Robust Hierarchical Bayesian Analysis Applied to Small Area Estimation

Fernando Moura

IM - UFRJ

joint work with Helio S. Migon and Kelly Cristina M. G.

ISI Satellite Meeting on Small area estimation, Bangkok, Thailand

September 2013
Summary

1. The problem
2. The main aim
3. The t-Student hierarchical model
4. The approximate Objective prior
5. Application
6. A Simulation Study
7. Concluding Remarks and Future Work
8. References
Hierarchical linear normal models are widely used to borrow strength from "exchangeable groups" in various stages of the hierarchy. However, it does not allow for the presence of atypical cases. This happens because the usual normal approach shrinks a fixed proportion to all groups and does not make any exception.
Some related Bayesian literature review

- [Datta and Lahiri(1995)]:
  - proposed to robustify the Fay-Herriot model by assuming a scale normal mixture;
  - provided conditions under which the joint posterior distribution is proper.
  - They noticed that these conditions hold for many distributions in the scale mixtures of normal family including t-student, certain distributions in the exponential power family, such as double exponential and logistic.

- [Bell and Huang(2006)] used hierarchical Bayes method based on t-distribution with $k > 2$ known degrees of freedom to deal with outliers either in the small area effect or in the sampling error effect.

- [Fabrizi and Trivisano(2010)] proposed to robustify the Fay-Herriot model by assuming that the random area effects are distributed according to either an exponential power (EP) distribution or a skewed EP distribution.
The main aim of this work is to propose a further extension of the Fay-Heriot model by assuming that the area random effects follow a t-distribution as [Bell and Huang(2006)], but with unknown degree of freedom.

We also developed an ”approximate” objective priors for all hyperparameters of the model based on [Sun and Berger(1998)] and exploited by Liseo et al.(2010) for accommodating latent structure.
The t-student hierarchical model

Notation:
- \( m \) denotes the number of small area selected;
- \( y_i \) denotes survey direct estimator for the \( ith \) small area and \( s_i^2 \) its respective sampling variance estimator.

The model
- First level
  \[
  y_i \mid \mu_i, v_i^{-1} \sim N(\mu_i, v_i^{-1}) \quad \text{where} \quad v_i^{-1} = \frac{\sigma_i^2}{n_i}
  \]
  \[
  s_i^2 \mid n_i, \sigma_i^2 \sim Ga \left\{ 0.5(n_i - 1), 0.5(n_i - 1)\sigma_i^{-2} \right\}, \quad \text{for} \quad i = 1, \ldots, m,
  \]

- Second level
  \[
  \mu_i \mid \alpha, \beta, \sigma_\nu \sim T(x_i^T \beta, \sigma_\nu^2, \alpha), \quad \text{or} \quad \mu_i = x_i^T \beta + \delta_i \quad \text{where} \quad \delta_i \sim T(0, \sigma_\nu^2, \alpha) \quad \text{and}
  \]
  \[
  v_i \mid a, b \sim Ga(a, b)
  \]
Objective Bayesian analysis has a strong appeal when prior information is absent.

Furthermore, it is particularly useful in applications from the frequentist perspective, since it produces point and interval estimation with good repeat sampling properties, see Berger et al. (2009) for examples.

Objective priors approaches are based on the calculation of the expected Fisher information matrix ($I(\theta)$), although an alternative method was proposed by Berger et al. (2009).

However, in some practical applications, the calculation of $I(\theta)$ is not feasible or cannot be obtained in closed form.

To overcome this problem, Liseo, et al. (2010) proposed the introduction of a vector of latent quantities $z$ and pretended that they are additional vector of observations.
The aim is to obtain an approximate objective priors for the hyperparameters $\theta = (\beta, \sigma_z, \nu, a, b)$.

Let $y = (y_1, ..., y_m)^T$ and $s^2 = (s^2_1, ..., s^2_m)^T$ be the bivariate responses of the model and $z = (\delta_1, ..., \delta_m, \nu_1, ..., \nu_m)^T$ be the random latent variables.

Thus the logarithm of the extended likelihood is given by:

$$l(\theta, z) = \log[f(y, s^2 | \theta, z)] = \sum_{i=1}^{m} \log[f(y_i | x_i, \beta, \nu_i, \nu_i) f(s^2_i | \nu_i) f(\nu_i | a, b) f(\delta_i | \sigma_\delta, \nu)]$$
Approximate Objective Priors

- Thus the approach consists in calculating $I(\theta)$ as equal to $E_{(y,s,z)} \left( - \frac{\partial l(\theta, z)}{\partial \theta} \right)$, where this expectation is evaluated over the joint distribution of the set of data $(y, s)$ and the latent vector $z$.
- It is shown that $I(\theta)$ is block-diagonal with the following structure:

$$I(\theta) = \begin{bmatrix} I_\beta & O_{p\times2} & O_{p\times2} \\ O_{2\times1} & I_{\sigma_\delta,\nu} & O_{4\times4} \\ O_{2\times1} & O_{4\times4} & I_{a,b} \end{bmatrix}$$

where:

$$I_\beta = \frac{b}{a} \sum_{i=1}^{m} x_i x_i^T$$

$$I_{\sigma_\delta,\nu} = \begin{bmatrix} \frac{2m}{\sigma_\delta^2} \frac{\nu}{\nu+3} & - \frac{2m}{\sigma_\delta} \frac{1}{(\nu+1)(\nu+3)} \\ - \frac{2m}{\sigma_\delta} \frac{1}{(\nu+1)(\nu+3)} & \frac{m}{4} \left\{ \psi' \left( \frac{\nu}{2} \right) - \psi' \left( \frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\alpha(\nu+1)(\nu+3)} \right\} \end{bmatrix}$$

$$I_{a,b} = \begin{bmatrix} m \psi'(a) & - \frac{m}{b} \\ - \frac{m}{b} & \frac{ma}{b^2} \end{bmatrix}$$

and $\psi(u) = \frac{d\log\Gamma(u)}{du}$ and $\psi'(u) = \frac{d\psi(u)}{du}$ are the digamma and trigamma functions, respectively.
Approximate Objective Priors

Applying Jeffreys-rule to the results obtained above, we obtain the following:

\[ p(\beta, \sigma_\delta, a, b) \propto |I(\theta)|^{1/2} = \left\{ |I_\beta| |I_{\sigma_\delta,\nu}| |I_{a,b}| \right\}^{1/2} \]

\[ \propto a^{-p/2} b^{\frac{p-2}{2}} (a \psi'(a) - 1)^{1/2} \left( \frac{\nu}{\nu + 3} \right)^{1/2} \left\{ \psi' \left( \frac{\nu}{2} \right) - \psi' \left( \frac{\nu + 1}{2} \right) - \frac{2(\nu + 3)}{\nu(\nu + 1)^2} \right\}^{1/2} \]

We can easily derive the "independence Jeffreys prior" by assuming that the marginal priors for \( \beta \) and \((a, b, \alpha)\) are independent a priori, and separately computing priors for each of these groups of parameters by applying a Jeffreys-rule prior. This yields to:

\[ p_I(\beta, \sigma_\delta, a, b) \propto b^{-1} (a \psi'(a) - 1)^{1/2} \left( \frac{\nu}{\nu + 3} \right)^{1/2} \left\{ \psi' \left( \frac{\nu}{2} \right) - \psi' \left( \frac{\nu + 1}{2} \right) - \frac{2(\nu + 3)}{\nu(\nu + 1)^2} \right\}^{1/2} \]
Although the prior are improper, it can be shown that the posterior are proper.

The marginal posterior for the degree of freedom has no mean.

The prior for the degree of freedom is the same as obtained by Fonseca, et al. (2008).
Trial Census in a certain Brazilian municipality

- 140 areas
- 38740 households (population units)
- characteristic of interest: head of household income
- area level covariates:
  - small area population means of the educational attainment of the Head of Household (ordinal scale of 0 – 5) and the number of rooms in the household (1 – 11+).
  - We center both covariates towards their respective overall population means.
  - The number of households per area in the population varies from 57 to 588.
Two sets of samples are used to evaluate our proposed model: 10% and 5% stratified random sample of households in each area.

A preliminary analysis of the income variable reveals that it has potential outliers.

This suggests that our proposed approach should be more adequate than the customary one based on the normal distribution.
## Parameter point estimates

Table 1: Summary statistics for the posterior distributions of the parameters for the data fitted under the student’s-t model for the 5% and 10% sample

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5% sample</th>
<th>10% sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td><strong>Student’s-t with independent prior</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>8.38</td>
<td>8.37</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.63</td>
<td>0.64</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>3.68</td>
<td>3.71</td>
</tr>
<tr>
<td>$\sigma_\nu$</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>$a$</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>$b$</td>
<td>2.31</td>
<td>2.28</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-</td>
<td>5.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5% sample</th>
<th>10% sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td><strong>Student’s-t with dependent prior</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>8.38</td>
<td>8.38</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.59</td>
<td>0.58</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>3.76</td>
<td>3.75</td>
</tr>
<tr>
<td>$\sigma_\nu$</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>$a$</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>$b$</td>
<td>2.30</td>
<td>2.28</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-</td>
<td>4.83</td>
</tr>
</tbody>
</table>
Some results

Table: Summary measurements of the point and interval estimation of the small area means for the income data fitted under the student’s-t and the normal models for the 10% sample and for the 5% sample.

<table>
<thead>
<tr>
<th>Model</th>
<th>5% sample AMSE</th>
<th>5% sample AARB (%)</th>
<th>10% sample AMSE</th>
<th>10% sample AARB (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>St-t with ind. prior</td>
<td>2.56</td>
<td>11.50</td>
<td>2.00</td>
<td>9.77</td>
</tr>
<tr>
<td>St-t with dep. prior</td>
<td>2.52</td>
<td>11.56</td>
<td>2.01</td>
<td>9.84</td>
</tr>
<tr>
<td>Normal</td>
<td>2.67</td>
<td>11.56</td>
<td>2.10</td>
<td>9.96</td>
</tr>
</tbody>
</table>
Some results

**Figure**: Plot of square error obtained in the student-t fit and the normal fit for the 5% and 10% samples. The student’s-t fit is presented with both priors.
We carry out a simulation study to evaluate the frequentist properties of the parameter estimators using our propose priors.

We generate 500 samples from the T-student model fixing the parameters as in the Table bellow.
Simulation Study Results

Table: Summary measurements for the point and interval estimation of the parameters for 500 samples generated under the student’s-t model fitted for a 5% sample.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>MSE</th>
<th>Coverage</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0 = 8$</td>
<td>8.01</td>
<td>8.01</td>
<td>0.03</td>
<td>0.94</td>
<td>0.68</td>
</tr>
<tr>
<td>$\beta_1 = 1$</td>
<td>0.99</td>
<td>0.99</td>
<td>0.15</td>
<td>0.94</td>
<td>1.48</td>
</tr>
<tr>
<td>$\beta_2 = 4$</td>
<td>4.02</td>
<td>4.02</td>
<td>0.52</td>
<td>0.94</td>
<td>2.83</td>
</tr>
<tr>
<td>$\sigma_\nu = 1$</td>
<td>0.96</td>
<td>0.97</td>
<td>0.24</td>
<td>0.94</td>
<td>1.03</td>
</tr>
<tr>
<td>$a = 1$</td>
<td>1.05</td>
<td>1.04</td>
<td>0.02</td>
<td>0.94</td>
<td>0.49</td>
</tr>
<tr>
<td>$b = 2$</td>
<td>2.06</td>
<td>2.08</td>
<td>0.17</td>
<td>0.95</td>
<td>1.42</td>
</tr>
<tr>
<td>$\alpha = 6$</td>
<td>-</td>
<td>5.64</td>
<td>-</td>
<td>0.98</td>
<td>47.14</td>
</tr>
</tbody>
</table>

Table: Summary measurements of the point and interval estimation of the small area means for the 500 samples generated under the student’s-t model fitted for a 5% sample.

<table>
<thead>
<tr>
<th>AMSE</th>
<th>ARE (%)</th>
<th>Coverage (%)</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.22</td>
<td>0.08</td>
<td>0.93</td>
<td>3.99</td>
</tr>
</tbody>
</table>
Figure: Boxplots of the percentage of times that the 95% credible intervals, produced by fitting student’s-t model, cover the true small area means for the 500 samples, the respective widths, the mean squared error and the relative absolute bias obtained in the fit with the 500 samples.
Concluding Remarks and Future Work

The evaluation studies with real data show that the t-Student area model are superior to the customary employed Normal area model when data has potential outliers.

As far as this simulation study is concerned, the model parameters are properly estimated.

However, further simulation study with smaller area sample size should be carried out to assess the frequent properties of our approach.

Fully simulation study will be carried out to assess the small area estimation procedure under different settings.

We intend to apply our approach to the unit level model.

Extensions to Skew-t models are also in progress (extension of [Ferraz and Moura(2012)]).

Another interest issue is to examine the implications of the proposal approach when dealing with more complex sample designs than simple random sampling (informative sampling).


