Robust Small Area Estimation untuk Model Campuran Semiparametrik dengan Approksimasi *P-Spline*

Frida Murtinasari
Mahasiswa Pascasarjana Universitas Jember
fnopiyanto@gmail.com

Abstrak

Small Area Estimation merupakan tehnik pendugaan area kecil dengan menggunakan data pada domain yang besar. Untuk memperoleh cukup informasi dan mendapatkan nilai keragaman yang kecil, maka dibutuhkan peubah penjelas sebagai penunjang. Salah satu metode pendugaan parameter yang digunakan dalam *Small Area Estimation* adalah *Empirical Best Unbiased Linear*. Metode pendekatan EBLUP pada penelitian terdahulu digunakan pada *unit level model*. Pada unit level model diasumsikan bahwa penduga langsung memiliki hubungan yang linier dengan peubah penyertanya. Selain itu, EBLUP *unit level model* ini hanya berlaku dengan baik pada data yang diasumsikan normal dan tidak memiliki *outlier*. Akan tetapi tidak semua data memiliki asumsi yang normal maupun memiliki hubungan yang linier antar variabelnya. Oleh karena itu dibutuhkan sebuah pendekatan untuk *Small Area Estimation* dengan data yang tidak linier dan memiliki *outlier*. Penelitian ini mengembangkan *Robust Small Area Estimation* dengan menggunakan pendekatan model campuran semiparametrik dan approksimasi *P-Spline* untuk menduga rata-ratanya.

Kata kunci : P-Spline, semiparametrik, EBLUP, *Robust Small Area Estimation*

1 Pendahuluan

Small Area Estimation merupakan salah satu metode yang paling sering digunakan untuk menduga parameter pada area kecil. *Small Area Estimation* mengalami banyak perkembangan dan perbaikan berdasarkan jenis data sampel yang diperoleh. *Small Area Estimation* dapat dilakukan dengan memodelkan penduga langsung terhadap variabel penyerta. Terdapat dua pendekatan (Rao, 2003) yaitu *basic area level model* dan *basic unit level model* . Menurut Mukhopadhyay & Maiti (2004) , kedua model tersebut mengasumsikan bahwa penduga langsung memiliki hubungan yang linier dengan peubah penyertanya. Data yang diperoleh di lapangan, tidak semuanya memilki hubungan linier antara penduga langsung dengan variabel penyerta atau pendukungnya. Selain itu data yang digunakan tidak selalu berdistribusi normal. Ada kalanya data tersebut tidak dapat diasumsikan normal dan memiliki *outlier*. Oleh karena itu *Small Area Estimation* pada penelitian ini didekati dengan model campuran semi parametrik. Menurut Wibowo (2009), model ini lebih fleksibel daripara model linear karena keberadaan dua komponen parametrik dan non parametrik ini akan mengakomodasi hubungan antara respon dengan prediktor yang bersifat linear, dan hubungan antar respon dengan prediktor yang bersifat nonlinear. Pendugaan parameter pada level unit dengan EBLUP akan diapproksimasi dengan *P-Spline* untuk mendapatkan nilai ragam yang kecil dan lebih halus.

2 *Small Area Estimation* dengan Model Campuran Semiparametrik

Small Area Estimation merupakan pendugaan parameter pada subpopulasi yang lebih kecil dengan memanfaatkan memanfaatkan informasi tambahan yang akan mempunyai sifat "meminjam kekuatan" (*borrowing strength*) dari hubungan antara rataan area kecil dan informasi tambahan tersebut. Informasi tambahan tersebut diperoleh dari variabel penjelas atau penyertanya. Dalam pendugaan area kecil terdapat dua jenis model dasar yang digunakan, yaitu model berbasis area dan model berbasis unit (Rao 2003).

Basic area level model atau dapat disebut sebagai model berbasis area merupakan model yang didasarkan pada ketersediaan data pendukung atau penyerta yang hanya ada untuk level area tertentu, misalkan $x_i = \left(x_{1_i}, x_{2_i}, x_{3_i}, \ldots, x_{p_i}\right)^T$ dengan parameter yang akan diduga adalah θ_i yang merupakan fungsi dari rata-rata peubah respon dan diasumsikan mempunyai keterkaitan dengan x_i. Data pendukung atau penyerta tersebut digunakan untuk membangun model

$$\theta_i = x_i{}^T\beta + b_i v_i$$

dimana $i = 1, 2, 3, \ldots, p$ dan $v_i \sim N(0, \sigma_v^2)$.

Sedangkan *basic unit level* atau model berbasis unit merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $x_{ij} = \left(x_{ij_1}, x_{ij_2}, x_{ij_3}, \ldots, x_{ij_p}\right)^T$ memiliki arti untuk masing-masing anggota populasi *j* berada dalam masing-masing area kecil *i*, namun terkadang cukup dengan rata-rata populasi $\hat{x}_i$ yang diketahui, didapatkan suatu model regresi tersarang sebagai berikut (Rao, 2003) :

$$y_{ij} = x_{ij}^T\beta + v_i + e_{ij}\,, i = 1,2,\ldots,m\,; j = 1,2,\ldots,N_i$$

dengan v_i merupakan komponen acak yang memiliki distribusi $v_i \sim N(0, \sigma_v^2)$ dan $e_{ij} \sim N(0, \sigma_e^2)$ yang acak saling bebas. Pengembangan model *Small Area Estimation* didasarkan pada model campuran sebagai berikut :

$$y_i = x_{ij}^T\beta + b_i v_i + e_i$$

Model campuran semiparametrik pada *small area estimation* didekati oleh dua moment dari respon y_{ij} yaitu y_{ij} dari rata-rata μ_i dari area kecil tersebut dan μ_i yang telah diduga. Sehingga diperoleh model semiparametrik sebagai berikut (Breslow and Clayton, 1993) :

$$z(\mu_{ij}) = x_{ij1}^T\beta + x_{ij2}^T v_i, v_i \sim N(0, \Sigma_v)$$

3 Estimasi EBLUP Model Campuran Semiparametrik dengan Pendekatan *P-Spline*

Terdapat tiga metode estimasi yang dapat digunakan untuk menduga parameter dalam *small area estimation*. Metode estimasi tersebut adalah *Empirical Bayes* (EB), *Hierarical Bayes* (HB) dan *Best Linear Unbiased Prediction* (EBLUP) . Metode yang digunakan pada pendugaan ini adalah *metode EBLUP*. Metode EBLUP merupakan suatu pendugaan parameter yang meminimumkan MSE diantara kelas-kelas pendugaan parameter linier tak bias lainnya. EBLUP dihasilkan dengan asumsi bahwa komponen ragam telah diketahui. Namun dalam prakteknya, komponen ragam tidak diketahui. Oleh karena itu, diperlukan pendugaan terhadap komponen ragam tersebut melalui data contoh. Metode EBLUP mensubtitusi komponen ragam yang tidak diketahui ini dengan penduganya . Misal jika diberikan $\theta = (\sigma_v^2, \sigma_e^2)^t$ dengan θ tidak diketahui maka EBLUP dari $\bar{Y}_i$ digantikan dengan penduga $\hat{\theta}$.

Apabila bentuk $m(x,\beta) = x^t\beta$ merupakan fungsi mean yang diketahui (fungsi parametrik mean), maka diasumsikan bahwa fungsi mean yang tidak diketahui yaitu $m_0(x)$ dapat diapproksimasi oleh P-Spline dengan spline polinomial truncated. P-Spline model untuk *small area estimation* dengan efek acak v_i dan menyertakan EBLUP dari $\bar{Y}_i$ adalah sebagai berikut (Opsomer et al, 2008):

$$y_{ij} = x_{ij}^T\beta + w_{ij}^T u + v_i + e_{ij}, i = 1,2,\dots,m\ ; j = 1,2,\dots,N_i$$

dimana $x_{ij} = (1, x_{ij}, \dots, x_{ij}{}^h)^t$ dan

$w_{ij} = \{(x_{ij} - q_1)^h + \cdots (x_{ij} - q_k)^h\}^t = (w_{ij1}, \dots, w_{ijk})^t$.

4 *Robust Small Area Estimation* dengan EBLUP

Dalam EBLUP sangat mungkin untuk terjadi outlier dalam v_i dan e_{ij}. Pendugaan *Robust Small Area* level unit dirumuskan oleh Sinha dan Rao pada tahun 2009. Berdasarkan pendekatan P-Spline pada persamaan sebelumnya maka model P-Spline dalam bentuk vektor matrik dapat dituliskan sebagai berikut :

$$y = \mathbf{X}\beta + \mathbf{W}\mathrm{u} + \mathbf{Z}\mathbf{v} + \mathbf{e}$$

Maka algoritma Robust EBLUP dari u, v dan *robust estimator* yang didapat dari Fellner (1986) dan bersifat simultan, sebagai berikut :

1. Memberikan nilai awal untuk $\theta^{(0)}$ dari $\theta = (\sigma_u^2, \sigma_v^2, \sigma_e^2)^t$, yaitu dengan memasukkan $\beta^{(0)}$, $u^{(0)}$, $e^{(0)}$, dan $v^{(0)}$ maka diperoleh
$$y = \mathbf{X}\beta^{(0)} + \mathbf{W}\mathbf{u}^{(0)} + \mathbf{Z}\mathbf{v}^{(0)} + \mathbf{e}^{(0)}$$
2. Menggunakan bentuk estimasi di atas untuk mendapatkan $\theta^{(1)}$ dari θ
$$y = \mathbf{X}\beta^{(1)} + \mathbf{W}\mathbf{u}^{(1)} + \mathbf{Z}\mathbf{v}^{(1)} + \mathbf{e}^{(1)}$$
3. Masukkan pseudo-value
$$\tilde{y} = \mathbf{X}\beta + \mathbf{W}\mathrm{u} + \mathbf{Z}\mathbf{v} + \sigma_e \boldsymbol{\psi}(\sigma_e^{-1} e)$$
$$\tilde{0}_v = \mathbf{v} - \sigma_v \boldsymbol{\psi}(\sigma_v^{-1} v)$$
dimana $\boldsymbol{\psi}(\mathbf{u}) = (\psi(u_1), \psi(u_2), \dots)^t$ dengan menggunakan $\beta^{(0)}$, $u^{(0)}$, $e^{(0)}$, dan $v^{(0)}$ untuk β, **u, e,** dan **v** juga $\theta^{(1)}$ untuk θ
4. Menyelesaikan persamaan robust mixed model
$$\begin{bmatrix} \sigma_e^{-2}\mathbf{X}^t\mathbf{X} & \sigma_e^{-2}\mathbf{X}^t\mathbf{W} & \sigma_e^{-2}\mathbf{X}^t\mathbf{Z} \\ \sigma_e^{-2}\mathbf{W}^t\mathbf{X} & \boldsymbol{\sigma}_\mathbf{u}^{-2}\mathbf{I}_\mathbf{K} + \sigma_e^{-2}\mathbf{W}^t\mathbf{W} & \sigma_e^{-2}\mathbf{W}^t\mathbf{Z} \\ \sigma_e^{-2}\mathbf{Z}^t\mathbf{X} & \sigma_e^{-2}\mathbf{Z}^t\mathbf{W} & \boldsymbol{\sigma}_\mathbf{v}^{-2}\mathbf{I}_m + \sigma_e^{-2}\mathbf{Z}^t\mathbf{Z} \end{bmatrix} = \begin{bmatrix} \beta \\ \mathrm{u} \\ \mathrm{v} \end{bmatrix} = \begin{bmatrix} \sigma_e^{-2}\mathbf{X}^t\tilde{\mathrm{y}} \\ \sigma_e^{-2}\mathbf{W}^t\tilde{\mathrm{y}} \\ \sigma_e^{-2}\mathbf{Z}^t\tilde{\mathrm{y}} + \sigma_v^{-2}\tilde{0}_v \end{bmatrix},$$
Menggunakan $\theta^{(1)}$ dan pseudo-values pada langkah ke 3. Selanjutnya didapatkan nilai yang baru untuk $\beta^{(1)}$, $u^{(1)}$, dan $v^{(1)}$ dan menggunakan nilai tersebut untuk mendapatkan
$$\mathbf{e}^{(1)} = \tilde{\boldsymbol{y}}^{(1)} + \mathbf{X}\beta^{(1)} + \mathbf{W}\mathrm{u}^{(1)} + \mathbf{Z}\mathbf{v}^{(1)}$$

Langkah ini akan diulangi sampai nilai error yang konvergen.

Persamaan matriks di atas tereduksi pada persamaan model campuran ketika $\psi(\sigma_e^{-1}e) = \sigma_e^{-1}e$ dan $\psi(\sigma_v^{-1}v) = \sigma_v^{-1}v$ serta $\breve{0}_v$ mendekati 0. Bentuk persamaan REBLUP dari $\bar{Y}_i$ adalah

$$\hat{\mu}_{iF} = \frac{1}{N_i}\left(\sum_{j\epsilon s_i} y_{ij} + \sum_{j\epsilon s_i} \hat{y}_{ijF}\right)$$

dimana $\hat{y}_{ijF} = \mathrm{x}_{ij}^T\hat{\beta}_F + \mathrm{w}_{ij}^T\hat{\mathrm{u}}_F + \hat{v}_{iF}$.

Untuk daerah *l* yang tidak memiliki sampel, hasil dari $\hat{Y}_i$ adalah

$$\hat{\mu}_{iF} = \frac{1}{N_i}\sum_{j=1}^{N_i} \hat{y}_{ijF} = \mathrm{x}_{ij}^T\hat{\beta}_F + \mathrm{w}_{ij}^T\hat{\beta}_F,$$

dengan $\hat{y}_{ijF} = \mathrm{x}_{ij}^T\hat{\beta}_F + \mathrm{w}_{ij}^T\hat{\mathrm{u}}_F\ , j = 1 \ldots N_i$ merupakan prediktor sintetik dari y_{ij}.

Daftar Pustaka

[1] Fellner, W.. 1986. Robust estimation of variance components. *Technometrics*, 28, 51-60.

[2] Rao JNK. 2003. Small Area Estimation. New York: John Wiley & Sons.

[3] Sinha, S. K. and Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, 37, 381-399.

[4] Wibowo , W. 2009. Metode Kuadrat Terkecil Untuk Estimasi Kurva Regresi Semiparametrik Spline. Prosiding Seminar Matematika dan Pendidikan Matematika FMIPA UNY 2009.