf{H} - \frac{1}{n}J_n)\mathbf{Y}
\end{aligned}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

$$\hat{\sigma}_e^2 = \hat{\mathbf{e}}^T\hat{\mathbf{e}}[\sum_{i=1}^{T}(n_i - 1) - 2]^{-1}$$

where $\hat{\mathbf{e}}^T\hat{\mathbf{e}} = (\hat{\mathbf{Y}}^* - \mathbf{Y}^*)^T(\hat{\mathbf{Y}}^* - \mathbf{Y}^*) = \mathbf{Y}^{*T}(\mathbf{H}^* - I_n)\mathbf{Y}^*$ with

$$\mathbf{Y}^* = \mathbf{Y} - \mathrm{diag}(\frac{1}{n_1}J_{n_1}, \frac{1}{n_2}J_{n_2}, \cdots, \frac{1}{n_{12}}J_{n_{12}})\mathbf{Y}$$

$$\mathbf{X}^* = \mathbf{X} - \mathrm{diag}(\frac{1}{n_1}J_{n_1}, \frac{1}{n_2}J_{n_2}, \cdots, \frac{1}{n_{12}}J_{n_{12}})\mathbf{X}$$

$$\mathbf{H}^* = \mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}$$

Denote $I_n - \text{diag}(\frac{1}{n_1}J_{n_1}, \frac{1}{n_2}J_{n_2}, \cdots, \frac{1}{n_{12}}J_{n_{12}}) = K_n$, then we have

$$\hat{\sigma}_e^2 = \text{constant} * \mathbf{Y}^{*T}(\mathbf{H}^* - I_n)\mathbf{Y}^* = \text{constant} * \mathbf{Y}^T K_n^T(\mathbf{H} - I_n)K_n\mathbf{Y}$$

Since $(\mathbf{H} - \frac{1}{n}J_n)K_n^T(\mathbf{H} - I_n)K_n = 0$, by the Cochran's theorem, we have that $\mathbf{Y}^T(\mathbf{H} - \frac{1}{n}J_n)\mathbf{Y}$ and $\mathbf{Y}^T K_n^T(\mathbf{H} - I_n)K_n\mathbf{Y}$ are independent, i.e. the weighted sum of squares $\hat{m}_{..}$ is independent of $\hat{\sigma}_e^2$. □

1.6 Derivation of (A.1) (A.2)

Proof. The predictor $\hat{y}_i$ is

$$\hat{y}_i = \overline{\mathbf{x}}_{(p)}\hat{\boldsymbol{\beta}} + \left(\overline{y}_{i.} - \overline{\mathbf{x}}_i\hat{\boldsymbol{\beta}}\right)\hat{g}_i$$

where

$$\hat{g}_i = 1 - \hat{h}_i$$

$$\hat{h}_i = \left[\hat{m}_i + \hat{k}_i + \left(n_i^{-1} - c\right)^2 \hat{w}_i\right]^{-1}\left[n_i^{-1}\hat{\sigma}_e^2 + \left(n_i^{-1} - c\right)n_i^{-1}\hat{w}_i\right]$$

$$\hat{m}_i = \hat{m}_{..} + \left(n_i^{-1} - c\right)\hat{\sigma}_e^2$$

$$\hat{w}_i = 2d_e^{-1}\hat{m}_i^{-1}\hat{\sigma}_e^4$$

$$\hat{k}_i = 2\hat{\sigma}_e^2\left(\ddot{\sigma}_{ff} + n_i^{-1}\right)^{-1}\left[\sum_{j=1}^{T} n_j b_j\right]^{-2}\left[\sum_{j=1}^{T} n_j^2 b_j\left(\ddot{\sigma}_{ff} + n_j^{-1}\right)^2\right]$$

$$\ddot{\sigma}_{ff} = \max\left[0, (T-5)^{-1}(T-3)\hat{\sigma}_e^{-2}\hat{m}_{.} - c\right]$$

Then we can compute the variance of the error of the predictor above as

$$\begin{aligned}
V(\hat{y}_i - y_i) &= \text{Var}(\overline{\mathbf{x}}_{(p)}\hat{\boldsymbol{\beta}} + \left(\overline{y}_{i.} - \overline{\mathbf{x}}_i\hat{\boldsymbol{\beta}}\right)\hat{g}_i - \overline{x}_i\beta - v_i) \\
&= \text{Var}[(\overline{\mathbf{x}}_{i(p)} - \hat{g}_i\overline{x}_{i.})\hat{\boldsymbol{\beta}} + \hat{g}_i\overline{y}_{i.} - v_i] \\
&= \text{Var}(\hat{c}_i\hat{\beta} + \hat{g}_i\overline{y}_{i.} - v_i) \\
&= \text{Var}([\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1} + \hat{g}_i Z_i^T]Y - Z_i^T v) \\
&= [\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1} + \hat{g}_i Z_i^T]V[\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T + \hat{g}_i Z_i] + \sigma_v^2 \\
&- \text{Cov}(\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1} + \hat{g}_i Z_i^T, Z_i^T V) \\
&= \hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1}V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T + 2\hat{g}_i Z_i^T V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}c_i^T \\
&+ {\hat{g}_i}^2 Z_i^T V Z_I + \sigma_v^2 - 2\text{Cov}(\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}\frac{1}{n_i\hat{\sigma}_v^2 + \hat{\sigma_e^2}}\overline{x}_{i.}^T + \hat{g}_i Z_i^T)n_{Z\,i} \\
&= \hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1}V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T + 2\hat{g}_i Z_i^T V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}c_i^T \\
&+ {\hat{g}_i}^2\frac{\sigma_v^2}{g_i} + \sigma_v^2 - 2\sigma_v^2(\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}\frac{\overline{x}_{i.}^T}{n_i\hat{\sigma}_v^2 + \hat{\sigma}_e^2} + \hat{g}_i) \\
&= \hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1}V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T \\
&+ 2(1 - \left[\hat{m}_i + \hat{k}_i + \left(n_i^{-1} - c\right)^2\hat{w}_i\right]^{-1}\left[n_i^{-1}\hat{\sigma}_e^2 + \left(n_i^{-1} - c\right)n_i^{-1}\hat{w}_i\right])Z_i^T V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}c_i^T \\
&+ (1 - \left[\hat{m}_i + \hat{k}_i + \left(n_i^{-1} - c\right)^2\hat{w}_i\right]^{-1}\left[n_i^{-1}\hat{\sigma}_e^2 + \left(n_i^{-1} - c\right)n_i^{-1}\hat{w}_i\right])^2\frac{\sigma_v^2}{g_i} + \sigma_v^2 \\
&- 2\sigma_v^2(\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}\frac{\overline{x}_{i.}^T}{n_i\hat{\sigma}_v^2 + \hat{\sigma}_e^2} + (1 - \left[\hat{m}_i + \hat{k}_i + \left(n_i^{-1} - c\right)^2\hat{w}_i\right]^{-1}\left[n_i^{-1}\hat{\sigma}_e^2 + \left(n_i^{-1} - c\right)n_i^{-1}\hat{w}_i\right]))
\end{aligned}$$

For the unknown parameters, we can use their estimator. For example, consider the first term

$$\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1}V\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T$$

we can use $\hat{V}$ as the estimator of V, then the estimator of this term will be

$$\begin{aligned}
&\hat{c}_i(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1}\hat{V}\hat{V}^{-1}X(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T \\
&= \hat{c}_i(X^T\hat{V}^{-1}X)^{-1}\hat{c}_i^T \\
&= \hat{c}_i\hat{V}(\hat{\beta})\hat{c}_i^T
\end{aligned}$$

which is the leading term in the estimated variance. For other part, we may use their estimator to get corresponding formula, respectively, to get the result:

$$\hat{V}\left\{\hat{y}_i - y_i\right\} = n_i^{-1}\hat{\sigma}_e^2 - \hat{\phi}_i + \hat{\mathbf{c}}_i\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\hat{\mathbf{c}}_i' + \hat{h}_i^2\hat{k}_i + d_e^{-1}\hat{r}_i^2\hat{\phi}_i + d_e^{-1}\hat{r}_i^2\hat{h}_i\hat{\sigma}_e^2$$

where

$$\hat{\mathbf{c}}_i = \overline{\mathbf{x}}_{(p)} - \hat{g}_i\overline{\mathbf{x}}_i$$

$$\hat{\phi}_i = (d_e+1)^{-1}\, d_e\ddot{\phi}_i - d_e^{-1}n_i^{-1}\hat{\sigma}_e^-\hat{h}_i$$

$$\ddot{\phi}_i = n_i^{-2}\left[\hat{\sigma}_e^2 + \left(n_i^{-1}-c\right)\hat{w}_i\right]^2\left[\hat{m}_i + \hat{k}_i + \left(n_i^{-1}-c\right)^2\hat{w}_i\right]^{-1}$$

$$\hat{r}_i = 1 - (1-n_ic)\,\hat{h}_i$$

□

2 Part I: Model Selection

2.1 AIC and BIC Criteria

The nested error regression model proposed here is

$$Y_{ij} = x_{ij}^{'}\beta + v_i + e_{ij}, i = 1, 2, .., 12, j = 1, 2, .., n_i$$

where y_{ij} is the reported hectares of corn(soybean) and the fixed effect is given by

$$x_{ij}^{'}\beta = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$$

The county-specific random effect is $v_i \sim N(0, \sigma_v^2)$ and the sample-specific random error is $e_{ij} \sim N(0, \sigma_e^2)$ and they are independent with both variance unknown.

In the original paper, the author discussed about the possibility to incorporate quadratic and interaction effects. So for the fixed effects we have the following candidate variable set in which we select the optimal model based on AIC/BIC:

$$\Theta = \{x_{1ij}, x_{2ij}, x_{1ij}^2, x_{2ij}^2, x_{1ij}x_{2ij}\}$$

As for preprocessing the data, as the original paper argues that the 33rd observation may be problematic, we delete this sample and fit our model on the rest of the data.

For implementation, we traverse all the possible cases including the all possible subsets in Θ for corn hectares and soybean hectares separately. We apply **AIC** and **BIC** criteria to perform the model comparison and select the model with the smallest AIC/BIC.

$$\text{AIC} = -2\log\text{Lik} + 2k \tag{1}$$

$$\text{BIC} = -2\log\text{Lik} + k\log(n) \tag{2}$$

where k is the number of parameters in the model and n is the number of samples.

Here we use the n_{total} as the "effective" sample size. However, it may not be the optimal choice for this data as we did not consider the effect size innate to the random effects.

In 2014, Maud Delattre proposed an alternative BIC for mixed effect model [2] with a more comprehensive consideration of the effective sample size for both random effects and fixed effects.

$$\text{BIC}_h = -2\log\text{Lik} + \dim(\theta_R)\log(N) + \dim(\theta_F)\log(n_{total}) \tag{3}$$

where $\dim(\theta_R)$ and $\dim(\theta_F)$ denotes the number of parameters of random effects and fixed effects and N denotes the number of subjects (corresponding to the number of county in this application).

We may argue that as the known correlation structure, directly estimating the effective sample size maybe very complex and data-specific. The alternative BIC indeed reweight the sample size of fixed effects and random effects and avoid estimating the 'effective sample size' directly.

One more important issue we would like to point out is that for fitting the models, comparing mixed model with different fixed effects by a likelihood-based criterion, we would like to use ML procedure instead of REML procedure as REML won't give us feasible comparison between the 'likelihood' of different models.

In the following sections, we will use (1)(3) to calculate AIC and BIC and select our corresponding models.

2.2 Results

The results in Table 1 show that for corn hectares both AIC and BIC select the same model with CornPix and CornPix:SoyBean as fixed effect, and for soybeans hectares both AIC and BIC select the same model with only SoyBeans as fixed effect. For further details, see Appendix A.1 Table 2.

Criterion	Selected Model
AIC	CornHec $\sim$ CornPix + CornPix:SoyBeanPix + (1|County)
BIC	CornHec $\sim$ CornPix + CornPix:SoyBeanPix + (1|County)
AIC	SoyBeanHec $\sim$ SoyBeansPix + (1|County)
BIC	SoyBeanHec $\sim$ SoyBeansPix + (1|County)

Table 1: Model Selection

2.3 Comments

- In real application BIC generally tends to select a smaller model (when the sample size n is relatively large) as it impose more penalty on the model complexity than AIC does. **However** in this data application we can see that they select the same model for both response variables.

- Note that in fitting the linear mixed model with higher order terms especially with quadratic terms, some of the predictor variables are on very different scales which may introduce some kinds of numerical instabilities.

- In general, BIC is argued to be appropriate for selecting the "true model" from the set of candidate models, whereas AIC is not appropriate. However, AIC is asymptotically optimal for selecting the model with the least mean squared error, under the assumption that the "true model" is not in the candidate set as AIC can be served as an estimator of the KL divergence between the selected model and the 'true model'.

- Since we hope to predict the reported hectares based on the data, AIC may of better interests in prediction.

- Alternative BIC considers the different penalty for the sample size and the group size, which seems more reasonable in model selection than the plain BIC criteria. But the selected models under BIC and alternative BIC are the same for this dataset, since the sample size is relatively small.

3 Part II: Application of Sumca

3.1 Sumca Method

Recently, a new method was proposed for estimating the MSPE of a complex predictor, known as **Sumca**: a Simple, Unified, Monte-Carlo Assisted Approach to Second-order Unbiased MSPE Estimation (Jiang, Torabi, 2018). [3, 4]

Here, we apply Sumca method for the problem under two different situations. The discrepancy between these situations depends on our unknown parameters ψ.

In the first situation, ψ only includes fixed effects and variance components i.e.

$$\psi = (\beta, \sigma_v^2, \sigma_e^2) \tag{4}$$

We refer to this method as "Plain" Sumca.

In the second situation, ψ includes model (M), fixed effects and variance components i.e.

$$\psi = (M, \beta, \sigma_v^2, \sigma_e^2) \tag{5}$$

We refer to this method as "M-parameterized" Sumca.

3.1.1 Plain Sumca

The post model selection predictor $\hat{\theta}_i$ be a predictor of θ_i, where $\theta_i = \beta_0 + \beta_1 \bar{X}_{1i(p)} + \beta_2 \bar{X}_{2i(p)} + v_i$ and $\hat{\theta}_i$ is a function of Y_i. Then, the MSPE of the post model selection prediction is

$$\text{MSPE}(\hat{\theta}_i) = E(\hat{\theta}_i - \theta)^2 = E[E(\hat{\theta}_i - \theta_i)^2 | Y_i] \tag{6}$$

Let

$$\begin{aligned} a(Y_i, \psi) = \text{E}[(\hat{\theta}_i - \theta_i)^2 | Y_i] &= (\hat{\theta}_i - \text{E}(\theta_i | Y_i))^2 + Var(\theta_i | Y_i) \\ &= (\hat{\theta}_i - a_1(Y_i, \psi))^2 + a_2(Y_i, \psi) \end{aligned} \tag{7}$$

where $a_1(Y_i, \phi) = \text{E}(\theta_i | Y_i)$ and $a_2(Y_i, \phi) = Var(\theta_i | Y_i)$

Under the first situation where $\psi = (\beta, \sigma_e^2, \sigma_v^2)$ and $a_1(Y_i, \hat{\psi}) \neq 0$, thus we need to derive $E(\theta_i | Y_i)$ and $\hat{\theta}_i$, and $E(\theta_i | Y_i)$ should be derived under the full model and $\hat{\theta}_i$ should be derived under the selected model.

$$\begin{aligned} a_1(Y_i, \psi) &= \text{E}(\theta_i | Y_i) \\ &= \beta_0 + \beta_1 \bar{X}_{1i(p)} + \beta_2 \bar{X}_{2i(p)} + \sigma_v^2 1_{n_i}^T (\sigma_e^2 I + \sigma_v^2 1_{n_i} 1_{n_i}^T)^{-1} (Y_i - x_i \beta) \\ &= \beta_0 + \beta_1 \bar{X}_{1i(p)} + \beta_2 \bar{X}_{2i(p)} + \frac{\sigma_v^2 n_i (\bar{Y}_i - \bar{x}_i \beta)}{\sigma_v^2 n_i + \sigma_e^2} \end{aligned} \tag{8}$$

$$a_2(Y_i, \psi) = \text{Var}(\theta_i | Y_i) = \frac{\sigma_v^2 \sigma_e^2}{\sigma_v^2 n_i + \sigma_e^2} \tag{9}$$

$$\hat{\theta}_i = [\beta_0 + \beta_1 \bar{X}_{1i(p)} + \beta_2 \bar{X}_{2i(p)} + \frac{\sigma_v^2 n_i (\bar{Y}_i - \bar{x}_i^* \beta)}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi = \hat{\psi}^S} \tag{10}$$

where $x_i \beta = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{1ij}^2 + \beta_4 x_{2ij}^2 + \beta_5 x_{1ij} x_{2ij}$, $\quad x_i^* \beta$ is the linear predictor of the selected model and $\hat{\psi}^S$ is obtained under the selected model.

Let $a(Y_i, \hat{\psi}^F)$ be an estimate of $a(Y_i, \psi)$ where $\hat{\psi}^F$ is estimated parameters under the full model. Thus, the **leading term** $a(Y_i, \hat{\psi}^F)$ is

$$a(Y_i, \hat{\psi}^F) = (\hat{\theta}_i - a_1(Y_i, \hat{\psi}^F))^2 + a_2(Y_i, \hat{\psi}^F) \tag{11}$$

where $\hat{\theta}_i$ in (10) is the predictor under the selected model.

$$a_1(Y_i, \hat{\psi}^F) = [\beta_0 + \beta_1 \bar{X}_{1i(p)} + \beta_2 \bar{X}_{2i(p)} + \frac{\sigma_v^2 n_i (\bar{Y}_i - \bar{x}_i \beta)}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^F} \tag{12}$$

$$a_2(Y_i, \hat{\psi}^F) = [\frac{\sigma_v^2 \sigma_e^2}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^F} \tag{13}$$

According to Sumca, therefore, a second-order unbiased estimator of MSPE($\hat{\theta}_i$) is

$$\hat{\text{MSPE}}(\hat{\theta}_i) = a(Y_i, \hat{\psi}^F) + d(\hat{\psi}^F) \tag{14}$$

where $d(\hat{\psi}^F) = \text{E}[a(Y_i, \psi) - a(Y_i, \hat{\psi})]_{\psi=\hat{\psi}^F}$ can be estimated by Monte-Carlo approach.

$$d(\hat{\psi}^F) = \frac{1}{K} \sum_{k=1}^{K} (a(Y_{i[k]}, \hat{\psi}^F) - a(Y_{i[k]}, \hat{\psi}^F_{[k]})) \tag{15}$$

where

$$a(Y_{i[k]}, \hat{\psi}^F) = (\hat{\theta}_i^{[k]} - a_1(Y_{i[k]}, \hat{\psi}^F))^2 + a_2(Y_{i[k]}, \hat{\psi}^F) \tag{16}$$

$$a(Y_{i[k]}, \hat{\psi}^F_{[k]}) = (\hat{\theta}_i^{[k]} - a_1(Y_{i[k]}, \hat{\psi}^F_{[k]}))^2 + a_2(Y_{i[k]}, \hat{\psi}^F_{[k]}) \tag{17}$$

where $Y_{[k]}$ is simulated under the full model as $\psi = \hat{\psi}^F$, $Y_{i[k]}$ is the corresponding response of the i-th county, and $\hat{\psi}^F_{[k]}$ is estimated based on $Y_{[k]}$ under the full model.

Algorithm 1: Plain Sumca

Input: Response Y, Predictor X, Population mean $\overline{X}_{(p)}$, Index i;
Select the optimal submodel according to AIC or BIC and obtain $\hat{\theta}_i$ by (10);
Fit the full model to obtain $\hat{\psi}^F$ and $a_1(Y_i, \hat{\psi}^F), a_2(Y_i, \hat{\psi}^F)$ by (12)(13);
Obtain the leading term $a(Y_i, \hat{\psi}^F)$ by (11);
for *k=1:K* **do**
  Simulate $Y_{[k]}$ based on the full model;
  Fit a new full model with response $Y_{[k]}$ and obtain the corresponding $\hat{\psi}^F_{[k]}$;
  Obtain $d_{[k]} = a(Y_{i[k]}, \hat{\psi}^F) - a(Y_{i[k]}, \hat{\psi}^F_{[k]})$
end
Set $\hat{\text{MSPE}} = a(Y_i, \hat{\psi}^F) + \sum_{k=1}^{K} d_{[k]}/K$;
if $\hat{\text{MSPE}} < 0$ **then**
  return $a(Y_i, \hat{\psi}^F)$
else
  return $\hat{\text{MSPE}}$
end

3.1.2 M-parameterized Sumca

Under the second situation in which we view the model as a parameter in ψ, we have the poste model selection predictor $\hat{\theta}_i = a_1(Y_i, \hat{\psi}^S)$. Thus, the **leading term** $a(Y_i, \hat{\psi}^S)$ is

$$\begin{aligned} a(Y_i, \hat{\psi}^S) = a_2(Y_i, \hat{\psi}^S) &= \text{Var}(\theta_i | Y_i)_{\psi=\hat{\psi}^S} \\ &= [\frac{\sigma_v^2 \sigma_e^2}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^S} \end{aligned} \tag{18}$$

where $\text{Var}(\theta_i|Y_i)_{\psi=\hat{\psi}^S}$ is the conditional variance under the selected model.

Based on the M-parameterized Sumca, a second-order unbiased estimator of $\text{MSPE}(\hat{\theta}_i)$ is

$$\hat{\text{MSPE}}(\hat{\theta}_i) = a(Y_i, \hat{\psi}^S) + d(\hat{\psi}^S) = a_2(Y_i, \hat{\psi}^S) + d(\hat{\psi}^S) \tag{19}$$

where $d(\hat{\psi}^S) = E[a(Y_i, \psi) - a(Y_i, \hat{\psi})]_{\psi=\hat{\psi}^S}$ and it can be estimated by Monte-Carlo approach.

$$d(\hat{\psi}^S) = \frac{1}{K}\sum_{k=1}^{K}[a(Y_{i[k]}, \hat{\psi}^S) - a(Y_{i[k]}, \hat{\psi}^S_{[k]})] \tag{20}$$

where

$$a(Y_{i[k]}, \hat{\psi}^S) = (\hat{\theta}_i^{[k]} - a_1(Y_{i[k]}, \hat{\psi}^S))^2 + a_2(Y_{i[k]}, \hat{\psi}^S) \tag{21}$$

$$a(Y_{i[k]}, \hat{\psi}^S_{[k]}) = a_2(Y_{i[k]}, \hat{\psi}^S_{[k]}) = \text{Var}(\theta_i|Y_{i[k]})_{\psi=\hat{\psi}^S_{[k]}} \tag{22}$$

Here $Y_{[k]}$ is simulated under the selected model as $\psi = \hat{\psi}^S$, $Y_{i[k]}$ is the corresponding response of the i-th county, and $\hat{\psi}^S_{[k]}$ is estimated based on $Y_{[k]}$ after model selection.

Algorithm 2: M-parameterized Sumca

Input: Response Y, Predictor X, Population mean $\overline{X}_{(p)}$, Index i;
Select the optimal submodel according to AIC or BIC and obtain $\hat{\psi}^S$;
Obtain the leading term $a(Y_i, \hat{\psi}^S)$ by (18);
for *k=1:K* **do**
  Simulate $Y_{[k]}$ based on the $\hat{\psi}^S$;
  Select the optimal submodel with response $Y_{[k]}$ and obtain the corresponding $\hat{\psi}^S_{[k]}$;
  Obtain $d_{[k]} = a(Y_{i[k]}, \hat{\psi}^S) - a(Y_{i[k]}, \hat{\psi}^S_{[k]})$
end
Set $\hat{\text{MSPE}} = a(Y_i, \hat{\psi}^S) + \sum_{k=1}^{K} d_{[k]}/K$;
if $\hat{\text{MSPE}} < 0$ **then**
  return $a(Y_i, \hat{\psi}^S)$
else
  return $\hat{\text{MSPE}}$
end

3.1.3 The Leading Term

The leading term for Sumca method above is guaranteed positive, since it is the summation of the squared conditional bias and the conditional variance. This is a desirable property for an MSPE estimator. If there are any negative estimates of the MSPEs by Sumca, an alternative estimator is the leading term of the Sumca estimator.

- **Plain Sumca**: The leading term is

$$a(Y_i, \hat{\psi}^F) = (\hat{\theta}_i - a_1(Y_i, \hat{\psi}^F))^2 + a_2(Y_i, \hat{\psi}^F)$$

  where $\hat{\theta}_i$ is the predictor under the selected model.

$$\hat{\theta}_i = [\beta_0 + \beta_1\bar{X}_{1i(p)} + \beta_2\bar{X}_{2i(p)} + \frac{\sigma_v^2 n_i(\bar{Y}_i - \bar{x}_i^*\beta)}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^S}$$

$$a_1(Y_i, \hat{\psi}^F) = [\beta_0 + \beta_1\bar{X}_{1i(p)} + \beta_2\bar{X}_{2i(p)} + \frac{\sigma_v^2 n_i(\bar{Y}_i - \bar{x}_i\beta)}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^F}$$

$$a_2(Y_i, \hat{\psi}^F) = [\frac{\sigma_v^2\sigma_e^2}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^F}$$

where $x_i\beta = \beta_0+\beta_1 x_{1ij}+\beta_2 x_{2ij}+\beta_3 x_{1ij}^2+\beta_4 x_{2ij}^2+\beta_5 x_{1ij}x_{2ij}$, $x_i^*\beta$ is the linear predictor of the selected model and $\hat{\psi}^S$ is obtained under the selected model.

- M**-parameterized Sumca**: The leading term is

$$a(Y_i, \hat{\psi}^S) = a_2(Y_i, \hat{\psi}^S) = \text{Var}(\theta_i|Y_i)_{\psi=\hat{\psi}^S} = [\frac{\sigma_v^2\sigma_e^2}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^S}$$

  where $\text{Var}(\theta_i|Y_i)_{\psi=\hat{\psi}^S}$ is the conditional variance under the selected model.

3.1.4 The choice of K

As for choosing K, the number of monte-carlo simulations we perform, there are two main issues we need to consider:

- **Computational cost**: we need to perform a nested model selection procedure for each simulated sample. Therefore, K should not be too large.
- **Robustness**: Simulation definitely introduces much more uncertainty so we need to repeat serveral times to improve robustness.

Based on the lecture notes, we choose the K that is proportional to the number of random effects. In this project, after several experiments, we finally choose $K = 100$ empirically. We believe that more work is required for investigating the asymptotic behavior of the MSPE.

3.2 Results

In this section, we perform the "Plain" and "M-parameterized" Sumca method in two ways:

- **Fixed**: The post model selection predictor $\hat{\theta}_i$ has the same form with (10) in which the fixed effect only considers two population mean of CornPix and SoyBeanPix, no matter which predictors the selected model includes.

$$\hat{\theta}_i = [\beta_0 + \beta_1\bar{X}_{1i(p)} + \beta_2\bar{X}_{2i(p)} + \frac{\sigma_v^2 n_i(\bar{Y}_i - \bar{x}_i^*\beta)}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^S}$$

  This follows the instructions of the Professor Jiang on this project. Note that if the selected model does not contain the linear terms, the corresponding coefficient $\hat{\beta}_1, \hat{\beta}_2$ in $\hat{\theta}_i$ degenerate to 0.

- **Nonfixed**: The post model selection predictor $\hat{\theta}_i$ has an alternative form in which the fixed effect considers all the possible predictors. Therefore, the estimates of the fixed effect and the random effects are consistent under the same selected model.

$$\hat{\theta}_i = [\beta_0 + \beta_1\bar{X}_{1i(p)} + \beta_2\bar{X}_{2i(p)} + \beta_3\overline{X_{1i(p)}^2} + \beta_4\overline{X_{2i(p)}^2} + \beta_5\overline{X_{1i}X_{2i}}_{(p)} + \frac{\sigma_v^2 n_i(\bar{Y}_i - \bar{x}_i^*\beta)}{\sigma_v^2 n_i + \sigma_e^2}]_{\psi=\hat{\psi}^S}$$

  In practice, since we do not have the information about the population mean of squared terms and interaction term, we directly replace them by the squared population means and the product of the population mean.

We compute the Sumca estimates in four ways (PlainFixed, PlainNonfixed, M-parFixed, M-parNonfixed) for all 12 EBLUPs of the population mean hectares of corn and soybeans, and under AIC and BIC model selections separately. The graph below presents the EBLUPs as well as the corresponding 2 times the square roots of the MSPE estimates, used as the margins of errors. For further details, see Appendix A.2.

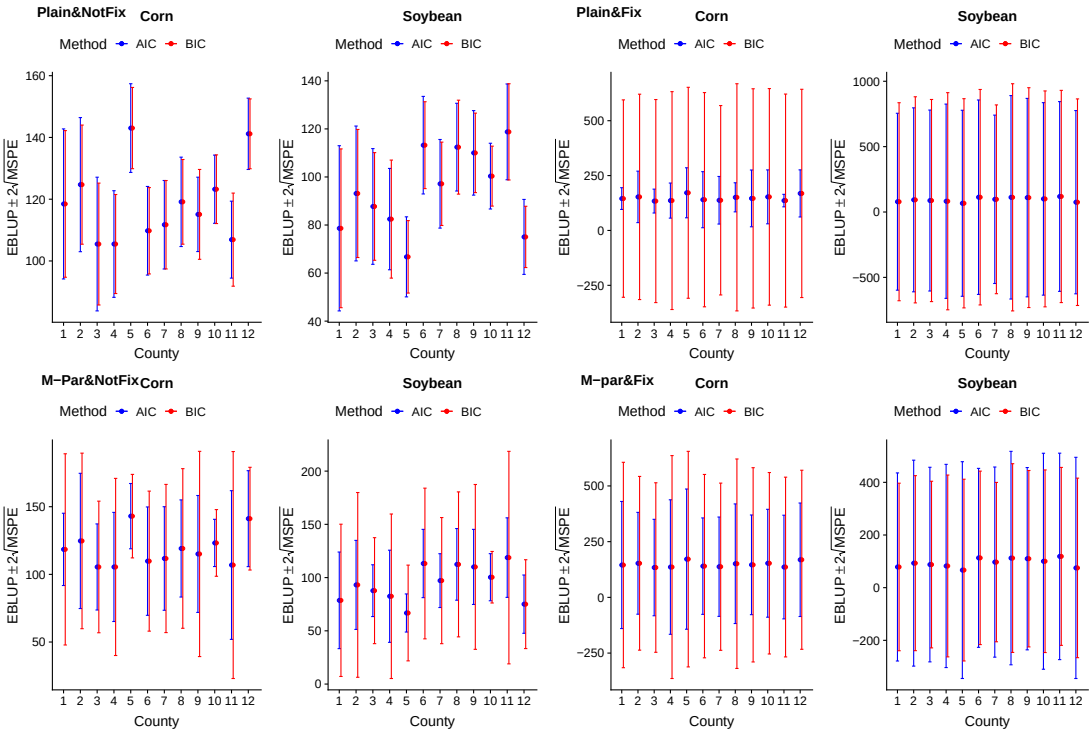

Figure 1: The EBLUPs and the corresponding 2 times the square roots of the MSPE estimates for 12 counties

As we can see from Figure 1,

- The nonfixed sumca method performs better than the fixed one for most cases, since the nonfixed sumca has larger variation;
- The MSPE estimates for counties with larger size ($n_{10} = n_{11} = n_{12} = 5$) is usually smaller than the ones for counties with smaller size ($n_1 = n_2 = n_3 = 1$), which corresponds to the intuition;
- When comparing the results based on different model selection criteria, AIC tends to find smaller MSPEs than BIC does in most cases, especially in M-parameterized Sumca. It may be caused by the variation of M in ψ when implementing the Monte-Carlo method.
- Comparing "Plain" Sumca and "M-parameterized" Sumca, the "M-parameterized" Sumca seems to have better performance than the plain one. One reason might be that the "Plain" Sumca assumes the full model is true which is not always resonable.

3.3 Comments

Based on the results above, we discussed about several intriguing points of the proposed Sumca method in this application.

- **Data Preprocessing**:

  For the two identical data points (32nd, 33rd), we just simple delete the second one (33rd) according to the paper, which may be harmful to the original data structure. We may consider better methods to deal with this issue, such as using the mean to replace it.

- **Inconsistency in $\hat{\theta}_i$**:

The above "fixed" Sumca method based on Professor Jiang's instructions only considers the population main effects in the post model selection predictor $\hat{\theta}_i$. We argue that $\hat{\theta}_i$ is definitely inconsistent. Here are some possible reasons for inconsistency:

1. We select the model based on all the possible variables and use the corresponding estimated coefficients. Therefore, the coefficients might be quite unreasonable after deleting the higher order terms because of the multicollinearity;
2. Population quadratic terms and interaction terms which removed by ourselves may also introduce some uncertainties;
3. Misspecification under the model with removed quadratic effects will definitely introduce some inflation for the MSPE estimation.

- **Difference between Mean of Square and Square of Mean**

  In the dataset we can't observe the mean of the sample-specific square. In practice, we directly replace them by the squared population means and the product of the population mean, which may introduce potential model misspecification.

- **Robustness**:

  We observe that the size of monte-carlo samples K does not make any difference to the robustness of the MSPE estimation. One possible explanation may be that the size of this data is quite small, especially for some counties, the size is 1. More simulations cannot reduce the systematic uncertainty for this data.

- **AIC vs. BIC**

  We observe that for part II, results based on AIC is significantly better than BIC, it may come from the nature of BIC. In the nested model selection, BIC would tend to select a too simple model which may lose some efficiency in capturing the variation of the dataset. Moreover,minimizing AIC is equivalent to minimizing the KL divergence so in the application domain choosing the small AIC means choosing a model approaching the 'true' one. So we may argue that AIC will be better than BIC for model selection aimed to do prediction.

References

[1] George E Battese, Rachel M Harter, and Wayne A Fuller. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401):28–36, 1988.

[2] Maud Delattre, Marc Lavielle, Marie-Anne Poursat, et al. A note on bic in mixed-effects models. *Electronic journal of statistics*, 8(1):456–475, 2014.

[3] Jiming Jiang. Sta 232b lecture notes.

[4] Jiming Jiang. *Linear and generalized linear mixed models and their applications*. Springer Science & Business Media, 2007.

A Appendix

A.1 Model Selection

Table 2: AIC and BIC of all submodels

CPix	SPix	$CPix^2$	$SPix^2$	CPix:SPix	CHec (AIC/BIC)	SHec (AIC/BIC)
0	0	0	0	0	357.90 / 360.46	365.04 / 367.60
1	0	0	0	0	307.53 / 311.66	349.43 / 353.57
0	**1**	0	0	0	330.74 / 334.88	**314.23** / **318.37**
1	1	0	0	0	304.03 / 309.75	316.01 / 321.73
0	0	1	0	0	305.72 / 309.86	348.39 / 352.53
1	0	1	0	0	307.58 / 313.30	350.28 / 356.00
0	1	1	0	0	302.83 / 308.55	316.11 / 321.83
1	1	1	0	0	304.66 / 311.96	317.78 / 325.08
0	0	0	1	0	336.58 / 340.72	323.06 / 327.20
1	0	0	1	0	306.68 / 312.40	324.87 / 330.59
0	1	0	1	0	331.04 / 336.76	315.76 / 321.48
1	1	0	1	0	303.41 / 310.71	317.48 / 324.78
0	0	1	1	0	303.97 / 309.69	324.31 / 330.03
1	0	1	1	0	305.97 / 313.27	322.23 / 329.54
0	1	1	1	0	304.60 / 311.90	317.47 / 324.78
1	1	1	1	0	305.41 / 314.30	319.47 / 328.36
0	0	0	0	1	358.24 / 362.37	352.67 / 356.80
1	0	0	0	**1**	**301.93** / **307.65**	323.44 / 329.17
0	1	0	0	1	320.86 / 326.58	315.85 / 321.57
1	1	0	0	1	303.88 / 311.18	317.84 / 325.14
0	0	1	0	1	304.63 / 310.35	327.72 / 333.44
1	0	1	0	1	303.93 / 311.23	320.49 / 327.79
0	1	1	0	1	304.83 / 312.13	317.83 / 325.14
1	1	1	0	1	305.76 / 314.64	319.77 / 328.66
0	0	0	1	1	335.67 / 341.39	321.57 / 327.29
1	0	0	1	1	303.81 / 311.12	319.14 / 326.44
0	1	0	1	1	306.09 / 313.40	317.63 / 324.93
1	1	0	1	1	305.41 / 314.30	319.10 / 327.98
0	0	1	1	1	305.45 / 312.75	319.44 / 326.74
1	0	1	1	1	305.62 / 314.51	320.92 / 329.81
0	1	1	1	1	305.52 / 314.41	319.36 / 328.25
1	1	1	1	1	307.41 / 317.88	320.23 / 330.70

A.2 EBLUPs and Sumca Estimates of MSPEs

Table 3: Post Model Selection EBLUP $\hat{\theta}_i$

County	Corn (NotFixed	Corn (Fixed)	Soybean (Nonfixed)	Soybean (Nonfixed)
1	118.47	145.21	78.63	78.63
2	124.72	152.92	93.13	93.13
3	105.48	133.86	87.72	87.72
4	105.49	136.05	82.46	82.46
5	143.03	171.6	66.75	66.75
6	109.78	140.12	113.23	113.23
7	111.73	137.55	97.15	97.15
8	119.15	150.98	112.41	112.41
9	115.08	146.01	110.02	110.02
10	123.23	153.04	100.35	100.35
11	106.9	136.08	118.77	118.77
12	141.18	168.73	75.05	75.05

Table 4: Sumca Estimates of MSPEs (Plain & Nonfixed)

County	Corn (AIC)	Corn (BIC)	Soybean (AIC)	Soybean (BIC)
1	147.58	140.96	295.41	272.97
2	117.89	93.04	196.88	177.61
3	117.14	97.35	144.96	125.48
4	74.51	64.26	111.17	150.82
5	51.55	43.25	69.36	56.84
6	51.46	48.87	103.03	81.73
7	51.32	51.25	85.03	75.21
8	52.37	47	83.31	95.48
9	36.25	52.9	77.34	68.39
10	30.48	30.93	46.79	38.91
11	38.8	56.81	99.33	100.32
12	33.35	31.97	60.85	40.64

Table 5: Sumca Estimates of MSPEs (Plain & Fixed)

County	Corn (AIC)	Corn (BIC)	Soybean (AIC)	Soybean (BIC)
1	613.1	50560.95	114418.78	143359.26
2	3436.62	54726.62	123722.32	155322.75
3	739.44	53494.91	119695.29	149355.79
4	1596.71	61598.21	137980.46	172381.91
5	3244.21	57741.46	126498.12	160000.95
6	4076.95	59544.96	138194.5	169732.51
7	2937.55	46536.44	103633.95	130294.17
8	1097.33	66953.4	151319.99	188651.37
9	4185.18	62383.48	143943.75	176692.01
10	3771.59	60859.21	135466.6	170439.33
11	202.91	58888.09	131610.64	164633.28
12	2880.17	56276.06	122770	156017

Table 6: Sumca Estimates of MSPEs (M-parameterized & Fixed)

County	Corn (AIC)	Corn (BIC)	Soybean (AIC)	Soybean (BIC)
1	178.26	1248.02	514.31	1278.12
2	624.35	1051.95	437.03	1881.78
3	252.98	591.3	147.75	619.12
4	406.52	1071.05	467.93	1491.56
5	145.41	238.04	80.03	505.32
6	400.89	669.39	256.95	1251.94
7	366.15	749.56	159.7	876.98
8	322.03	869.88	282.89	1158.41
9	466.59	1437.86	310.77	1498.06
10	76.09	151.75	122.33	146.49
11	754.87	1758.29	348.77	2489.82
12	314.14	360.25	187.33	435.13

Table 7: Sumca Estimates of MSPEs (M-parameterized & Fixed)

County	Corn (AIC)	Corn (BIC)	Soybean (AIC)	Soybean (BIC)
1	20326.87	53099.28	31885.2	25301.25
2	13047.53	37979.49	38225.5	27609.28
3	11699.99	36167.72	34130.21	25027.25
4	22769.89	62531.56	37219.38	29800.26
5	24736.81	58473.46	42358.14	29769.57
6	11697.82	42310.92	28865.21	27119.35
7	12418.92	35180.55	32582.7	22890.09
8	18005.95	55214.99	41068.02	32101.4
9	12502.89	47492.18	29953.04	28053.34
10	14645.46	41428.5	42045.91	30055.72
11	13518.2	40656.23	38417.68	28546.12
12	16203.19	40349.33	44097.82	29050.68

A.3 R Code

```
##

# Global Variables ------------------------------------------------------

library(sae)
data("cornsoybean")
data("cornsoybeanmeans")
# delete 33rd sample point
cornsoybean <- cornsoybean[-33,]
m <- 12
PX1 <- cornsoybeanmeans$MeanCornPixPerSeg
PX2 <- cornsoybeanmeans$MeanSoyBeansPixPerSeg
PX1sq <- PX1^2
PX2sq <- PX2^2
PX1X2 <- PX1*PX2
PX = cbind(PX1,PX2,PX1sq,PX2sq,PX1X2)
ni <- c(1,1,1,2,3,3,3,3,4,5,5,5)
X1 <- cornsoybean$CornPix
X2 <- cornsoybean$SoyBeansPix
X1sq <- X1^2
X2sq <- X2^2
X1X2 <- X1*X2
Y <- cornsoybean$CornHec
#Y <- cornsoybean$SoyBeansHec
county = as.factor(rep(1:12,ni,each=T))
fit_F = lmer(Y~X1+X2+X1sq+X2sq+X1X2+(1|county),REML = F)
### control whether to use AIC/BIC
aic=TRUE
# aic=FALSE
K <- 100

# Model Selection -------------------------------------------------------

### model_selection

nested_model_selection<-function(response,
                                 data_matrix,
                                 candidate_variable=
                                   c("CornPix","SoyBeansPix","I(CornPix^2)","I(SoyBeansPix^2)","CornPix:SoyBeansPix
                                       "),
                                 AIC= TRUE){
  #candidate_variable<-c("CornPix","SoyBeansPix","I(CornPix^2)","I(SoyBeansPix^2)","CornPix:SoyBeansPix")
  n<-length(response)
  data_matrix<-cbind(response,data_matrix)
  colnames(data_matrix)[1]<-'response'
  County<-data_matrix$County
  n_county<-max(County)
  n_var<-length(candidate_variable)
  n_possible<-32
  candidate_encoding<-expand.grid(replicate(5,0:1,simplify=FALSE))[2:n_possible,]
  list_criterion<-rep(0,nrow(candidate_encoding))
  #Traverse all the possible conditions:
  for(i in 1:nrow(candidate_encoding)){
    involved_terms<-which(candidate_encoding[i,]!=0)
    #current_data<-data_matrix[involved_variables,]
    formula_1<-""
    for(k in 1:length(involved_terms)){
      formula_1<-paste(formula_1,"+",candidate_variable[involved_terms[k]],sep="")
    }
    current_formula<-paste("response~ (1|County) ",formula_1,sep="")
    current_model<-lmer(formula=current_formula,data=data_matrix,REML=FALSE)#,REML=FALSE)
    k1<-ncol(model.matrix(current_model))+2
    if(AIC==TRUE){
      list_criterion[i]<-(-2*logLik(current_model))+2*k1
    }
    else{
      #For BIC we use the alternative BIC criterion:
      list_criterion[i]<-(-2*logLik(current_model)+log(n)*(k1-2)+log(n_county)*2)
    }
  }
  #####
  #select the indice of what we select:
  variables_indice<-which(candidate_encoding[which.min(list_criterion),]!=0)
  formula_1<-""
  for(k in 1:length(variables_indice)){
    formula_1<-paste(formula_1,"+",candidate_variable[variables_indice[k]],sep="")
  }
  selected_formula<-paste("response~ (1|County) ",formula_1,sep="")
  selected_model<-lmer(formula=selected_formula,data=data_matrix,REML=FALSE)
  variable_selected<-candidate_variable[variables_indice]
  return(list(selected_model=selected_model,variables_indice=variables_indice,variable_selected=variable_selected))
}




# theta_with_quadr (non-Jiming) -----------------------------------------

##### theta_hat;; here out is the out of nested_model_selection
theta_with_quadr= function(out)
{
  candidate = out$variables_indice
  fit = out$selected_model
  beta = fixef(fit)
  ref = unlist(ranef(fit))
  result = beta[1]+as.matrix(PX[,candidate])%*%beta[2:(length(beta))]+ref
  result
}


# theta_no_quadr (Jiming) -----------------------------------------------
```

```
theta_no_quadr= function(out)
{
  candidate = out$variables_indice
  fit = out$selected_model
  beta_n = fixef(fit)
  beta =rep(0,3)
  beta[1] = beta_n[1]
  if(1 %in% candidate&2 %in% candidate)
  {
    beta[2:3] = beta_n[2:3]
  }
  if( 1%in% candidate &!2%in% candidate)
  {
    beta[2] = beta_n[2]
  }
  if( !1%in% candidate & 2 %in% candidate)
  {
    beta[3] = beta_n[2]
  }

  ref = unlist(ranef(fit))
  result = beta[1]+beta[2]*PX[,1]+beta[3]*PX[,2]+ref
  result
}


# variance ----------------------------------------------------

###### variance fit is the full model
variance = function(fit)
{
  sigma2_v = unlist(VarCorr(fit))
  sigma2_e = sigma(fit)^2
  sigma2_v*sigma2_e/(ni*sigma2_v+sigma2_e)
}



# cond_exp_with_quadr (Plain Sumca) --------------------------------

cond_exp <- function(Y, lmerF){
  f_beta <- fixef(lmerF)
  f_sig_v <- sqrt(unlist(VarCorr(lmerF)))
  f_sig_e <- sigma(lmerF)
  PX_new = cbind(1,PX)

  diag <- list()
  for (ns in ni) {
    diag <- c(diag, list(rep(1/ns,ns)))
  }
  Z <- bdiag(diag)
  ybar <- t(Z)%*%Y
  f_xbar <- t(Z)%*%cbind(X1,X2,X1sq,X2sq,X1X2)

  exp_result <- PX_new %*% f_beta + ni*f_sig_v^2/(ni*f_sig_v^2+f_sig_e^2)*
    (ybar-f_xbar%*%f_beta[2:6]-f_beta[1])

  return(exp_result)
}


# cond_exp_no_quadr (Plain Sumca; Jiming) ----------------------------------------------------------
cond_exp_no_quadr <- function(Y, lmerF){
  f_beta <- fixef(lmerF)
  f_sig_v <- sqrt(unlist(VarCorr(lmerF)))
  f_sig_e <- sigma(lmerF)
  PX_new = cbind(1,PX)

  diag <- list()
  for (ns in ni) {
    diag <- c(diag, list(rep(1/ns,ns)))
  }
  Z <- bdiag(diag)
  ybar <- t(Z)%*%Y
  f_xbar <- t(Z)%*%cbind(X1,X2,X1sq,X2sq,X1X2)

  exp_result <- PX_new[,1:3] %*% f_beta[1:3] + ni*f_sig_v^2/(ni*f_sig_v^2+f_sig_e^2)*
    (ybar-f_xbar%*%f_beta[2:6]-f_beta[1])

  return(exp_result)
}


# Compute a ---------------------------------------------------------------

#cond_exp: out 1, variance: out 2, theta: out 3.
a<-function(out1,out2,out3){
  a<-(out3-out1)^2+out2
  return(a)
}

# d_k_with_quadr (Plain Sumca; non-Jiming) ----------------------------------------

###
d_k_with_quadr = function(fit_F,aic=TRUE)
{
  Y_new  = simulate(fit_F)
  out_new = nested_model_selection(response = Y_new,cornsoybean,AIC=aic)
  names(Y_new) = "Y_new"
  fit_F_new = lmer(Y_new~X1+X2+X1sq+X2sq+X1X2+(1|county),data = cbind(Y_new,X1,X2,X1sq,X2sq,X1X2,county),REML = F)
  out1 = cond_exp(unlist(Y_new),fit_F)
  out2 = variance(fit_F)
  out3 = theta_with_quadr(out_new)
  a_former = a(out1,out2,out3)
  out12 = cond_exp(unlist(Y_new),fit_F_new)
  out22 = variance(fit_F_new)
  out32 = theta_with_quadr(out_new)
  a_latter = a(out12,out22,out32)
```

```
  as.vector(a_former-a_latter)
}

# MSPE1 (Plain Sumca; non-Jiming) ------------------------------------------------

MSPE1_fun <- function(Y, aic){
  fit_selected = nested_model_selection(Y,cornsoybean,AIC=aic)
  fit_F = lmer(Y~X1+X2+X1sq+X2sq+X1X2+(1|county),REML = F)
  out1 = cond_exp(Y,fit_F)
  out2 = variance(fit_F)
  out3 = theta_with_quadr(fit_selected)

  d = rep(0,12)
  for(i in 1:K)
  {
    d = d+d_k_with_quadr(fit_F,aic)
  }

  #For negative MSPE, we repeat them as the:
  mspe_estimation<-d/K+a(out1,out2,out3)
  indice<-which(mspe_estimation[,1]<0)
  mspe_estimation[indice,1]<-a(out1,out2,out3)[indice,1]
  print(mspe_estimation)
  return(mspe_estimation)
}




# EBLUP1 -------------------------------------------------------------------

EBLUP1 <- NULL
fit_selected = nested_model_selection(cornsoybean$CornHec,cornsoybean,AIC=T)
out3 = theta_with_quadr(fit_selected)
EBLUP1 <- cbind(EBLUP1, out3)
fit_selected = nested_model_selection(cornsoybean$CornHec,cornsoybean,AIC=F)
out3 = theta_with_quadr(fit_selected)
EBLUP1 <- cbind(EBLUP1, out3)
fit_selected = nested_model_selection(cornsoybean$SoyBeansHec,cornsoybean,AIC=T)
out3 = theta_with_quadr(fit_selected)
EBLUP1 <- cbind(EBLUP1, out3)
fit_selected = nested_model_selection(cornsoybean$SoyBeansHec,cornsoybean,AIC=F)
out3 = theta_with_quadr(fit_selected)
EBLUP1 <- cbind(EBLUP1, out3)


# store MSPE1 --------------------------------------------------------------

MSPE1 <- NULL
mspe1_tmp <- MSPE1_fun(cornsoybean$CornHec, aic=T)
MSPE1 <- cbind(MSPE1, mspe1_tmp)
mspe1_tmp <- MSPE1_fun(cornsoybean$CornHec, aic=F)
MSPE1 <- cbind(MSPE1, mspe1_tmp)
mspe1_tmp <- MSPE1_fun(cornsoybean$SoyBeansHec, aic=T)
MSPE1 <- cbind(MSPE1, mspe1_tmp)
mspe1_tmp <- MSPE1_fun(cornsoybean$SoyBeansHec, aic=F)
MSPE1 <- cbind(MSPE1, mspe1_tmp)

colnames(MSPE1) <- c("Corn+AIC", "Corn+BIC", "Soybean+AIC", "Soybean+BIC")
library(gridExtra)
library(grid)
pdf("MSPE1(Non-model, non-Jiming).pdf")
p <- grid.table(round(MSPE1,2))
dev.off()

# PRINT OUT ----------------------------------------------------------------


write.matrix(MSPE1, file = "MSPE1.csv", sep = ',')
write.matrix(MSPE2, file = "MSPE2.csv", sep = ',')
write.matrix(MSPE3, file = "MSPE3.csv", sep = ',')
write.matrix(MSPE4, file = "MSPE4.csv", sep = ',')
write.matrix(EBLUP1, file = "EBLUP1.csv", sep = ',')
write.matrix(EBLUP2, file = "EBLUP2.csv", sep = ',')
write.matrix(EBLUP3, file = "EBLUP3.csv", sep = ',')
write.matrix(EBLUP4, file = "EBLUP4.csv", sep = ',')


# plot1 --------------------------------------------------------------------

pl1 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP1[,1], color = "blue")) +
  geom_point(aes(x = (1:m+0.1), y = EBLUP1[,2], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Corn') +
  geom_errorbar(aes(x = 1:m, ymin=EBLUP1[,1]-2*sqrt(MSPE1[,1]), ymax=EBLUP1[,1]+2*sqrt(MSPE1[,1]), width=.2,
                color = 'blue')) +
  geom_errorbar(aes(x = (1:m+0.1), ymin=EBLUP1[,2]-2*sqrt(MSPE1[,2]), ymax=EBLUP1[,2]+2*sqrt(MSPE1[,2]), width=.2,
                color = 'red')) +
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC')) +
  theme(legend.position="top")

pl2 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP1[,3], color = "blue")) +
  geom_point(aes(x = 1:m+0.1, y = EBLUP1[,4], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Soybean')+
  geom_errorbar(aes(x = 1:m, ymin=EBLUP1[,3]-2*sqrt(MSPE1[,3]), ymax=EBLUP1[,3]+2*sqrt(MSPE1[,3]), width=.2,
                color = 'blue')) +
  geom_errorbar(aes(x = 1:m+0.1, ymin=EBLUP1[,4]-2*sqrt(MSPE1[,4]), ymax=EBLUP1[,4]+2*sqrt(MSPE1[,4]), width=.2,
                color = 'red')) +
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
```

```
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC')) +
  theme(legend.position="top")

pdf("plot1.pdf")
library(gridExtra)
grid.arrange(p11, p12, ncol=2)
dev.off()


# d_k_no_quadr (Plain Sumca; Jiming) ----------------------------------------

###
d_k_no_quadr = function(fit_F,aic=TRUE)
{
  Y_new  = simulate(fit_F)
  out_new = nested_model_selection(response = Y_new,cornsoybean,AIC=aic)
  names(Y_new) = "Y_new"
  fit_F_new = lmer(Y_new~X1+X2+X1sq+X2sq+X1X2+(1|county),data = cbind(Y_new,X1,X2,X1sq,X2sq,X1X2,county))
  # out1 = cond_exp(unlist(Y_new),fit_F)
  out1 = cond_exp_no_quadr(unlist(Y_new),fit_F)
  out2 = variance(fit_F)
  out3 = theta_no_quadr(out_new)
  a_former = a(out1,out2,out3)
  # out12 = cond_exp(unlist(Y_new),fit_F_new)
  out12 = cond_exp_no_quadr(unlist(Y_new),fit_F_new)
  out22 = variance(fit_F_new)
  out32 = theta_no_quadr(out_new)
  a_latter = a(out12,out22,out32)
  as.vector(a_former-a_latter)

}

# MSPE2_fun (Plain Sumca; Jiming) -------------------------------------------------

MSPE2_fun <- function(Y, aic){
  fit_selected = nested_model_selection(Y,cornsoybean,AIC=aic)
  # out1 = cond_exp(Y,fit_F)
  fit_F = lmer(Y~X1+X2+X1sq+X2sq+X1X2+(1|county),REML = F)
  out1 = cond_exp_no_quadr(Y,fit_F)
  out2 = variance(fit_F)
  out3 = theta_no_quadr(fit_selected)

  d = rep(0,12)
  for(i in 1:K)
  {
    d = d+d_k_no_quadr(fit_F,aic)
  }
  #For negative MSPE, we repeat them as the:
  mspe_estimation<-d/K+a(out1,out2,out3)
  indice<-which(mspe_estimation[,1]<0)
  mspe_estimation[indice,1]<-a(out1,out2,out3)[indice,1]
  print(mspe_estimation)
  return(mspe_estimation)
}




# EBLUP2 ----------------------------------------------------------------------

EBLUP2 <- NULL
fit_selected = nested_model_selection(cornsoybean$CornHec,cornsoybean,AIC=T)
out3 = theta_no_quadr(fit_selected)
EBLUP2 <- cbind(EBLUP2, out3)
fit_selected = nested_model_selection(cornsoybean$CornHec,cornsoybean,AIC=F)
out3 = theta_no_quadr(fit_selected)
EBLUP2 <- cbind(EBLUP2, out3)
fit_selected = nested_model_selection(cornsoybean$SoyBeansHec,cornsoybean,AIC=T)
out3 = theta_no_quadr(fit_selected)
EBLUP2 <- cbind(EBLUP2, out3)
fit_selected = nested_model_selection(cornsoybean$SoyBeansHec,cornsoybean,AIC=F)
out3 = theta_no_quadr(fit_selected)
EBLUP2 <- cbind(EBLUP2, out3)

# store MSPE2 -----------------------------------------------------------------

MSPE2 <- NULL
mspe2_tmp <- MSPE2_fun(cornsoybean$CornHec, aic=T)
MSPE2 <- cbind(MSPE2, mspe2_tmp)
mspe2_tmp <- MSPE2_fun(cornsoybean$CornHec, aic=F)
MSPE2 <- cbind(MSPE2, mspe2_tmp)
mspe2_tmp <- MSPE2_fun(cornsoybean$SoyBeansHec, aic=T)
MSPE2 <- cbind(MSPE2, mspe2_tmp)
mspe2_tmp <- MSPE2_fun(cornsoybean$SoyBeansHec, aic=F)
MSPE2 <- cbind(MSPE2, mspe2_tmp)

colnames(MSPE2) <- c("Corn+AIC", "Corn+BIC", "Soybean+AIC", "Soybean+BIC")
library(gridExtra)
library(grid)
pdf("MSPE2(Non-model, Jiming).pdf")
p <- grid.table(round(MSPE2,2))
dev.off()


# plot2 -----------------------------------------------------------------------

p21 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP2[,1], color = "blue")) +
  geom_point(aes(x = (1:m+0.1), y = EBLUP2[,2], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Corn') +
  geom_errorbar(aes(x = 1:m, ymin=EBLUP2[,1]-2*sqrt(MSPE2[,1]), ymax=EBLUP2[,1]+2*sqrt(MSPE2[,1]), width=.2,
                color = 'blue')) +
  geom_errorbar(aes(x = (1:m+0.1), ymin=EBLUP2[,2]-2*sqrt(MSPE2[,2]), ymax=EBLUP2[,2]+2*sqrt(MSPE2[,2]), width=.2,
                color = 'red')) +
```

```
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC')) +
  theme(legend.position="top")


p22 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP2[,3], color = "blue")) +
  geom_point(aes(x = 1:m+0.1, y = EBLUP2[,4], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Soybean')+
  geom_errorbar(aes(x = 1:m, ymin=EBLUP2[,3]-2*sqrt(MSPE2[,3]), ymax=EBLUP2[,3]+2*sqrt(MSPE2[,3]), width=.2,
                color = 'blue')) +
  geom_errorbar(aes(x = 1:m+0.1, ymin=EBLUP2[,4]-2*sqrt(MSPE2[,4]), ymax=EBLUP2[,4]+2*sqrt(MSPE2[,4]), width=.2,
                color = 'red')) +
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC')) +
  theme(legend.position="top")


pdf("plot2.pdf")
library(gridExtra)
grid.arrange(p21, p22, ncol=2)
dev.off()



# cond_exp_AddModel (Model Based; non-Jiming) ----------------------------------

cond_exp_AddModel <- function(Y, s_out){
  lmerS <- s_out$selected_model
  candidate <- s_out$variables_indice

  s_beta <- fixef(lmerS)
  s_sig_v <- sqrt(unlist(VarCorr(lmerS)))
  s_sig_e <- sigma(lmerS)

  s_PX <- PX[, candidate]
  s_PX_new <- cbind(1, s_PX)
  x <- cbind(X1,X2,X1sq,X2sq,X1X2)
  s_x <- x[, candidate]

  diag <- list()
  for (ns in ni) {
    diag <- c(diag, list(rep(1/ns,ns)))
  }
  Z <- bdiag(diag)
  ybar <- t(Z) %*% Y
  s_xbar <- t(Z) %*% s_x

  result <- s_PX_new %*% s_beta + ni * s_sig_v^2 / (ni*s_sig_v^2+s_sig_e^2)*
    (ybar-s_xbar%*%s_beta[2:length(s_beta)]-s_beta[1])

  # out <- s_out
  # candidate = out$variables_indice
  # fit = out$selected_model
  # beta_n = fixef(fit)
  # beta =rep(0,3)
  # beta[1] = beta_n[1]
  # if(1 %in% candidate&2 %in% candidate)
  # {
  #   beta[2:3] = beta_n[2:3]
  # }
  # if( 1%in% candidate &!2%in% candidate)
  # {
  #   beta[2] = beta_n[2]
  # }
  # if( !1%in% candidate & 2 %in% candidate)
  # {
  #   beta[3] = beta_n[2]
  # }
  #
  # PXbeta = beta[1]+beta[2]*PX[,1]+beta[3]*PX[,2]

  # result <- PXbeta + ni * s_sig_v^2 / (ni*s_sig_v^2+s_sig_e^2)*
  #   (ybar-s_xbar%*%s_beta[2:length(s_beta)]-s_beta[1])

}



# d_k_AddModel (Model Based; non-Jiming) ---------------------------------------------

d_k_AddModel <- function(s_out, aic=TRUE)
{
  fit_selected <- s_out$selected_model
  Y_new <- simulate(fit_selected)
  out_new <- nested_model_selection(response = Y_new, cornsoybean,AIC=aic)
  fit_selected_new <- out_new$selected_model
  names(Y_new) <- "Y_new"
  out1 <- cond_exp_AddModel(unlist(Y_new), s_out)
  out2 <- variance(fit_selected)
  out3 <- theta_with_quadr(out_new)
  a_former <- a(out1, out2, out3)
  out12 <- cond_exp_AddModel(unlist(Y_new),out_new)
  out22 <- variance(fit_selected_new)
  out32 <- theta_with_quadr(out_new)
  a_latter <- a(out12, out22, out32)
  return(a_former - a_latter)
}
```

```r
# MSPE3 (Model Based; non-Jiming) ----------------------------------------------------

MSPE3_fun <- function(Y, aic){
  s_out <- nested_model_selection(Y, cornsoybean, AIC=aic)
  fit_selected <- s_out$selected_model
  out1 <- cond_exp_AddModel(Y, s_out)
  out2 <- variance(fit_selected)
  out3 <- theta_with_quadr(s_out)

  d <- rep(0,12)
  for (i in 1:K) {
    d <- d + d_k_AddModel(s_out,aic)
  }
  #a(out1, out2, out3) + d/K
  #For negative MSPE, we repeat them as the:
  mspe_estimation<-d/K+a(out1,out2,out3)
  print(mspe_estimation)
  indice<-which(mspe_estimation[,1]<0)
  mspe_estimation[indice,1]<-a(out1,out2,out3)[indice,1]
  print(mspe_estimation)
  return(mspe_estimation)
}


# EBLUP3 ----------------------------------------------------------------------

EBLUP3 <- NULL
s_out <- nested_model_selection(cornsoybean$CornHec, cornsoybean, AIC=T)
out3 <- theta_with_quadr(s_out)
EBLUP3 <- cbind(EBLUP3, out3)
s_out <- nested_model_selection(cornsoybean$CornHec, cornsoybean, AIC=F)
out3 <- theta_with_quadr(s_out)
EBLUP3 <- cbind(EBLUP3, out3)
s_out <- nested_model_selection(cornsoybean$SoyBeansHec, cornsoybean, AIC=T)
out3 <- theta_with_quadr(s_out)
EBLUP3 <- cbind(EBLUP3, out3)
s_out <- nested_model_selection(cornsoybean$SoyBeansHec, cornsoybean, AIC=F)
out3 <- theta_with_quadr(s_out)
EBLUP3 <- cbind(EBLUP3, out3)



# store MSPE3 -----------------------------------------------------------------

MSPE3 <- NULL
mspe3_tmp <- MSPE3_fun(cornsoybean$CornHec, aic = T)
MSPE3 <- cbind(MSPE3, mspe3_tmp)
mspe3_tmp <- MSPE3_fun(cornsoybean$CornHec, aic = F)
MSPE3 <- cbind(MSPE3, mspe3_tmp)
mspe3_tmp <- MSPE3_fun(cornsoybean$SoyBeansHec, aic = T)
MSPE3 <- cbind(MSPE3, mspe3_tmp)
mspe3_tmp <- MSPE3_fun(cornsoybean$SoyBeansHec, aic = F)
MSPE3 <- cbind(MSPE3, mspe3_tmp)

colnames(MSPE3) <- c("Corn+AIC", "Corn+BIC", "Soybean+AIC", "Soybean+BIC")
library(gridExtra)
library(grid)
pdf("MSPE3(Model-based, non-Jiming).pdf")
p <- grid.table(round(MSPE3,2))
dev.off()


# plot3 -----------------------------------------------------------------------

p31 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP3[,1], color = "blue"), show.legend = T) +
  geom_point(aes(x = (1:m+0.1), y = EBLUP3[,2], color = "red"), show.legend = T) +
  geom_errorbar(aes(x = 1:m, ymin=EBLUP3[,1]-2*sqrt(MSPE3[,1]), ymax=EBLUP3[,1]+2*sqrt(MSPE3[,1]), color = 'blue'),
          width=.2,
                show.legend = T) +
  geom_errorbar(aes(x = (1:m+0.1), ymin=EBLUP3[,2]-2*sqrt(MSPE3[,2]), ymax=EBLUP3[,2]+2*sqrt(MSPE3[,2]), color = '
        red'), width=.2,
                show.legend = T) +
  scale_x_continuous(breaks = seq(0, 13, by = 1))+
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC'))+
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Corn') +
  theme(legend.position="top")


p32 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP3[,3], color = "blue")) +
  geom_point(aes(x = 1:m+0.1, y = EBLUP3[,4], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Soybean')+
  geom_errorbar(aes(x = 1:m, ymin=EBLUP3[,3]-2*sqrt(MSPE3[,3]), ymax=EBLUP3[,3]+2*sqrt(MSPE3[,3]), color = 'blue'),
          width=.2) +
  geom_errorbar(aes(x = 1:m+0.1, ymin=EBLUP3[,4]-2*sqrt(MSPE3[,4]), ymax=EBLUP3[,4]+2*sqrt(MSPE3[,4]), color = 'red
        '), width=.2) +
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC')) +
  theme(legend.position="top")


pdf("plot3.pdf")
library(gridExtra)
grid.arrange(p31, p32, ncol=2)
dev.off()
```

```
# cond_exp_new (Model Based; Jiming) ----------------------------------------

cond_exp_new = function(Y_new,out)
{
  difference = tapply(Y_new-Y,INDEX =county,FUN = mean)
  fit = out$selected_model
  sigma2_e = sigma(fit)^2
  sigma2_v = unlist(VarCorr(fit))
  t1 = theta_no_quadr(out = out)
  t2 = sigma2_v*ni/(sigma2_v*ni+sigma2_e)*(difference)
  t1+t2
}

# d_k_new (Model Based; Jiming) ---------------------------------------------

d_k_new = function(out, aic=TRUE)
{
  fit = out$selected_model
  Y_new  = simulate(fit)
  out_new = nested_model_selection(response = Y_new,cornsoybean,AIC=aic)
  names(Y_new) = "Y_new"
  fit_new = out_new$selected_model
  out1 = cond_exp_new(unlist(Y_new),out)
  out2 = variance(fit)
  out3 = theta_no_quadr(out_new)
  a_former = a(out1,out2,out3)
  out22 = variance(fit_new)
  as.vector(a_former-out22)
}

# MSPE4 (Model Based; Jiming) -----------------------------------------------

MSPE4_fun <- function(Y, aic){
  fit_selected = nested_model_selection(Y,cornsoybean,AIC=aic)
  out2 = variance(fit_selected$selected_model)
  out3 = theta_no_quadr(fit_selected)

  d = rep(0,12)
  for(i in 1:K)
  {
    d = d+d_k_new(fit_selected,aic)
  }

  d/K+out2

  #For negative MSPE, we repeat them as the:
  mspe_estimation<-d/K+out2
  indice<-which(mspe_estimation[1]<0)
  mspe_estimation[indice]<-out2[indice]
  print(as.matrix(mspe_estimation))
  return(as.matrix(mspe_estimation))
}

# EBLUP4 --------------------------------------------------------------------

EBLUP4 <- NULL
fit_selected <- nested_model_selection(cornsoybean$CornHec, cornsoybean, AIC=T)
out3 = theta_no_quadr(fit_selected)
EBLUP4 <- cbind(EBLUP4, out3)
fit_selected <- nested_model_selection(cornsoybean$CornHec, cornsoybean, AIC=F)
out3 = theta_no_quadr(fit_selected)
EBLUP4 <- cbind(EBLUP4, out3)
fit_selected <- nested_model_selection(cornsoybean$SoyBeansHec, cornsoybean, AIC=T)
out3 <- theta_no_quadr(fit_selected)
EBLUP4 <- cbind(EBLUP4, out3)
fit_selected <- nested_model_selection(cornsoybean$SoyBeansHec, cornsoybean, AIC=F)
out3 <- theta_no_quadr(fit_selected)
EBLUP4 <- cbind(EBLUP4, out3)



# store MSPE4 ---------------------------------------------------------------

MSPE4 <- NULL
mspe4_tmp <- MSPE4_fun(cornsoybean$CornHec, aic=T)
MSPE4 <- cbind(MSPE4, mspe4_tmp)
mspe4_tmp <- MSPE4_fun(cornsoybean$CornHec, aic=F)
MSPE4 <- cbind(MSPE4, mspe4_tmp)
mspe4_tmp <- MSPE4_fun(cornsoybean$SoyBeansHec, aic=T)
MSPE4 <- cbind(MSPE4, mspe4_tmp)
mspe4_tmp <- MSPE4_fun(cornsoybean$SoyBeansHec, aic=F)
MSPE4 <- cbind(MSPE4, mspe4_tmp)

colnames(MSPE4) <- c("Corn+AIC", "Corn+BIC", "Soybean+AIC", "Soybean+BIC")
library(gridExtra)
library(grid)
pdf("MSPE4(Model-based, Jiming).pdf")
p <- grid.table(round(MSPE4, 2))
dev.off()


# plot4 ---------------------------------------------------------------------

p41 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP4[,1], color = "blue")) +
  geom_point(aes(x = (1:m+0.1), y = EBLUP4[,2], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Corn') +
  geom_errorbar(aes(x = 1:m, ymin=EBLUP4[,1]-2*sqrt(MSPE4[,1]), ymax=EBLUP4[,1]+2*sqrt(MSPE4[,1]), width=.2,
                color = 'blue')) +
  geom_errorbar(aes(x = (1:m+0.1), ymin=EBLUP4[,2]-2*sqrt(MSPE4[,2]), ymax=EBLUP4[,2]+2*sqrt(MSPE4[,2]), width=.2,
                color = 'red')) +
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
```

```
                    labels = c('AIC','BIC')) +
  theme(legend.position="top")


p42 = ggplot() +
  geom_point(aes(x = 1:m, y = EBLUP4[,3], color = "blue")) +
  geom_point(aes(x = 1:m+0.1, y = EBLUP4[,4], color = "red")) +
  xlab('County') +
  ylab(expression(EBLUP %+-% 2*sqrt(MSPE))) +
  ggtitle('Soybean')+
  geom_errorbar(aes(x = 1:m, ymin=EBLUP4[,3]-2*sqrt(MSPE4[,3]), ymax=EBLUP4[,3]+2*sqrt(MSPE4[,3]), width=.2,
                color = 'blue')) +
  geom_errorbar(aes(x = 1:m+0.1, ymin=EBLUP4[,4]-2*sqrt(MSPE4[,4]), ymax=EBLUP4[,4]+2*sqrt(MSPE4[,4]), width=.2,
                color = 'red')) +
  scale_x_continuous(breaks = seq(0, 13, by = 1)) +
  scale_color_manual(name = 'Method', guide = 'legend',
                     values = c('blue', 'red'),
                     labels = c('AIC','BIC')) +
  theme(legend.position="top")


pdf("plot4.pdf")
library(gridExtra)
grid.arrange(p41, p42, ncol=2)
dev.off()



# plot all ------------------------------------------------------------------

require(cowplot)
g1 <- plot_grid(p11, p12)
g2 <- plot_grid(p21, p22)
g3 <- plot_grid(p31, p32)
g4 <- plot_grid(p41, p42)

pdf('AllPlots.pdf', width = 15, height = 10)
plot_grid(g1,g2,g3,g4,ncol = 2, labels = c('Plain&NotFix','Plain&Fix','M-Par&NotFix','M-par&Fix'))
dev.off()
```