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

The Problem

- Hierarchical linear normal models are widely used to borrow strength from "exchangeable groups" in various stage of the hierarchy
- However it does not allow for the presence of atypical cases.
- This happens cause 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.

Our approach

- 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

$$
\begin{aligned}
y_i \mid \mu_i, v_i^{-1} &\sim N(\mu_i, v_i^{-1}) \text{ where } v_i^{-1} = \frac{\sigma_i^2}{n_i} \\
s_i^2 \mid n_i, \sigma_i^2 &\sim Ga\left\{0.5(n_i - 1), 0.5(n_i - 1)\sigma_i^{-2}\right\}, \quad \text{for } i = 1, ..., m,
\end{aligned} \tag{1}
$$

- Second level

$$
\begin{aligned}
\mu_i \mid \alpha, \beta, \sigma_\nu &\sim T\left(x_i^T \beta, \sigma_\nu^2, \alpha\right), \quad \text{or} \\
\mu_i &= x_i^T \beta + \delta_i \quad \text{where} \quad \delta_i \sim T(0, \sigma_\nu^2, \alpha) \quad \text{and}
\end{aligned} \tag{2}
$$

$$
v_i \mid a, b \sim Ga(a, b)
$$

Approximate Objective Priors

- Objective Bayesian analysis has a strong appeal when prior information is absent.
- Furthermore, it is particular 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.

Approximate Objective Priors

- The aim is to obtain an approximate objective priors for the hyperpameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_z, \nu, a, b)$.
- Let $y = (y_1, ..., y_m)^T$ and $s^2 = (s_1^2, ..., s_m^2)^T$ be the bivariate responses of the model and $z = (\delta_1, ..., \delta_m, v_1, ..., v_m)^T$ be the random latent variables.
- Thus the logarithm of the extended likelihood is given by :

$$\begin{aligned} l(\boldsymbol{\theta}, z) &= log[f(y, s^2|\boldsymbol{\theta}, z)] \\ &= \sum_{i=1}^{m} log[f(y_i|x_i, \boldsymbol{\beta}, \nu_i, v_i) f(s_i^2|v_i) f(v_i|a, b) f(\delta_i|\sigma_\delta, \nu)] \end{aligned}$$

Approximate Objective Priors

- Thus the approach consists in calculating $I(\boldsymbol{\theta})$ as equal to $E_{(y,s,z)}\left(-\frac{\partial l(\boldsymbol{\theta},z)}{\partial\boldsymbol{\theta}}\right)$, 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 $I(\boldsymbol{\theta})$ is block-diagonal with the following structure:

$$I(\boldsymbol{\theta}) = \begin{bmatrix} I_{\boldsymbol{\beta}} & O_{p\times 2} & O_{p\times 2} \\ O_{2\times 1} & I_{\sigma_\delta,\nu} & O_{4\times 4} \\ O_{2\times 1} & O_{4\times 4} & I_{a,b} \end{bmatrix}$$

where:

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

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

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

Approximate Objective Priors

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

$$
\begin{aligned}
p(\boldsymbol{\beta}, \sigma_\delta, a, b) &\propto |I(\boldsymbol{\theta})|^{1/2} = \left\{ |I_{\boldsymbol{\beta}}||I_{\sigma_\delta,\nu}||I_{a,b}| \right\}^{1/2} \\
&\propto a^{-p/2} b^{\frac{p-2}{2}} (a\psi'(a) - 1)^{1/2} \\
&\quad \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{ \psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}
\end{aligned}
$$

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

$$
\begin{aligned}
p_I(\boldsymbol{\beta}, \sigma_\delta, a, b) &\propto b^{-1} (a\psi'(a) - 1)^{1/2} \\
&\quad \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{ \psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}
\end{aligned}
$$

Approximate Objective Priors

- 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).

Aplications

- 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.

Aplication

- 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

	5% sample				10% sample			
	Mean	Median	Std	$CI_{95\%}$	Mean	Median	Std	$CI_{95\%}$
Student's-t with independent prior								
β_0	8.38	8.37	0.21	(7.99,8.80)	8.61	8.60	0.17	(8.29,8.94)
β_1	0.63	0.64	0.40	(-0.16,1.42)	0.81	0.79	0.35	(0.11,1.48)
β_2	3.68	3.71	0.71	(2.28,5.01)	3.38	3.37	0.61	(2.17,4.58)
σ_ν	0.87	0.87	0.25	(0.39,1.34)	0.98	0.99	0.20	(0.59,1.35)
a	1.08	1.08	0.13	(0.85,1.35)	1.24	1.23	0.15	(0.98,1.53)
b	2.31	2.28	0.39	(1.63,3.18)	1.52	1.51	0.23	(1.10,1.99)
α	-	5.82	17.42	(1.49,72.28)	-	7.31	26.07	(1.90,107.83)
Student's-t with dependent prior								
β_0	8.38	8.38	0.21	(7.97,8.75)	8.60	8.61	0.17	(8.28,8.94)
β_1	0.59	0.58	0.40	(-0.15,1.38)	0.78	0.78	0.35	(0.12,1.47)
β_2	3.76	3.75	0.74	(2.35,5.18)	3.40	3.38	0.64	(2.11,4.66)
σ_ν	0.85	0.85	0.25	(0.39,1.34)	0.98	0.98	0.21	(0.56,1.39)
a	1.07	1.07	0.13	(0.84,1.33)	1.22	1.22	0.14	(0.96,1.50)
b	2.30	2.28	0.38	(1.63,3.11)	1.51	1.50	0.22	(1.12,1.97)
α	-	4.83	9.46	(1.46,36.57)	-	7.73	23.23	(1.83,88.76)

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.*

	5% sample		10% sample	
Model	AMSE	AARB (%)	AMSE	AARB (%)
St-t with ind. prior	2.56	11.50	2.00	9.77
St-t with dep. prior	2.52	11.56	2.01	9.84
Normal	2.67	11.56	2.10	9.96

Some results

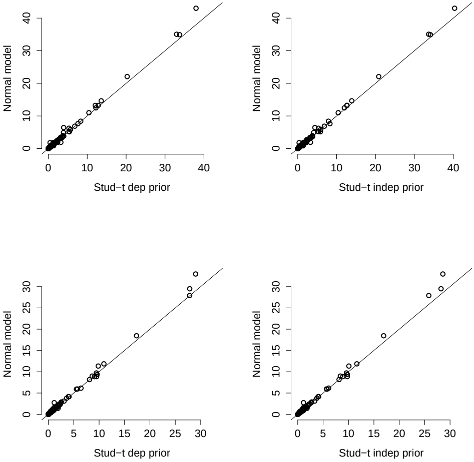


Figure : *Plot of square error obtained in the student-t fit and the normal fit for the 5% and 10% samples. The student's-t fit is presented with both priors.*

Simulation Study

- 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

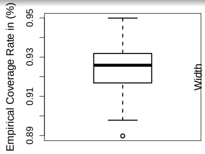
Table : *Summary measurements for the point and interval estimation of the parameters for 500 samples generated under the student's-t model fitted for a* 5% *sample.*

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

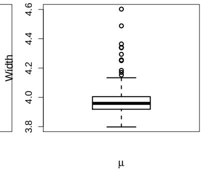
Table : *Summary measurements of the point and interval estimation of the small area means for the 500 samples generated under the student's-t model fitted for a* 5% *sample.*

AMSE	ARE (%)	Coverage (%)	Width
2.22	0.08	0.93	3.99

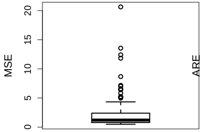
Simulation Study Results



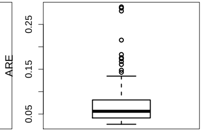
(a) Coverage



(b) Width



(c) MSE



(d) ARE

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).

- Bell, W. R. and Huang, E. T. (2006) Using the t-distribution to deal with outliers in small area estimation.
  In *Proceedings of Statistics Canada Symposium*.
- Berger, J. O., Bernardo, J. M. and Sun, D. (2009) The formal definition of reference priors.
  The Annals of Statistics, 905–938.
- Datta, G. S. and Lahiri, P. (1995) Robust hierarchical bayes estimation of small area characteristics in the presence of covariates and outliers.
  Journal of Multivariate Analysis, **54**, 310–328.
- Fabrizi, E. and Trivisano, C. (2010) Robust linear mixed models for small area estimation.
  Journal of Statistical Planning and Inference, **140**, 433–443.
- Ferraz, V. R. S. and Moura, F. A. S. (2012) Small area estimation using skew normal models.
  Computational Statistics and Data Analysis, **56**, 2864–2874.

- Fonseca, T., Ferreira, M. and Migon, H. (2008) Objective bayesian analysis for the student-t regression model.
  Biometrika, **95**, 325–333.

- Liseo, B., Tancredi, A. and Barbieri, M. (2010) Approximated reference priors in the presence of latent structure.
  In *Frontiers of statistical decision making and bayesian analysis, in honor of James O. Berger* (eds. P. M. D. S. M. Chen, D. K. Dey and K. Ye), 23–42. Springer, New York.

- Sun, D. and Berger, J. (1998) Reference priors with partial information.
  Biometrika, **85**, 55–71.