

Perbandingan *Mean Squared Error* (MSE) Metode Prasad-Rao dan Jiang-Lahiri-Wan Pada Pendugaan Area Kecil

Widiarti^{1*}, Rifa Rahma Pertiwi², & Agus Sutrisno³

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung
Jl. Soemarti Brojonegoro No.1 Rajabasa, Bandar Lampung 35141
E-mail: widiarti08@gmail.com

Abstrak – Metode *Empirical Best Linear Unbiased Prediction* (EBLUP) merupakan salah satu metode pendugaan area kecil yang digunakan pada data kontinu dengan mensubstitusikan komponen ragam yang tidak diketahui ke dalam penduga BLUP. Keakuratan penduga salah satunya dapat diperoleh dengan cara mengukur *Mean Squared Error* (MSE). Beberapa metode telah dikembangkan dalam pendugaan MSE EBLUP. Prasad dan Rao (1990) mengembangkan penduga bagi MSE dengan menggunakan ekspansi deret Taylor. Jiang-Lahiri-Wan (2002) mengembangkan penduga bagi MSE dengan menggunakan metode *Jackknife*. Dalam penelitian ini pendugaan MSE bagi EBLUP dilakukan dengan kedua metode tersebut. Pendugaan MSE dilakukan secara empiris melalui data simulasi yang berdistribusi Normal dengan bantuan *software R 3.3.3*. Hasil penelitian ini menunjukkan bahwa metode pendugaan MSE EBLUP dengan metode Jiang-Lahiri-Wan relatif lebih baik karena menghasilkan nilai yang lebih kecil dibanding MSE Prasad dan Rao dan besarnya nilai ragam pengaruh acak menyebabkan nilai MSE yang dihasilkan semakin besar.

Kata kunci: *Pendugaan Area Kecil; Empirical Best Linear Unbiased Prediction (EBLUP); Mean Squared Error (MSE)*

1 Pendahuluan

Suatu area disebut kecil apabila contoh yang diambil pada area tersebut tidak mencukupi untuk melakukan pendugaan langsung dengan hasil yang akurat. Pendekatan klasik untuk menduga parameter area kecil didasarkan pada aplikasi model desain penarikan sampel (*design-based*) yang dikenal sebagai pendugaan langsung (*direct estimation*). Dalam konteks survei, penduga dikatakan langsung apabila pendugaan terhadap parameter populasi di suatu area hanya didasarkan pada data sampel yang diperoleh dari area tersebut. Pendugaan langsung pada suatu area kecil merupakan penduga tak bias tetapi memiliki ragam yang besar karena diperoleh dari ukuran sampel yang kecil [1].

Pendugaan tidak langsung (*indirect estimation*) merupakan salah satu upaya untuk menekan ragam yang besar pada area kecil yaitu dengan memanfaatkan informasi dari area sekitarnya yang berhubungan dengan

parameter yang diamati. Pendugaan tidak langsung tersebut dikenal sebagai pendugaan area kecil atau lebih dikenal dengan *Small Area Estimation* (SAE). Berbagai metode pendugaan area kecil (*small area estimation*) telah dikembangkan khususnya menyangkut metode yang berbasis model (*model-based estimator*). Beberapa metode yang tergolong dalam metode berbasis model adalah metode *Empirical Bayes* (EB), *Hierarchical Bayes* (HB), dan *Empirical Best Linear Unbiased Prediction* (EBLUP). Metode EB dan HB digunakan untuk data biner atau cacahan sedangkan metode EBLUP digunakan pada data kontinu.

Metode EBLUP merupakan perluasan dari metode BLUP. Pada metode BLUP diasumsikan komponen ragam dari pengaruh acak diketahui. Namun dalam kenyataannya, komponen ragam sulit untuk diketahui

sehingga diperlukan pendugaan terhadap komponen ragam melalui data sampel. Metode EBLUP mensubstitusikan komponen ragam yang tidak diketahui ke dalam penduga BLUP.

Keakuratan penduga dapat diperoleh dengan cara mengukur *mean squared error*-nya. Semakin kecil *mean squared error* suatu penduga maka penduga semakin akurat. Beberapa metode telah dikembangkan dalam pendugaan MSE ($\hat{\theta}^{EBLUP}$). Prasad dan Rao (1990) [2] mengembangkan penduga bagi MSE ($\hat{\theta}^{EBLUP}$) dengan menggunakan ekspansi deret Taylor. Jiang-Lahiri-Wan (2002) [3] mengembangkan penduga bagi MSE ($\hat{\theta}^{EBLUP}$) dengan menggunakan metode *jackknife*. *Jackknife* merupakan suatu teknik *resampling* yang secara khusus digunakan untuk menentukan ragam dan bias dugaan. Prinsip metode *Jackknife* adalah dengan cara menghilangkan satu buah data dan mengulanginya sebanyak jumlah yang ada. Pada penelitian ini penulis tertarik untuk membandingkan nilai dugaan MSE pada penduga EBLUP menggunakan metode yang dikembangkan Prasad dan Rao dengan metode yang dikembangkan Jiang-Lahiri-Wan dengan mengikutsertakan peubah penyerta. Dalam penelitian ini juga akan dikaji apakah ada pengaruh besarnya ragam pengaruh acak atau ragam area terhadap besarnya MSE. Perbandingan dilakukan secara empiris melalui data simulasi dengan bantuan *software R 3.3.3* dengan sebaran data berdistribusi normal.

2 Landasan Teori

Model area kecil merupakan model dasar dalam pendugaan area kecil. Dalam pendugaan area kecil terdapat dua jenis model dasar yang digunakan [1] yaitu :

2. 1. Basic Area Level (Type A) Model

Basic Area Level Model atau dapat disebut sebagai model berbasis area merupakan model yang didasarkan pada ketersediaan data pendukung yang hanya ada untuk level area tertentu, yaitu $\mathbf{x}_i^T = (x_{1i}, x_{2i}, \dots, x_{pi})$. Parameter *small area* yang ingin diamati adalah θ_i . Parameter *small area* ini berhubungan linear dengan \mathbf{x}_i^T mengikuti model linear berikut :

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + b_i v_i, \quad i = 1, 2, \dots, m \quad (1)$$

dengan $v_i \sim N(0, A)$ sebagai pengaruh acak yang diasumsikan menyebar normal, sedangkan b_i merupakan konstanta positif yang diketahui dan $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ adalah vektor koefisien regresi berukuran $p \times 1$. Kesimpulan mengenai θ_i dapat diketahui dengan mengasumsikan bahwa model penduga langsung y_i telah tersedia, yaitu :

$$y_i = \theta_i + e_i, \quad i = 1, 2, \dots, m \quad (2)$$

dengan *sampling error* $e_i \sim N(0, D_i)$ dan D_i diketahui. Dari kombinasi persamaan (2.1) dan (2.2) didapatkan model gabungan :

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + b_i v_i + e_i, \quad i = 1, 2, \dots, m \quad (3)$$

dengan asumsi v_i dan e_i saling bebas. Rao (2003) [1] menyatakan bahwa model tersebut merupakan bentuk khusus dari model linear campuran.

2. 2. Basic Unit Level (Type A) Model

Basic Unit Level Model atau model berbasis unit merupakan suatu model dimana data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ artinya untuk masing-masing anggota populasi j dalam masing-masing area kecil i , namun terkadang cukup dengan rata-rata populasi \bar{X}_i diketahui saja.

Model yang digunakan pada penelitian ini adalah model Fay-Herriot. Model ini diperkenalkan oleh Fay dan Herriot (1979) [4] sebagai model dasar untuk menaksir pendapatan per kapita pada *small area* (dengan populasi yang kurang dari 1.000 jiwa penduduk) di Amerika Serikat menggunakan model dua level berikut :

$$\text{Level 1 : } y_i | \theta_i \sim N(\theta_i, D_i)$$

$$\text{Level 2 : } \theta_i \sim N(\bar{X}_i^T \boldsymbol{\beta}, A)$$

Model dua level diatas dapat dituliskan sebagai model linear campuran :

$$y_i = \theta_i + e_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, 2, \dots, m \quad (4)$$

Model Fay-Herriot ini merupakan kasus model area level seperti pada persamaan (3) dengan $b_i = 1$, dimana :

y_i : nilai pendugaan langsung berdasarkan rancangan survei

\mathbf{x}_i : vektor variabel pendukung yang elemennya diketahui

$\boldsymbol{\beta}$: vektor parameter berukuran $p \times 1$

v_i : pengaruh acak area kecil dengan asumsi $v_i \sim N(0, A)$

e_i : *sampling error* yang tidak terobservasi dengan asumsi $e_i \sim N(0, D_i)$

Menurut Rao (2003) [1] penduga BLUP yang terbentuk bagi $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i$ adalah :

$$\hat{\theta}_i^{BLUP} = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \hat{v}_i = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \gamma_i (y_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}) \quad (5)$$

dengan

$$\gamma_i = \frac{A}{A + D_i}$$

$\tilde{\boldsymbol{\beta}}$ dapat diperoleh dengan metode *generalized least square* sehingga diperoleh

$$\tilde{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

3

dimana $X = x_i^T$, $V = D_i + A$

Penduga BLUP pada persamaan (2) masih bergantung pada komponen ragam A yang pada prakteknya tidak diketahui nilainya, sehingga harus ditaksir dari data. Dengan mensubstitusi \hat{A} ke A pada persamaan (2) diperoleh penduga EBLUP bagi $\theta_i = x_i^T \beta + v_i$ sebagai berikut :

$$\hat{\theta}_i^{EBLUP} = x_i^T \hat{\beta} + \hat{\gamma}_i (y_i - x_i^T \hat{\beta}) \quad (6)$$

dengan

$$\hat{\gamma}_i = \frac{\hat{A}}{\hat{A} + D_i}, \text{ untuk } i = 1, 2, \dots, m.$$

dan

$$\hat{\beta} = \left(\sum_{i=1}^m \frac{x_i x_i^T}{(D_i + \hat{A})} \right)^{-1} \left(\sum_{i=1}^m \frac{x_i y_i}{(D_i + \hat{A})} \right) \quad (7)$$

Dimana $\hat{\gamma}_i$ dan $\hat{\beta}$ merupakan nilai γ_i dan β saat A disubstitusikan dengan nilai dugaannya \hat{A} . Menurut Wan (1999) [5] nilai \hat{A} tersebut dapat diperoleh dengan metode *moment* yaitu $\hat{A} = \max(0, \hat{A})$

Dimana

$$\hat{A} = \frac{1}{m-p} \sum_{i=1}^m \left[(y_i - x_i^T \hat{\beta}_{OLS})^2 - D_i(1 - h_i) \right] \quad (8)$$

5

Dengan

$$h_i = x_i^T \left(\sum_{i=1}^m x_i x_i^T \right)^{-1} x_i$$

Dan

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^m x_i x_i^T \right)^{-1} \left(\sum_{i=1}^m x_i y_i \right)$$

2. 3. Mean Squared Error EBLUP

Prasad dan Rao (1990) [2] menggunakan ekspansi deret Taylor untuk menduga *mean squared error* EBLUP sehingga diperoleh sebagai berikut:

$$MSE^{PR}(\hat{\theta}_i^{EBLUP}) = g1i(\hat{A}) + g2i(\hat{A}) + 2g3i(\hat{A}) \quad (9)$$

Dengan

$$g1i(\hat{A}) = \frac{\hat{A} D_i}{\hat{A} + D_i},$$

$$g2i(\hat{A}) = \left(1 - \frac{\hat{A}}{\hat{A} + D_i} \right)^2 x_i^T \left(\frac{\hat{A} + D_i}{x_i x_i^T} \right) x_i,$$

$$g3i(\hat{A}) = \frac{2D_i^2}{m(\hat{A} + D_i)}$$

Sementara Jiang-Lahiri-Wan (2002) [3] menggunakan konsep *jackknife* untuk mengoreksi bias dugaannya sehingga diperoleh sebagai berikut:

$$MSE^{JLW}(\hat{\theta}_i^{EBLUP}) = g1i(\hat{A}) - \frac{m-1}{m} \sum_{u=1}^m \left(g1i(\hat{A}_{-u}) - g1i(\hat{A}) \right) + \frac{m-1}{m} \sum_{u=1}^m \left(\hat{\theta}(y_i; \hat{A}_{-u}) - \hat{\theta}(y_i; \hat{A}) \right)^2 \quad (10)$$

Dimana

$$\hat{\theta}(y_i; \hat{A}) = \hat{\theta}_i^{EBLUP} = x_i^T \hat{\beta} + \frac{\hat{A}}{\hat{A} + D_i} (y_i - x_i^T \hat{\beta})$$

$$\hat{\theta}(y_i; \hat{A}_{-u}) = x_i^T \hat{\beta}_{-u} + \frac{\hat{A}_{-u}}{\hat{A}_{-u} + D_i} (y_i - x_i^T \hat{\beta}_{-u})$$

6

$$g1i(\hat{A}) = \frac{\hat{A} D_i}{\hat{A} + D_i}$$

$$g1i(\hat{A}_{-u}) = \frac{\hat{A}_{-u} D_i}{\hat{A}_{-u} + D_i}$$

\hat{A}_{-u} dan $\hat{\beta}_{-u}$ merupakan penduga \hat{A} dan $\hat{\beta}$ setelah menghapus data area ke-u.

Sehingga pada $MSE^{JLW}(\hat{\theta}_i^{EBLUP})$ ini, untuk mendapatkan nilai $\hat{\theta}(y_i; \hat{A}_{-u})$ harus dilakukan pendugaan ulang terhadap \hat{A}_{-u} dan $\hat{\beta}_{-u}$ untuk setiap area ke-u yang dihapus sejumlah banyaknya area.

3 Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah berupa data simulasi dengan bantuan *software R 3.3.3* dan sebaran data berdistribusi normal. Pada kajian ini penulis menetapkan nilai A dan D_i serta banyaknya area m berdasarkan penelitian sebelumnya dimana Jiang, Lahiri dan Wan (2002) [3] dalam Rao (2003) [1] menetapkan nilai $D_i = 1$ dan banyaknya area $m = 30, 60$ dan 90 . Untuk melihat apakah ada pengaruh besarnya ragam pengaruh acak atau ragam area terhadap perolehan hasil MSE digunakan tiga nilai A yang berbeda yaitu $A = 1, 5$ dan 10 . Metode pendugaan MSE yang digunakan dalam penelitian ini yaitu *mean squared error* Prasad dan Rao (1990) [2] yang dikembangkan dengan aproksimasi deret Taylor serta Jiang-Lahiri-Wan (2002) [3] dengan konsep *jackknife*.

Simulasi pada software R

1. Membangkitkan peubah acak x_i sebagai peubah penyerta bagi variabel respon y_i , digunakan empat peubah penyerta sebagai berikut :

$$x_1 \sim N(1480, 387158)$$

$$x_2 \sim N(721.8, 69525.7) \quad x_3 \sim N(14691, 53264948)$$

$$x_4 \sim N(20.76, 40.69)$$

Nilai tengah dan ragam masing masing variabel x tersebut diambil dari nilai tengah dan ragam variabel penyerta yang memengaruhi pengeluaran per kapita setiap kecamatan di Kabupaten Brebes tahun 2013 dimana pada penelitian sebelumnya Ningtyas *et al* (2015) [6] menggunakan data jumlah kelahiran penduduk (x_1), jumlah kematian penduduk (x_2), jumlah penduduk yang memiliki kendaraan roda 2 (x_3) dan jumlah sarana kesehatan (puskesmas, poliklinik kesehatan desa, balai pengobatan, rumah sakit khusus, rumah bersalin dan rumah sakit umum) (x_4). Masing masing dibangkitkan sejumlah banyaknya area yaitu 30, 60 dan 90.

2. Membangkitkan data $e_i = (e_1, e_2, \dots, e_m)$ sebagai *sampling error* dengan $e_i \sim N(0,1)$
3. Menetapkan nilai β , pada penelitian ini digunakan empat nilai β yaitu -0.000380 , 0.001278 , 0.000122 , -0.022920 yang diperoleh dari penelitian sebelumnya oleh Ningtyas *et al* (2015) [6] menggunakan data pengeluaran per kapita setiap kecamatan di Kabupaten Brebes tahun 2013.
4. Membangkitkan data $\theta_i \sim N(x_i^T \beta, A)$ sejumlah banyaknya area yaitu 30, 60 dan 90 juga untuk masing-masing nilai A=1, 5 dan 10.
5. Memperoleh data y_i sebagai nilai pendugaan langsung dimana $y_i = \theta_i + e_i$
6. Menghitung nilai dugaan ragam pengaruh acak (\hat{A})
7. Menghitung nilai dugaan *mean squared error* Prasad dan Rao (1990) [2] $MSE^{PR}(\hat{\theta}_i^{EBLUP})$ dengan mensubstitusikan hasil dugaan ragam pengaruh acak (\hat{A}) (langkah 6) ke persamaan (9). Kemudian menghitung rata-rata MSE Prasad dan Rao untuk setiap area (m=30, 60 dan 90) dan ragam pengaruh acak yang berbeda (A=1,5 dan 10).
8. Menghitung nilai dugaan *mean squared error* Jiang-Lahiri-Wan (2002) [3] $MSE^{JLW}(\hat{\theta}_i^{EBLUP})$ dengan persamaan (10) dengan langkah sebagai berikut :
 - a. Menghitung $\hat{\beta}$ dengan persamaan (7)
 - b. Menghitung $\hat{\theta}(y_i; \hat{A}) = \hat{\theta}_i^{EBLUP}$ dengan mensubstitusikan hasil \hat{A} dan $\hat{\beta}$ pada langkah (6) dan (8.a) ke persamaan (6)
 - c. Menghitung nilai dugaan ragam pengaruh acak untuk setiap area ke-u yang dihapus (\hat{A}_{-u})
 - d. Menghitung $\hat{\beta}$ untuk setiap area ke-u yang dihapus ($\hat{\beta}_{-u}$)
 - e. Menghitung $\hat{\theta}(y_i; \hat{A}_{-u})$ dengan mensubsstitusikan hasil \hat{A}_{-u} dan $\hat{\beta}_{-u}$
 - f. Mensubstitusikan hasil \hat{A} , \hat{A}_{-u} , $\hat{\theta}(y_i; \hat{A})$ dan $\hat{\theta}(y_i; \hat{A}_{-u})$ ke persamaan (10) sehingga diperoleh hasil dugaan *mean squared error* Jiang-Lahiri-Wan $MSE^{JLW}(\hat{\theta}_i^{EBLUP})$ kemudian menghitung rata-rata MSE untuk setiap area (m=30, 60 dan 90) dan

ragam pengaruh acak yang berbeda (A=1, 5 dan 10).

9. Membandingkan hasil (7) dan (8)

4 Hasil dan Pembahasan

Seperti yang telah diuraikan sebelumnya, bahwa untuk membandingkan nilai MSE Prasad dan Rao dengan MSE Jiang-Lahiri-Wan, penulis menggunakan data simulasi yang dibangkitkan dengan *software R* dengan sebaran data berdistribusi normal.

Nilai MSE Prasad dan Rao, diperoleh dengan terlebih dahulu mencari nilai dugaan bagi ragam pengaruh acak (\hat{A}). Sementara MSE Jiang-Lahiri-Wan tergolong lebih rumit, karena setelah didapatkan nilai dugaan bagi ragam pengaruh acak (\hat{A}) selanjutnya menghitung $\hat{\beta}$ dan mensubstitusikan nya ke $\hat{\theta}(y_i; \hat{A}) = \hat{\theta}_i^{EBLUP}$. Setelah itu dihitung nilai \hat{A}_{-u} dan $\hat{\beta}_{-u}$ yaitu nilai \hat{A} dan $\hat{\beta}$ untuk setiap area ke-u yang dihapus untuk mendapatkan nilai $\hat{\theta}(y_i; \hat{A}_{-u})$. Hasil tersebut kemudian disubstitusikan ke persamaan (10) sehingga diperoleh MSE Jiang-Lahiri-Wan.

Setelah diperoleh nilai MSE untuk masing-masing jumlah area (m= 30, 60, dan 90) dan ragam pengaruh acak atau ragam area (A= 1, 5, dan 10) kemudian dihitung rata-rata nilai MSE kedua metode tersebut. Hasil rata-rata MSE tersaji pada Tabel 1 berikut :

Tabel 1 Rata - rata Perolehan MSE Prasad-Rao dan Jiang-Lahiri-Wan

| Area (m) | Rata-Rata MSE | | | | | |
|----------|---------------|-----------|----------|-----------|----------|-----------|
| | A=1 | | A=5 | | A=10 | |
| | PR | JLW | PR | JLW | PR | JLW |
| 30 | 0,9151977 | 0,7295627 | 4,871824 | 0,9872269 | 15,93285 | 0,9968345 |
| 60 | 0,6435772 | 0,5298622 | 2,564093 | 0,963263 | 7,076253 | 0,9903925 |
| 90 | 0,6250407 | 0,5681524 | 2,348756 | 0,9713256 | 6,507008 | 0,9927885 |

Berdasarkan Tabel 1 terlihat bahwa untuk masing-masing jumlah area (m) dan ragam pengaruh acak (A), MSE dengan metode Jiang-Lahiri-Wan menghasilkan nilai yang lebih kecil dibandingkan Prasad-Rao. Selain itu terlihat pula bahwa semakin banyak jumlah area rata-rata nilai MSE yang dihasilkan dari kedua metode tersebut pun relatif semakin kecil.

Rata-rata MSE yang diperoleh baik dengan metode Prasad-Rao maupun Jiang-Lahiri-Wan untuk masing-masing nilai ragam pengaruh acak atau ragam area (A) menghasilkan nilai yang semakin besar. Ini menunjukkan bahwa nilai ragam pengaruh acak atau ragam area berpengaruh terhadap perolehan MSE, dimana semakin besar nilai ragam pengaruh acak maka semakin besar pula nilai MSE yang dihasilkan.

5 Simpulan dan Saran

Berdasarkan uraian sebelumnya, maka dapat disimpulkan bahwa metode pendugaan *mean squared error* EBLUP dengan metode Jiang-Lahiri-Wan lebih baik karena menghasilkan nilai yang lebih kecil dibandingkan MSE Prasad dan Rao. Semakin besar jumlah area (m) maka nilai MSE EBLUP dengan kedua metode relatif semakin kecil. Sebaliknya, semakin besar nilai ragam area (A) maka nilai MSE menjadi semakin besar.

Kepustakaan

- [1] Rao, J.N.K. 2003. *Small Area Estimation*. John Willey and Sons, Inc., New York.
- [2] Prasad, N.G.N., and Rao, J.N.K. 1990. The Estimation of Mean Squared Errors of Small Area. Estimators. *Journal of the American Statistical Association* . 85, 163-171.
- [3] Jiang, J., Lahiri, P., and Wan, S. 2002. A Unified Jackknife Method. *Annals of Statistics*. 30, 1782-1810.
- [4] Fay, R.E., and Herriot, R.A. 1979. Estimates of Income for Small Places: an Application of James-Stein Procedure to Cencus Data. *Journal of American Statistical Association*. 74, 269-277.
- [5] Wan, S. M. 1999. Jackknife Methods in Small Area Estimation and Related Problems. (Dissertation). University of Nebraska. Lincoln.
- [6] Ningtyas, R., Rahmawati, R., dan Wilandari, Y. 2015. Penerapan Metode Empirical Best Linear Unbiased Prediction (EBLUP) Pada Model Penduga Area Kecil Dalam Pendugaan Pengeluaran Per Kapita Di Kabupaten Brebes. *Jurnal Gaussian*. 4, 977-986.