Predictors of Average Incomes under Area Level Gamma Mixed Model

Ondřej Faltys

Faculty of Nuclear Sciences and Physical Engineering

June 24, 2019

Small Area Estimation

- What is a small area?
- What is the essence of small area estimation (SAE)?
- What approaches are taken into account in SAE?
  - Design-based approach
  - Model-based approach
    - Unit level model
    - Area level model

Gamma Model

- Area level model
- Conditional distribution of the target variable y_d given v_d:

$$y_d|v_d \sim Gamma\left(\nu_d, a_d = \frac{\nu_d}{\mu_d}\right), \quad d = 1, \dots, D$$

- Conditional expectation and variance of the random variable y_d given v_d:

$$E[y_d|v_d] = \frac{\nu_d}{a_d} \equiv \mu_d, \quad \mathrm{var}[y_d|v_d] = \frac{\nu_d}{a_d^2} = \frac{\mu_d^2}{\nu_d}$$

Gamma Model

- Investigated model:

$$g(\mu_d) = \frac{1}{\mu_d} = \boldsymbol{x}_d^T \boldsymbol{\beta} + \phi v_d, \quad d = 1, \ldots, D$$

where $\beta = (\beta_1, \ldots, \beta_p)^T$ and $\boldsymbol{x}_d^T = (x_{d1}, \ldots, x_{dp})$

- Unexplained variability - random effects:

$$\boldsymbol{v} = (v_1, \ldots, v_D)^T \sim N_D(\mathbf{0}, \mathbf{I}_D)$$

- What is the distribution of $\boldsymbol{y} = (y_1, \ldots, y_D)^T$?

$$f(\boldsymbol{y}) = \int_{\mathbb{R}^D} f(\boldsymbol{y}|\boldsymbol{v}) f_{\boldsymbol{v}}(\boldsymbol{v}) \mathrm{d}\boldsymbol{v} = \int_{\mathbb{R}^D} \psi(\mathbf{y}, \mathbf{v}) \mathrm{d}\mathbf{v}$$

- Laplace approximation to the likelihood

Empirical Best Predictor

- Conditional expectation of y_d given v_d:

$$\mu_d \equiv E[y_d|v_d] = (\boldsymbol{x}_d^T\boldsymbol{\beta} + \phi v_d)^{-1}, \quad \mu_d = \mu_d(\boldsymbol{\theta}, v_d), \, \boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$$

- Best predictor (BP) of μ_d is $\hat{\mu}_d(\boldsymbol{\theta}) = E_\theta[\mu_d|y_d]$
- Empirical Best Predictor (EBP) of μ_d is $\hat{\mu}_d(\hat{\boldsymbol{\theta}})$

$$E_\theta[\mu_d|y_d] = \frac{\int_{\mathbb{R}}(\boldsymbol{x}_d^T\boldsymbol{\beta} + \phi v_d)^{-1} f(y_d|v_d) f(v_d) \mathrm{d}v_d}{\int_{\mathbb{R}} f(y_d|v_d) f(v_d) \mathrm{d}v_d}$$

- Plug-in predictor

$$\mu_d = \mu_d(\hat{\boldsymbol{\beta}}, \hat{v}_d) = (\boldsymbol{x}_d^T\hat{\boldsymbol{\beta}} + \hat{\phi}\hat{v}_d)^{-1}$$

Simulation Experiments

- **First task**: Validation of the gamma model
- For known $\theta \in \{\beta_0, \beta_1, \phi\}$, we compute

$$Mse = \frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta)^2$$

	$D = 26$	$D = 52$	$D = 104$
$\hat{\beta}_0$	6.04e-03	2.89e-03	1.55e-03
$\hat{\beta}_1$	8.27e-02	3.68e-02	1.66e-02
$\hat{\phi}$	1.69e-04	8.25e-05	4.57e-05

Table: MSE of parameter estimates computed by the Laplace approximation to the likelihood for $K = 1000$ repetitions.

Simulation Experiments

- **Second task**: investigating EBP and plug-in predictor
- For $d = 1, \dots, D$:

$$E_d = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_d^{(k)} - \mu_d^{(k)})^2, \quad E = \frac{1}{D}\sum_{d=1}^{D} E_d$$

	$D = 26$	$D = 52$	$D = 104$
Plug	3.17e-03	3.63e-03	3.36e-03
EBP	3.22e-03	3.70e-03	3.44e-03

Table: The average MSE of plug-in and EBP predictor, E, computed for $D = 26, 52, 104$ domains and for $K = 1000$ repetitions.

Simulation Experiments

- **Third task**: estimation of MSE of $mse^{\star}(\hat{\mu}_d) = E_d$
- Parametric bootstrap
- For $d = 1, \ldots, D$:

$$e_d = \frac{1}{K}\sum_{k=1}^{K}(mse_d^{\star(k)} - E_d)^2, \;\; e = \frac{1}{D}\sum_{d=1}^{D} e_d,$$

	$D = 26$	$D = 52$	$D = 104$
Plug	5.43e-07	5.32e-07	3.76e-07
EBP	5.72e-07	5.68e-07	4.22e-07

Table: Parametric bootstrap estimates for $D = 26, 52, 104$ domains and $K = 1000$ repetitions.

Application to real data

- Data for 104 Spanish domains are available
- Investigated model

$$g(\mu_d) = \frac{1}{\mu_d} = \boldsymbol{x}_d^T \boldsymbol{\beta} + \phi v_d, \quad d = 1, \ldots, D$$

- Fay-Herriot (1979) model

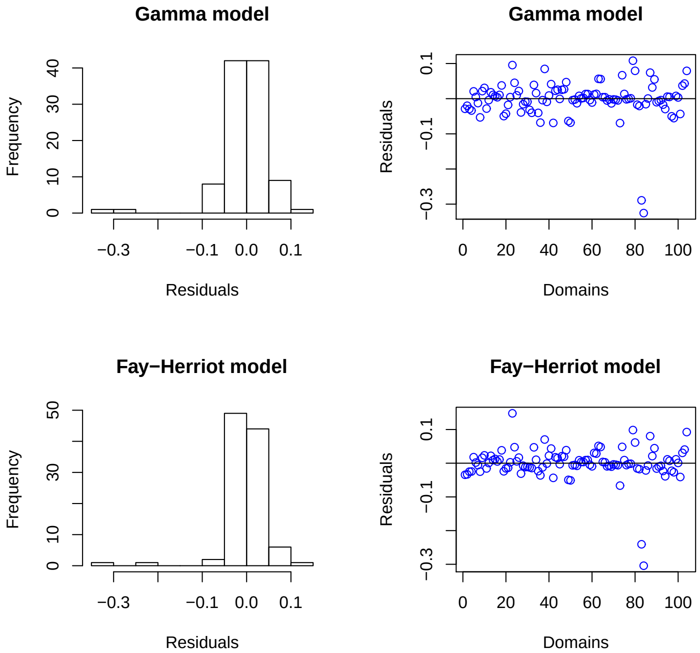
$$y_d = \boldsymbol{x}_d^T \boldsymbol{\beta} + \phi v_d + e_d,$$

- Employed auxiliary variables
  - Working status
  - Age
  - Education

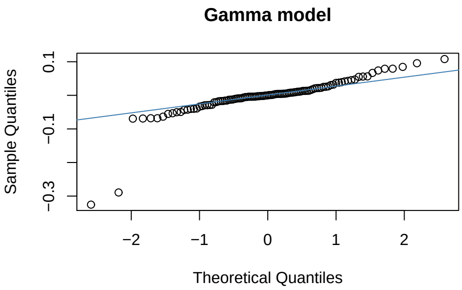
Gamma model				
	Estimate	Standard error	Wald (t-test) statistic	*p*-value
$\hat{\beta}_0$	1.58e-01	5.14e-02	9.39	2.18e-03
$\hat{\beta}_1$	1.90	2e-01	91.81	9.53e-22
$\hat{\beta}_2$	3.28	3.95e-01	68.89	1.04e-16
$\hat{\beta}_3$	1.43	8.14e-02	306.48	1.28e-68
$\hat{\beta}_4$	-9.3e-01	8.25e-02	127.14	1.73e-29
$\hat{\phi}$	5.87e-02	3.15e-03	348.67	8.27e-78

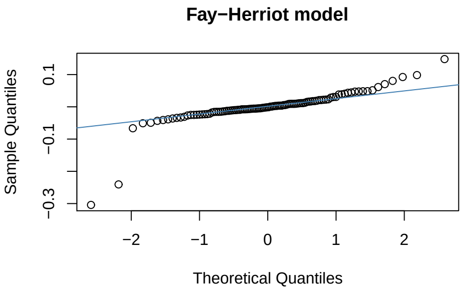
Fay-Herriot model				
	Estimate	Standard error	Wald (t-test) statistic	*p*-value
$\hat{\beta}_0$	2.44	2.49e-01	9.82	9.18e-23
$\hat{\beta}_1$	-3.71	8.72e-01	-4.25	2.16e-05
$\hat{\beta}_2$	-6.12	1.79	-3.43	6.11e-04
$\hat{\beta}_3$	-2.78	3.61e-01	-7.71	1.27e-14
$\hat{\beta}_4$	2.20	5.08e-01	4.33	1.47e-05
$\hat{\phi}$	1.19e-01			

Table: Regression parameter estimates for gamma model and Fay-Herriot model.



Gamma model
Frequency
0
10
20
30
40
−0.3
−0.1
0.0
0.1
Residuals
Gamma model
Residuals
0.1
−0.1
−0.3
0
20
40
60
80
100
Domains
Fay−Herriot model
Frequency
0
10
30
50
−0.3
−0.1
0.0
0.1
Residuals
Fay−Herriot model
Residuals
0.1
−0.1
−0.3
0
20
40
60
80
100
Domains


Gamma model
Sample Quantiles
0.1
−0.1
−0.3
−2
−1
0
1
2
Theoretical Quantiles


Fay−Herriot model
Sample Quantiles
0.1
−0.1
−0.3
−2
−1
0
1
2
Theoretical Quantiles

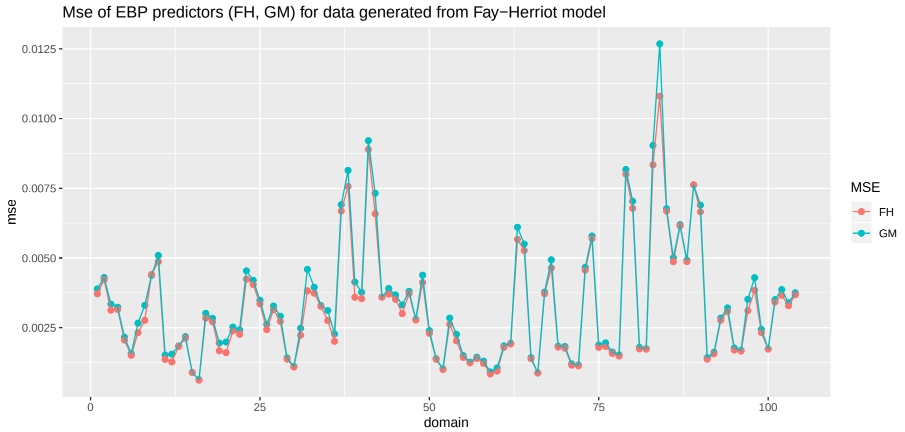
Sum of Squares

- Gamma model

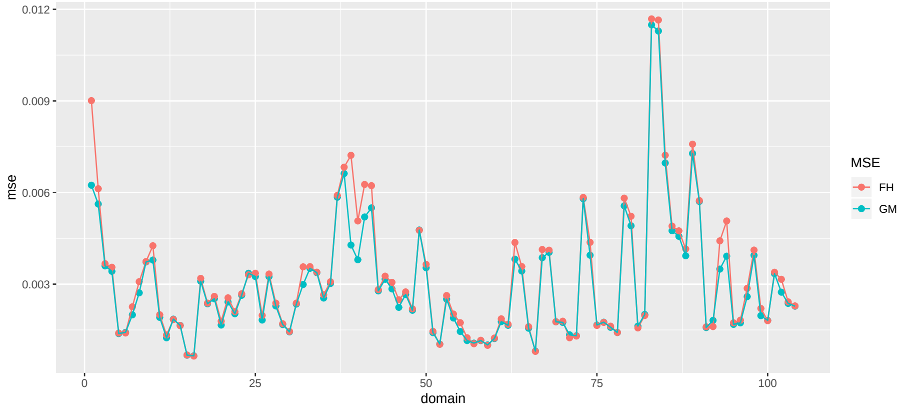
$$r^2 = \sum_{d=1}^{D} (y_d - \hat{\mu}_d^{(EBP)})^2 = 0.3315$$

- Fay-Herriot model

$$r^2 = \sum_{d=1}^{D} (y_d - \hat{\mu}_d^{(EBLUP)})^2 = 0.2615$$


Mse of EBP predictors (FH, GM) for data generated from Fay−Herriot model
0.0125
0.0100
0.0075
0.0050
0.0025
mse
0
25
50
75
100
domain
MSE
FH
GM

Mse of EBP predictors (FH, GM) for data generated from gamma model



Thank you for your attention.