An overview of Fay Herriot model with our package smallarea

Abhishek Nandy

May 3, 2013

1 The Fay Herriot Model

1.1 Model Notations

The Fay Herriot model can be written as follows:

$$y_i = x_i^T \beta + v_i + e_i \qquad i = 1, ..., m \tag{1}$$

where x_i is a vector of known covariates, β is a vector of unknown regresson coefficients, v_i's are area specific random effects and e_i's represent sampling errors. It is assumed that the $v_i \sim N(o, \psi)$ and $e_j \sim N(0, D_j)$ are independent for all pairs (i,j). We further assume that ψ is unknown but the D_i's are known. Each of the i's correspond to a small area. The ultimate goal of this package is to fit the fay herriot model to a given data and return estimates of several parameters noteworthy of which are the small area means, and mean square error of the small are mean. Below we give the definitions of the quantities of interest.

1.2 Vector Notation of the model

$$y = X\beta + v + \epsilon \tag{2}$$

where $v \sim N_m(0, \psi I)$ and $\epsilon \sim N_m(0, D)$ where N_m stands for the m dimensional Multivariate Normal Distribution I is the $m \times m$ identity matrix and D is an $m \times m$ diagonal matrix with elements $D_1, ..., D_m$. Also we will call the variance covariance matrix of Y as V, a diagonal matrix with $i - th$ diagonal element = $\psi + D_i$

1.3 Variance component

There is a more general definition of variance component, but for our purposes ψ, the variance of the random effects is the variance component.

1.4 estimate of the variance component

There are four popular ways of estimating the variance components

- The Prasad Rao estimator denoted by $\widehat{\psi}_{PR}$ has the form $\frac{y^T(I-P_X)y-tr((I-P_X)D)}{m-p}$ where $P_X = X(X^TX)^{-1}X^T$ is the projection matrix onto the column space of X.
- The Fay Herriot estimator denoted by $\widehat{\psi}_{FH}$ is obtained by solving the equation

$$\frac{y^TQ(\psi)y}{m-p} = 1 \tag{3}$$

  iteratively, where

$$y^TQ(\psi)y = \sum_{i=1}^{m} \frac{(y_i - x_i^T\beta)^2}{\psi + D_i}$$

  We will use unirrot function in the stats package of R to solve the equation.

- The Maximum likelihood estimator denoted by $\widehat{\psi}_{ML}$ is obtained by maximizing the likelihood function. Note that in our model assumptions we have already incorporated normality assumption. So the likelihood would be a normal likelihood. hence it should be made sure while fitting the model that the normality assumptions are met.
- The REML (aka Residual Maximum Likelihood aka Restricted Maximum Likelihood) encompasses the main drawback of the likelihood approach of estimating by adjusting for the degrees of freedom involved in estimating the fixed effects. In broad generality By multiplying with vectors k that belong to the null space of the matrix X on both sides of equation 2, we can get

$$k^Ty = k^Tv + kTv$$

  But this makes us vulnerable to the option that the estimator might depend on the choice of the vector k. However that problem can be resolved by finding REML estimators as a solution to the REML equations. For details see Searle, Casella, McCulloch (Wiley, 2006), sec 6.6. We then use the scoring method to solve the REML equations. Without going into the general theory of the REML we will write down the equation for implementation of the scoring algorithm here. The iterates in fisher scoring are given by the following formula

$$\psi^{(n+1)} = \psi^{(n)} + [I_R\psi^{(n)}]^{-1}s_R^{-1}(\psi^{(n)}) \tag{4}$$

where

$$
\begin{aligned}
I_R(\psi) &= \frac{1}{2}tr(PBPB)\\
s_R(\psi) &= -\frac{1}{2}tr(PB) + \frac{1}{2}y^T PBPy\\
P &= V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}
\end{aligned}
$$

where V is the variance covariance matrix of Y, a diagonal matrix with $i-th$ diagonal element $= \psi + D_i$. There is a more general definition of B (see the references), But for our purposes B is the Identity matrix.

1.5 Estimates of the Regression Coefficients

Estimates of the regression coefficients are given by the formula $(X^T V^{-1} X)^{-1} X^T V^{-1} y$ and is denoted by $\tilde{\beta}$. However note that V is a function of unknown ψ and known D_i's. So ψ's have to be estimated as we discussed in the section 1.2.5. we will often denote by $\tilde{\beta}$ by $\tilde{\beta}(\psi)$ when the V in the above formula is has not been estimated from data, whereas if the V in the above formula has been estimated from data(denote by $\widehat{V}$), it will be denoted by $\tilde{\beta}(\widehat{\psi})$ and sometimes by $\widehat{\beta}$. In Fay herriot Model the expressions of $\tilde{\beta}$ is $(\sum_{i=1}^{n} \frac{x_i x_i^T}{\psi + D_i})^{-1}(\sum_{i=1}^{n} \frac{x_i y_i}{\psi + D_i})$ and that $\widehat{\beta}$ is the same expression with ψ replaced by $\widehat{\psi}$

1.6 Small Area mean

The small area mean is a quantity $\eta = X\beta + v$ for our purposes.
The small area mean for the i-th small area is thus $\eta_i = x_i^T \beta + v_i$

1.7 The estimated Small Area mean: The BLUP and EBLUP

If β and ψ are both known the best predictor of η is given by

$$
\begin{aligned}
E(\eta|y) &= X\beta + E(v|y)\\
&= X\beta + \begin{pmatrix} \frac{\psi}{\psi + D_1}(y_1 - x_1^T \beta) \\ . \\ . \\ . \\ \frac{\psi}{\psi + D_m}(y_m - x_m^T \beta) \end{pmatrix}
\end{aligned}
$$

However if only β is unknown, it is customary to replace β in the above expression by $\tilde{\beta}$ and the result will be called BLUP.

Further if both β and ψ are unknown, we replace ψ by $\widehat{\psi}$ and β by $\widehat{\beta}$ and the result will be called EBLUP.

1.8 Estimate of the Mean Squared Prediction error(MSPE) of EBLUP

Another feature of our package is that it gives second order unbiassed estimator of the mean squared prediction error(MSPE) of the EBLUP of the small area means. The MSPE of the EBLUP of the $i-th$ small area is given by

$$\begin{aligned} MSPE(\widehat{\eta}_i) &= E(\tilde{\eta}_{B,i}-\eta_i)^2+E(\tilde{\eta}_i-\tilde{\eta}_{B,i})^2+E(\widehat{\eta}_i-\tilde{\eta}_i)^2 \\ & \quad \text{where} \quad \tilde{\eta}_{B,i}=\frac{\psi}{\psi+D_i}y_i-\frac{D_i}{\psi+D_i}x_i^T\beta \\ &= g_{1i}(\psi)+g_{2i}(\psi)+E(\widehat{\eta}_i-\tilde{\eta}_i)^2 \end{aligned}$$

Further analytic expressions can be obtained of the first two terms

$$\begin{aligned} g_{1i}(\psi) &= \frac{\psi D_i}{\psi+D_i} \\ g_{2i}(\psi) &= (\frac{D_i}{\psi+D_i})^2 x_i^T[\sum_{j=1}^{n}(\frac{x_j x_j^T}{\psi+D_j})]^{-1}x_i \end{aligned}$$

Note that $g_{1i}(\psi)=O(1)$ and $g_{2i}(\psi)=O(m^{-1})$ And the third term can be written as

$$E(\widehat{\eta}-\tilde{\eta})^2=g_{3i}(\psi)+o(m^{-1})$$

The expression of g_{3i} varies from case to case. We list below the different expressions

$$\begin{aligned} g_{3i}(\psi) &= \frac{2D_i^2}{(\psi+D_i)^3m^2}\sum_{j=1}^{m}(\psi+D_j)^2 \quad \text{for Prasad Rao} \\ &= \frac{2D_i^2}{(\psi+D_i)^3}[\sum_{j=1}^{m}(\psi+D_j)^{-2}]^{-1} \quad \text{for ML and REML} \\ &= \frac{2D_i^2 m}{(\psi+D_i)^3}[\sum_{j=1}^{m}(\psi+D_j)^{-1}]^{-2} \quad \text{for Fay Herriot} \end{aligned}$$

Finally we list below the second order approximated estimates of the MSPE we described above

$$
\begin{aligned}
\widehat{MSPE}(\widehat{\eta}_i) &= g_{1i}(\widehat{\psi}_{PR}) + g_{2i}(\widehat{\psi}_{PR}) + 2g_{3i,PR}(\widehat{\psi}_{PR}) \quad \text{for prasad rao} \\
\widehat{MSPE}(\widehat{\eta}_i) &= g_{1i}(\widehat{\psi}_{REML}) + g_{2i}(\widehat{\psi}_{REML}) + 2g_{3i,REML}(\widehat{\psi}_{REML}) \quad \text{for REML} \\
&= g_{1i}(\widehat{\psi}_{ML}) + g_{2i}(\widehat{\psi}_{ML}) + 2g_{3i,ML}(\widehat{\psi}_{ML}) \\
&\quad -(\frac{D_i}{\widehat{\psi}_{ML} + D_i})^2 \{\sum_{j=1}^{m} (\widehat{\psi}_{ML} + D_j)^{-2}\}^{-1} \\
&\quad \times tr\{(\sum_{j=1}^{m} \frac{x_j x_j^T}{\widehat{\psi}_{ML} + D_j})^{-1} \sum_{j=1}^{m} \frac{x_j x_j^T}{(\widehat{\psi}_{ML} + D_j)^2}\} \quad \text{for ML} \\
\widehat{MSPE}(\widehat{\eta}_i) &= g_{1i}(\widehat{\psi}_{FH}) + g_{2i}(\widehat{\psi}_{FH}) + 2g_{3i,FH}(\widehat{\psi}_{FH}) \\
&\quad -2(\frac{D_i}{\widehat{\psi}_{FH} + D_i})^2 \{\sum_{j=1}^{m} (\widehat{\psi}_{FH} + D_j)^{-1}\}^{-3} \\
&\quad \times [m \sum_{j=1}^{m} (\widehat{\psi}_{FH} + D_j)^{-2} - \{\sum_{j=1}^{m} (\widehat{\psi}_{FH} + D_j)^{-1}\}^2]
\end{aligned}
$$

For the details on the derivation of these results please see the references.

2 Examples

Now let us look at a simple examples.

2.1 First Example: prasadraoest

In the first example we will demonstrate what our function prasadraoest does. This is an auxiliary function of our package. So it has not been designed to be very user friendly as its main purpose is to compute the prasad rao estimate of variance component ψ mentioned in section 1.4. The arguements that this function takes are as follows

- response : This is the y vector. This must be a numeric vector.
- desgnmatrix: This is the matrix X (matrix of covariates with the first column needs to have all entries equal to one). This must be a numeric matrix.
- sampling.var: This is the vector consisting of the $D_i$values. This must be a numeric vector.

```
> library(smallarea)
> response=c(1,2,3,4,5)  # response vector
> designmatrix=cbind(c(1,1,1,1,1),c(1,2,4,4,1),c(2,1,3,1,5))
> # designmatrix with 5 rows and 3 columns,
```

```
> # the first column has all entries equal to one
> sampling.var=c(0.5,0.7,0.8,0.4,0.5)
> # This is the vector of sampling variances
> answer=prasadraoest(response,designmatrix,sampling.var)
> answer

$estimate
[1] 1.780361

> answer$estimate

[1] 1.780361
```

This function returns a list with only one element in it the estimate of the variance component

2.2 Second example: fayherriot

This function is pretty similar to the `prasadraoest` function in appearance, i.e takes the same arguments (example 1 gives a description of the arguments in details) and returns a list with only one element, the estimate of the variance component, except for the fact thatthe method of estimation is the one that was proposed by Fay Herriot, the formula is given in section 1.4. That is the estimate is obtained by numerically solving the equation (3) using the `uniroot` function in the `stats` package in R. Uniroot searches for the root of that equation in an interval which we have specified as $(0, \widehat{\psi}_{PR} + 3\sqrt{m})$. If no root is found in that interval, we have truncated our $\widehat{\psi}_{FH}$ at 0.0001 as suggested by Datta, Rao Smith (2005). To demonstrate the working of this function, we have used the same data as in `prasadraoest`

```
> response=c(1,2,3,4,5)
> designmatrix=cbind(c(1,1,1,1,1),c(1,2,4,4,1),c(2,1,3,1,5))
> sampling.var=c(0.5,0.7,0.8,0.4,0.5)
> fayherriot(response,designmatrix,sampling.var)

$estimate
[1] 1.793244
```

2.3 Third Example : maximlikelihood

This is an example which demonstrates the function`maximlikelihood`. The arguments are again same as the last two examples. It returns a list that has three elements

- estimate: is $\widehat{\psi}_{ML}$.
- reg.coefficients: a vector of the MLE of the regression coefficients i.e. β vector

- loglikeli.optimum : the value of the log likelihood function at the maximized value.

we have used the `optim` function in the `stats` package in R and used the BFGS algorithm to minimize the negative log likelihood. The maximum number of iterations of the algorithm is 100(the default in optim). Since the likelihhod function is a normal likelihood, care should be taken to check the normality assumptions of the data. Here the response has been generted from a normal distribution with mean 3 and standard deviation 1.5. The designmatrix and the sampling variances are however kept the same as in the last two examples.

```
> set.seed(55)
> response=rnorm(5,3,1.5)
> designmatrix=cbind(c(1,1,1,1,1),c(1,2,4,4,1),c(2,1,3,1,5))
> sampling.var=c(0.5,0.7,0.8,0.4,0.5)
> maximlikelihood(response,designmatrix,sampling.var)

$estimate
[1] 1.217849

$reg.coefficients
[1] 0.61531211 0.07861667 0.58275733

$loglikeli.optimum
[1] -2.459195
```

2.4 Fourth Example: resimaxilikelihood

The getup of this function is again very similar to the first two examples, except here there is one additional argument `maxiter` wherein the user can specify the maximum number of Fisher scoring iterations. Also here using Fisher's scoring iteration described in equation (4) in section 1.4 we find the REML of the variance component. A description of the data used is already given in third example. It returns a list of two elements

- estimate: $\widehat{\psi}_{REML}$
- iterations: number of fisher scoring iterations

```
> set.seed(55)
> response=rnorm(5,3,1.5)
> designmatrix=cbind(c(1,1,1,1,1),c(1,2,4,4,1),c(2,1,3,1,5))
> sampling.var=c(0.5,0.7,0.8,0.4,0.5)
> resimaxilikelihood(response,designmatrix,sampling.var,maxiter=100)

$estimate
[1] 1.207907

$iterations
[1] 2
```

2.5 Fifth Example: smallareafit

This is the main function in our library. It takes the following arguments

- formula: a formula similar in appearance to that of in `lm` function in R. You have to make sure that the data contains a column of the sampling variances, and that while specifying the formula the the name of the variable that contains the sampling variances should preceede the variables which are the covariates. e.g In the following example `response~D+x1+x2` is a correct way of specifying the formula where as `response~x1+D+x2` is not.(note D is the variabe that contains the values of sampling variances and x1 and x2 are covariates). In general the first of the variables on the right hand side of $\sim$ will be treated as the vector of sampling variance.
- data : It is an optional data.frame. In absense of this argument our function will accept variables from the global environment
- method : It can be one of the four methods discussed in this article and can be any one of the following options "PR","FH","ML","REML"

In usage of each of the four methods, the following will be the output. The function will return a list that has the following

- smallmean.est: a numeric vector of The EBLUP of the small area means discussed in section 1.7
- smallmean.mse: a numeric vector of The estimated Mean squared Prediction error $\widehat{MSPE}(\widehat{\eta}_i)$ discussed in section 1.8
- var.comp : an estimate of the variance component discussed in section 1.4
- est.coef: a numeric vector containg the estimates of the regression coefficients as discussed in section 1.5 and is denoted by $\widehat{\beta}$ beginning with the intercept in the order as specified in the formula.

The following example just illustrates what we have just discussed. Also note that the maximum number of Fisher scoring iterations have been set to 100 for the REML proceedure.

```
> data=data.frame(response=rnorm(5,3,1.5),
+ x1=c(1,2,4,4,1),x2=c(2,1,3,1,5),D=c(0.5,0.7,0.8,0.4,0.5))
> data

  response x1 x2   D
1 4.782778  1  2 0.5
2 2.241984  2  1 0.7
3 2.851148  4  3 0.8
4 3.458030  4  1 0.4
5 3.297615  1  5 0.5
```

```
> ans=smallareafit(response~D+x1+x2,data,method="FH")
> ans1=smallareafit(response~D+x1+x2,data,method="REML")
> ans2=smallareafit(response~D+x1+x2,data,method="PR")
> ans3=smallareafit(response~D+x1+x2,data,method="ML")
> ans # FH method

$smallmean.est
[1] 4.424377 2.821458 2.873995 3.336231 3.380075

$smallmean.mse
        1         2         3         4         5
0.5729548 0.7319444 0.8677240 0.4727353 0.6264921

$var.comp
[1] 0.9183763

$est.coef
[1]  4.18501040 -0.26255996 -0.07818307

> ans1 # REML method

$smallmean.est
[1] 4.421117 2.826725 2.874197 3.335126 3.380827

$smallmean.mse
        1         2         3         4         5
0.5730635 0.7309004 0.8665975 0.4731850 0.6270866

$var.comp
[1] 0.9047237

$est.coef
[1]  4.18625952 -0.26266982 -0.07843891

> ans2 # PR method

$smallmean.est
[1] 4.427651 2.816168 2.873791 3.337340 3.379320

$smallmean.mse
        1         2         3         4         5
0.5770696 0.7388315 0.8753713 0.4755707 0.6301189

$var.comp
[1] 0.9323386

$est.coef
[1]  4.18375604 -0.26244965 -0.07792615
```

```
> ans3 # ML method

$smallmean.est
[1] 4.403782 2.780007 2.871596 3.319181 3.384635

$smallmean.mse
        1         2         3         4         5
0.4620107 0.5662407 0.6757235 0.3901495 0.5150600

$var.comp
[1] 0.9323386

$est.coef
[1]  4.04772592 -0.25053063 -0.05005895
```

2.6 An example where smallarea accepts variables from the global environment

Our final example is that of usage of small area function that accepts variables from Global environment

```
> data=data.frame(response=rnorm(5,3,1.5),
+ x1=c(1,2,4,4,1),D=c(0.5,0.7,0.8,0.4,0.5))
> attach(data)

The following object(s) are masked _by_ '.GlobalEnv':

    response

> ans=smallareafit(response~D+x1,method="FH")
> ans1=smallareafit(response~D+x1,method="REML")
> ans2=smallareafit(response~D+x1,method="PR")
> ans3=smallareafit(response~D+x1,method="ML")
> ans

$smallmean.est
[1] 3.0150163 0.8869773 2.7774122 1.4308082 2.8797010

$smallmean.mse
        1         2         3         4         5
0.5321873 0.6819030 0.8236020 0.4367694 0.5321873

$var.comp
[1] 1.609707

$est.coef
[1]  2.686916 -0.203722
```

```
> ans1

$smallmean.est
[1] 3.0085987 0.9102301 2.7600286 1.4350709 2.8750459

$smallmean.mse
        1         2         3         4         5 
0.5335419 0.6799558 0.8221450 0.4388396 0.5335419 

$var.comp
[1] 1.524801

$est.coef
[1]  2.6905866 -0.2053299

> ans2

$smallmean.est
[1] 3.0215923 0.8631283 2.7952330 1.4264405 2.8844817

$smallmean.mse
        1         2         3         4         5 
0.5319471 0.6855882 0.8270283 0.4355317 0.5319471 

$var.comp
[1] 1.703841

$est.coef
[1]  2.6831519 -0.2020736

> ans3

$smallmean.est
[1] 3.009898 0.859762 2.817091 1.439448 2.872787

$smallmean.mse
        1         2         3         4         5 
0.4833025 0.6056660 0.7309263 0.4013165 0.4833025 

$var.comp
[1] 1.703841

$est.coef
[1]  2.5916196 -0.1620874
```

2.7 A final example

Now we will actually use a more interesting example taken forom the paper Reference 1 .We take the total number of small areas, n=15, $\psi = 1$ and three sampling variance, D_i-patterns: 0.7, 0.6, 0.5, 0.4, 0.3; There are five groups $G_1, ..., G_5$ and three small areas in each group. The sampling variances D_i are the same for areas within the same group. We will investigate through simulation for the area level model without covariates $x_i^T \beta = \mu$. Since the mean squared error is translation invariant, we set $\mu = 0$ without loss of generality. However, to account for the estimation of unknown regression parameters that arise in applications, we will still estimate this zero mean. We will consider distributions for the variance components v_i's, namely N(0, 1). The sampling error, e_i, will be generated from $N(0, D_i)$ for D_i as specified above.

```
> set.seed(55)
> # the sampling variances
> D=c(rep(0.7,3),rep(0.6,3),rep(0.5,3),rep(0.4,3),rep(0.3,3))
> # generating the errors
> e1=rnorm(3,0,sqrt(D[1]))
> e2=rnorm(3,0,sqrt(D[4]))
> e3=rnorm(3,0,sqrt(D[7]))
> e4=rnorm(3,0,sqrt(D[10]))
> e5=rnorm(3,0,sqrt(D[13]))
> e=c(e1,e2,e3,e4,e5)
> psi=1
> # generating the random small area effects
> v=rnorm(15,0,psi)
> # response
> y=v+e
> data1=data.frame(response=y,D=D)
> head(data1)

    response   D
1 -0.2657625 0.7
2  0.8390206 0.7
3  1.2202007 0.7
4 -0.5811039 0.6
5  0.9951359 0.6
6 -0.5986495 0.6

> fit1.pr=smallareafit(response~D,data1,method="PR")
> fit1.pr

$smallmean.est
 [1]  0.01933153  0.44238767  0.58835347 -0.13020013  0.53174366 -0.13756843
 [7]  0.63489543  0.45342849  0.70099509  0.54141111  0.57794174 -0.54392266
[13]  0.23947479 -1.00295090  0.02844395

$smallmean.mse
        1         2         3         4         5         6         7         8
0.3711601 0.3711601 0.3711601 0.3499072 0.3499072 0.3499072 0.3228506 0.3228506
```

```
        9        10        11        12        13        14        15
0.3228506 0.2878219 0.2878219 0.2878219 0.2417183 0.2417183 0.2417183

$var.comp
[1] 0.4343958

$est.coef
[1] 0.196251
```

We have used only Prasad Rao method in the last example, similarly other methods can also specified.

3 References

- Datta-Rao-Smith **On measuring the variability of small area estimators under a basic area level model** , *Biometrika* (2005), 92, 1, pp. 183-196.
- Large Sample Techniques for Statistics, Jiming Jiang, Springer Texts in Statistics
- REML Estimation: Asymptotic Behavior and Related Topics, *Annals of Statistics,1996, Vol 24, no.1, 255-286*, Jiang
- Small Area Estimation, JNK Rao, Wiley 2003.
- Variance Component, Searle, Casella, McCulloch, 2006, Wiley Series in Probability and Statistics

4 Acknowledgments

- **Snigdhansu Chatterjee, School of Statistics, University of Minnesota.**
- **Galin Jones, School of Statistics, University of Minnesota.**
- **Charles Geyer, School of Statistics, University of Minnesota**