

Received: 24 April 2022

Revised: 12 June 2022

Accepted: 29 June 2022

Published: 30 June 2022

Perbandingan Metode SAE EBLUP dan SAE HB Pada Pendugaan Area Kecil (Studi Kasus Pendugaan Kemiskinan di Provinsi Jawa Timur)

Widya Larasati^{1, a)}, Novia Permatasari^{2, b)}

^{1,2}*Badan Pusat Statistik*

Email: ^{a)} widya.larasati@bps.go.id, ^{b)} novia.permatasari@bps.go.id

Abstract

The demand for estimation at the small area level was increasing, while the success of estimation using an indirect method in reducing RSE was very dependent on the data condition and selection of the proper method. This study aimed to compare the results of estimating the percentage of poor people using the EBLUP and HB Beta methods in reducing RSE with direct estimation results using a case study of poverty data at the sub-district level of East Java Province. The data used are 2019 Susenas and 2019 Podes. The clustering technique was used to estimate several sub-districts not sampled in the 2019 Susenas. The prior distribution used in the HB method in this study was the Beta distribution. As a result, the EBLUP and HB methods could overcome the estimation of unsampled areas by utilizing cluster information. In addition, both SAE EBLUP and HB methods were able to reduce the average and median values of the estimated RSE compared to direct estimates. From 173 sub-districts due to direct estimation with RSE >50%, the EBLUP method could reduce it to 116 sub-districts. Meanwhile, the HB method reduced the RSE of all sub-districts to below 50%. So, it can be concluded that the HB method improved the direct estimation results better than the EBLUP method in estimating the percentage of poor people at the sub-district level of East Java Province in 2019.

Keywords: poverty, small area estimation, EBLUP, HB.

Abstrak

Permintaan estimasi pada level area kecil semakin meningkat, sedangkan keberhasilan pendugaan menggunakan metode tidak langsung dalam menurunkan RSE sangat bergantung pada kondisi data dan pemilihan metode yang tepat. Penelitian ini bertujuan untuk membandingkan hasil pendugaan persentase penduduk miskin menggunakan metode EBLUP dan HB Beta dalam menurunkan RSE hasil pendugaan langsung menggunakan studi kasus data kemiskinan level kecamatan Provinsi Jawa Timur. Data yang digunakan yaitu Susenas 2019 dan Podes 2019. Teknik *clustering* digunakan untuk mengestimasi beberapa kecamatan yang tidak ter sampel dalam Susenas 2019. Distribusi prior yang digunakan pada metode SAE HB di penelitian ini adalah distribusi Beta. Hasilnya, metode EBLUP dan HB dapat mengatasi pendugaan pada area yang tidak ter sampel dengan memanfaatkan informasi *cluster*. Selain itu, baik metode SAE EBLUP maupun HB mampu menurunkan nilai rata-rata dan median dari RSE

hasil estimasi jika dibandingkan dengan pendugaan langsung. Dari 173 kecamatan hasil pendugaan langsung yang memiliki RSE $>50\%$, metode EBLUP dapat mereduksinya menjadi 116 kecamatan. Sedangkan, metode HB mampu menurunkan RSE seluruh kecamatan menjadi di bawah 50% . Hal ini menunjukkan bahwa metode HB mampu memperbaiki hasil estimasi langsung lebih baik dari metode EBLUP pada pendugaan persentase penduduk miskin level kecamatan Provinsi Jawa Timur tahun 2019.

Kata-kata kunci: kemiskinan, *small area estimation*, EBLUP, HB.

PENDAHULUAN

Berkaitan dengan masalah kesejahteraan, kemiskinan menjadi persoalan yang dihadapi banyak negara-negara di dunia, termasuk Indonesia. Kemiskinan merupakan permasalahan yang kompleks karena mencakup multidimensi kehidupan. Urgensi masalah kemiskinan terlihat dari dicakupnya kemiskinan dalam tujuan pertama agenda *Sustainable Development Goals* (SDGs). Selain itu, upaya penanggulangan kemiskinan juga menjadi prioritas nasional pada setiap rencana kerja pemerintah, setidaknya dalam empat periode Rencana Pembangunan Jangka Menengah (RPJMN) (Bappenas, 2019). Pemerintah membutuhkan data yang akurat guna *monitoring* dan evaluasi program penanggulangan kemiskinan.

Data kemiskinan yang disediakan oleh BPS memiliki keterbatasan, yaitu hanya representatif pada level nasional, provinsi, dan kabupaten/kota akibat keterbatasan ukuran sampel. Ukuran sampel yang terlalu kecil pada level wilayah kecil membuat pengukuran kemiskinan secara langsung menghasilkan *standard error* yang besar sehingga statistik tersebut tidak dapat dipercaya (Ubaidillah, 2014). Menurut Mulia, dkk. (2007) dalam Kurniawan, dkk. (2020), nilai RSE yang dianggap akurat adalah RSE $\leq 25\%$, sedangkan estimasi dengan nilai RSE $>25\%$ tetapi $\leq 50\%$ perlu hati-hati jika ingin digunakan dan RSE $>50\%$ dianggap sangat tidak akurat. Di sisi lain, tuntutan akan data kemiskinan yang akurat dan tersedia hingga level wilayah yang lebih kecil (seperti kecamatan) semakin meningkat. Maka dari itu, dikembangkan metode *Small Area Estimation* (SAE) atau disebut juga pendugaan area kecil untuk mengakomodasi kebutuhan data pada level wilayah yang lebih rendah.

SAE merupakan salah satu upaya untuk meningkatkan keakuratan hasil estimasi menggunakan pendugaan tidak langsung (*indirect estimation*), yaitu dengan meminjam kekuatan daerah tetangga dan memanfaatkan hubungan antara variabel penjelas dan variabel yang diteliti (Rao (2003), Wardani (2008), Molina (2015), Ubaidillah, dkk. (2019), Permatasari & Ubaidillah (2021)). Di Indonesia, *small area estimation* telah diaplikasikan sejak tahun 2001 guna menghasilkan peta kemiskinan provinsi DKI Jakarta, Jawa Timur dan Kalimantan Timur oleh *Smeru Research Institute* yang berkolaborasi dengan BPS (Notodiputro & Kurnia, 2010). Pengembangan berbagai pendekatan metode SAE untuk menghitung kemiskinan wilayah kecil terus dilakukan dalam banyak penelitian. Beberapa metode yang bisa digunakan, yaitu *Empirical Bayes* (EB), *Empirical Best Linear Unbiased Predictions* (EBLUP), dan *Hierarchical Bayes* (HB) (Noviyanti & Zain (2014), Ubaidillah (2014), Ikhsan dkk. (2018), Yuliani dkk. (2019)). Penelitian ini akan membahas lebih lanjut mengenai metode SAE EBLUP dan HB dalam mengestimasi kemiskinan level kecamatan dengan menggunakan studi kasus data kemiskinan level kecamatan di Provinsi Jawa Timur, kemudian membandingkan efektivitas hasil pendugaan kedua metode tersebut dalam menurunkan RSE.

METODOLOGI

Bahan dan Data

Data yang digunakan pada penelitian ini yaitu data Susenas tahun 2019 dan Data Podes tahun 2019. Susenas merupakan survei dengan pendekatan rumah tangga yang mengumpulkan berbagai informasi seperti demografi, pendidikan, kesehatan, perumahan, pengeluaran untuk konsumsi, dll. Variabel respons pada penelitian ini, yaitu persentase penduduk miskin pada level kecamatan di Provinsi Jawa Timur, diperoleh dari data Susenas.

Di sisi lain, variabel penjelas diperoleh dari data Podes. Data Podes yang merupakan data sensus dengan pendekatan institusi desa/kelurahan. Variabel penjelas yang digunakan yaitu sebanyak 11 variabel dengan rincian sebagai berikut:

1. Proporsi desa yang terdapat akses internet
2. Proporsi desa yang terdapat sinyal HP
3. Rasio SD terhadap jumlah keluarga
4. Rasio SMP terhadap jumlah keluarga
5. Rasio Rumah Sakit terhadap jumlah keluarga
6. Rasio Poliklinik terhadap jumlah keluarga
7. Rasio Polindes terhadap jumlah keluarga
8. Rasio apotek terhadap jumlah keluarga
9. Rasio Rumah Sakit terhadap jumlah keluarga
10. Rasio kasus gizi buruk terhadap jumlah keluarga
11. Indeks Kesulitan Geografis (IKG) 2019

Metode Penelitian

Estimasi Langsung Indikator Kemiskinan

BPS menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*) untuk mengukur kemiskinan. Dengan pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Maka dari itu, penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan. Sumber data utama yang dipakai adalah data Susenas Modul Konsumsi dan Pengeluaran (BPS, 2021).

Garis kemiskinan merupakan batas pendapatan yang diperlukan untuk memenuhi kebutuhan minimal kalori yang diperlukan tubuh untuk beraktivitas, ditambah dengan kebutuhan non makanan (perumahan, pakaian, pendidikan, kesehatan, transpor, dan kebutuhan pokok lainnya). Karena data pendapatan tidak tersedia maka dipakai pendekatan data konsumsi/pengeluaran. Termasuk pengeluaran adalah perkiraan barang dan jasa yang dikonsumsi berasal dari hasil produksi sendiri dan pemberian dari pihak lain (BPS, 2021). Rumus penghitungan GK, sebagai berikut:

$$GK = GKM + GKNM \quad (1)$$

Keterangan:

GK = Garis Kemiskinan

GKM = Garis Kemiskinan Makanan

GKNM = Garis Kemiskinan Non Makan

Garis Kemiskinan Makanan (GKM) merupakan nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2100 kilokalori per kapita per hari. Paket komoditas kebutuhan dasar makanan diwakili oleh 52 jenis komoditas (padi-padian, umbi-umbian, ikan, daging, telur dan susu, sayuran, kacang-kacangan, buah-buahan, minyak dan lemak, dll.). Sedangkan, Garis Kemiskinan Non Makanan (GKNM) adalah kebutuhan minimum untuk perumahan, sandang, pendidikan dan kesehatan. Paket komoditas kebutuhan dasar non makanan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di pedesaan.

Tahapan untuk menghitung garis kemiskinan, yaitu: (1) Menghitung rata-rata tertimbang harga kalori yang diperlukan dari 52 komoditas makanan, (2) Mengalikan harga tersebut dengan 2.100 yang merupakan batas kemiskinan makanan per kapita per hari, (3) Menghitung nilai pengeluaran makanan dan non makanan per kapita, yang dinamakan garis kemiskinan, (4) Menghitung proporsi penduduk miskin dengan cara membagi jumlah penduduk miskin dengan jumlah penduduk (dinyatakan dalam persentase). Persamaan untuk menghitung persentase penduduk miskin adalah sebagai berikut

$$P_0 = \text{Banyaknya penduduk miskin} / \text{Jumlah penduduk} \times 100\% \quad (2)$$

Keterangan:

P_0 = Proporsi penduduk miskin

Small Area Estimation

Survei sampel memiliki beberapa keunggulan dibandingkan metode sensus dalam segi biaya, sumber daya manusia, waktu, dan informasi yang dikumpulkan. Dalam bukunya, Rao (2003) menjelaskan bahwa survei sampel secara luas digunakan dalam praktik pengumpulan data untuk memperoleh perkiraan/estimasi tidak hanya untuk total populasi yang diminati, tetapi juga untuk berbagai sub populasi (domain). Domain dapat didefinisikan sebagai area geografis (provinsi, kabupaten/kota, dan lain-lain) atau kelompok tertentu seperti kelompok sosio-demografis, jenis industri, dan sub populasi lainnya (Rao (2003), Pfeffermann (2013)).

Proses pendugaan pada suatu area atau sub populasi dapat dibagi menjadi dua macam, yaitu pendugaan berbasis rancangan (*design based estimator*) dan pendugaan berbasis model (*model based estimator*) (Wardani (2008), Pfeffermann (2013)). Pendugaan berbasis rancangan dilakukan berdasarkan data contoh dari area tempat survei dilakukan. Di sisi lain, pendugaan pada metode berbasis model merupakan pendugaan pada suatu area dengan cara menghubungkan informasi pada area tersebut dengan area lain melalui model yang tepat.

Salah satu metode yang termasuk pendugaan berbasis rancangan adalah pendugaan langsung (*direct estimator*), yaitu estimator yang dihitung hanya berdasarkan data sampel dari domain yang bersangkutan, menggunakan bobot survei (*weight*), dan kesimpulannya didasarkan pada distribusi probabilitas yang disebabkan oleh desain sampling dengan nilai populasi tetap (Wardani, 2008).

Sering kali pendugaan langsung dapat digunakan untuk memperkirakan karakteristik pada domain tertentu, namun tidak untuk karakteristik yang sama pada domain yang lebih kecil. Domain (area) dianggap “besar” jika ukuran sampel cukup besar untuk menghasilkan pendugaan langsung dengan presisi yang memadai. Presisi dapat dilihat dari nilai *standard error* atau *Relative Standard Error* (RSE)-nya. Sebaliknya, domain dianggap “kecil” jika ukuran sampel tidak cukup besar untuk mendukung estimasi langsung dari ketepatan (presisi) yang memadai. Penghitungan estimasi untuk area kecil dengan tingkat presisi yang memadai, sering kali perlu untuk menggunakan “*indirect estimator*” atau “estimator tidak langsung” dengan “meminjam kekuatan” dari area terkait dan atau area sekitarnya serta periode waktu tertentu sehingga meningkatkan efektivitas ukuran sampel.

Nilai dari variabel yang akan diestimasi atau variabel respons (Y) ini dimasukkan ke dalam proses estimasi melalui model (baik implisit atau eksplisit) yang menyediakan tautan ke area terkait dan atau periode waktu melalui penggunaan informasi tambahan yang berkorelasi dengan Y, seperti data yang berasal dari sensus terakhir dan/atau catatan administrasi saat ini. Penggunaan informasi tambahan dengan daya prediksi yang baik sangat penting untuk pembentukan estimator tidak langsung (Pfeffermann, 2013).

Beberapa metode estimasi yang termasuk estimator area kecil berbasis model, yaitu: estimator *Empirical Best Linear Unbiased Prediction* (EBLUP), *Empirical Bayes* (EB) dan *Hierarchical Bayes* (HB). Model penghubung eksplisit yang dibangun berdasarkan efek spesifik area acak yang memperhitungkan variasi area di luar yang dijelaskan oleh variabel tambahan (penjelas) yang termasuk dalam model akan disebut “model area kecil”. Model area kecil diklasifikasikan menjadi dua tipe besar, yaitu model level unit dan model level area.

Kesimpulan dari estimator berbasis model merujuk pada distribusi yang tersirat dalam model yang diasumsikan. Sehingga pemilihan dan validasi model memainkan peran penting dalam estimasi berbasis model. Jika model yang diasumsikan tidak memberikan kecocokan yang baik dengan data, maka estimator berbasis model akan menjadi bias, yang dapat menyebabkan kesalahan dalam pengambilan kesimpulan.

Metode EBLUP

Metode EBLUP merupakan pengembangan dari metode BLUP yang masih mengasumsikan bahwa komponen varians *area random effect* diketahui. Namun, pada kenyataannya varians *random effect* sulit diketahui sehingga harus diduga dari sampel yang ada. Dengan mengganti varians *random effect* populasi dengan varians *random effect* dari sampel maka didapat metode baru yaitu Empirical BLUP (EBLUP) (Molina, 2015).

Metode *Empirical Best Linier Unbiased Predictors* (EBLUP) yang diprakarsai Fay dan Herriot (1979), memiliki asumsi bahwa variabel penyerta tidak boleh memiliki error. Apabila variabel penyerta diperoleh dari data survei, model EBLUP tidak bisa dipakai karena data survei mengandung *error*.

Model EBLUP merupakan modifikasi dari model BLUP yang mengestimasi nilai σ_v^2 dari sampel dengan menggunakan pendekatan *Maximum Likelihood* (ML) atau *Restricted Maximum Likelihood* (REML) (Molina, 2015). Model EBLUP dinyatakan sebagai berikut:

$$\hat{\theta}_i^H = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \mathbf{z}_i^T \hat{\beta} \tag{3}$$

dimana,

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \psi_i}$$

Mean squared error (MSE) dari penduga EBLUP adalah sebagai berikut:

$$MSE(\hat{\theta}_i^H) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2) \tag{4}$$

Mean Squared Error (MSE) adalah rata-rata dari kuadrat perbedaan antara hasil estimasi dan nilai sebenarnya. MSE merupakan ukuran yang digunakan untuk melihat kualitas dari suatu estimator. Semakin rendah nilai MSE, maka perbedaan antara nilai estimasi dengan nilai sebenarnya semakin kecil, oleh karena itu suatu estimator dianggap baik jika mempunyai nilai MSE yang kecil.

Metode Hierarki Bayes

Teorema bayesian didasarkan pada distribusi posterior yang merupakan perpaduan antara distribusi prior (informasi masa lalu sebelum dilakukan observasi) dan data observasi yang digunakan untuk menyusun fungsi Likelihood (Box dan Tiao, 1973 dalam Ubaidillah, 2014). Hubungan distribusi posterior dengan distribusi prior dan Likelihood dapat ditulis sebagai berikut:

$$\text{Distribusi posterior} \propto \text{likelihood} \times \text{distribusi prior} \tag{5}$$

Pada teorema Bayes, apabila terdapat parameter θ yang diberikan oleh data observasi y , maka distribusi probabilitas untuk posterior θ pada data y akan proporsional dengan perkalian antara distribusi prior θ dan fungsi Likelihood θ yang diberikan oleh data y . Secara matematis dapat ditulis sebagai berikut.

$$f(\theta|y) \propto f(y|\theta).f(\theta) \tag{6}$$

dimana $f(\theta|y)$ merupakan distribusi posterior yang proporsional dengan perkalian antara fungsi Likelihood $f(y|\theta)$ dan distribusi prior $f(\theta)$.

Salah satu distribusi data yang dapat digunakan dalam SAE HB adalah distribusi Beta. Domain dari distribusi Beta bernilai antara 0 dan 1. Penduga model HB Beta dan varians penduga HB Beta diperoleh dari formula posterior-nya, sebagai berikut.

$$\hat{\theta}_i^{HB} = E(\theta_i | \hat{\theta}_i) \tag{7}$$

$$V(\hat{\theta}_i^{HB}) = V(\theta_i | \hat{\theta}_i) \tag{8}$$

Nilai penduga HB Beta dan varians penduga HB Beta tersebut diperoleh dengan metode numerik *Markov Chain Monte Carlo* (MCMC). Menurut Carlin (1992), pendekatan MCMC sangat efektif untuk mengurangi beban komputasi dalam menyelesaikan persamaan integrasi yang kompleks (Ubaidillah, 2014).

Tahapan Analisis

Tahapan-tahapan yang dilakukan pada penelitian ini, yaitu:

1. Identifikasi data dilakukan dengan mengidentifikasi data yang digunakan untuk menghasilkan peubah respons yaitu Susenas 2019 dan data yang menghasilkan *auxiliary variables* yaitu data Podes 2019.
2. Studi literatur dilakukan untuk mendapatkan dasar penggunaan *auxiliary variables* yang akan digunakan pada model SAE. Pada tahap ini ditentukan *auxiliary variables* yang akan digunakan.
3. Penyiapan data dilakukan dengan memilih lokus Provinsi Jawa Timur baik pada Data Susenas, Podes. Pendugaan langsung yaitu persentase penduduk miskin dihitung dari Susenas beserta statistik pendukung lainnya (*Standard Error, Variance, Lower Bound, Upper Bound, RSE, dan Design Effect*). Sedangkan, *auxiliary variables* dihitung dari data Podes.
4. Penerapan metode EBLUP dan HB untuk mengestimasi persentase penduduk miskin pada level kecamatan di Provinsi Jawa Timur. Pada metode HB, variabel respons berupa persentase penduduk miskin dibagi dengan 100 untuk mendapatkan nilai proporsi penduduk miskin dengan *range* 0 sampai dengan 1. Estimasi dilakukan dalam bahasa R menggunakan *package sae* untuk EBLUP (r-project, 2020).
5. Membandingkan hasil pendugaan metode EBLUP dan HB berdasarkan nilai RSE-nya dengan hasil pendugaan langsung.

HASIL DAN PEMBAHASAN

Hasil Eksplorasi Data

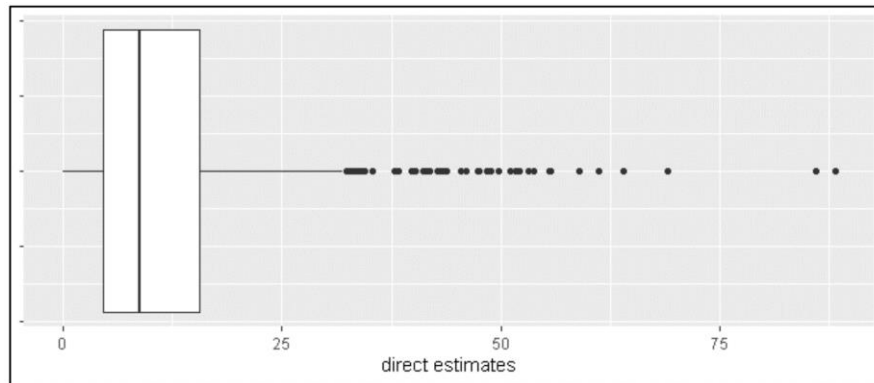
Eksplorasi data dilakukan guna mengetahui karakteristik data sehingga analisis dapat dilakukan dengan tepat. Hasilnya, diketahui terdapat sebanyak 666 kecamatan di Provinsi Jawa Timur. Dari total kecamatan tersebut, tidak seluruhnya ter sampel pada Susenas sehingga beberapa kecamatan pendugaan langsung persentase penduduk miskinnya bernilai NA, yaitu sebanyak 9 kecamatan. Selain itu, 25 kecamatan memiliki nilai pendugaan langsung 0, sisanya yaitu 632 kecamatan nilai pendugaan langsungnya lebih dari 0.

Dari eksplorasi pada nilai pendugaan langsung juga diketahui bahwa persentase penduduk miskin level kecamatan di Jawa Timur cukup beragam. Hal tersebut terlihat pada Tabel 1 yang menyajikan nilai statistik pendugaan langsung pada data ter sampel.

TABEL 1. Nilai Statistik Pendugaan Langsung Persentase Penduduk Miskin Level Kecamatan di Provinsi Jawa Timur, 2019

| Karakteristik | Persentase Penduduk Miskin |
|---------------|----------------------------|
| Minimum | 0 |
| Rata-rata | 12,31% |
| Median | 8,73% |
| Maksimum | 88,28% |

Pada *boxplot* Gambar 1, terlihat bahwa sebaran data pendugaan langsung persentase penduduk miskin tidak mengikuti distribusi normal melainkan menceng kanan dengan jumlah *outlier* atas sebanyak 47 kecamatan.



GAMBAR 1. Sebaran Nilai Pendugaan langsung persentase Penduduk Miskin di Jawa Timur, 2019

RSE digunakan untuk mengukur sejauh mana sampel yang digunakan sudah cukup menggambarkan keadaan parameter populasi yang digunakan atau dengan kata lain RSE yang dinyatakan dalam persentase mengukur presisi dari suatu estimasi. Dari Tabel 2, terlihat bahwa nilai RSE pendugaan langsung persentase penduduk miskin Jawa Timur tahun 2019 memiliki rentang antara 0 – 98,37%. Hal ini mengindikasikan bahwa hasil estimasi langsung masih belum cukup akurat. Selain itu juga terdapat sebanyak 25 kecamatan yang memiliki RSE NA akibat pendugaan langsungnya bernilai nol. Terdapat pula nilai RSE sama dengan 0 akibat dari pembulatan pada proses perhitungan.

TABEL 2. Nilai Statistik RSE Hasil Pendugaan langsung persentase Penduduk Miskin di Jawa Timur, 2019

| Statistik | RSE Pendugaan Langsung |
|-----------|------------------------|
| Minimum | 0 |
| Rata-rata | 40,61% |
| Median | 38,52% |
| Maksimum | 98,37% |

TABEL 3. Jumlah Kecamatan Ter sampel menurut Kelompok RSE Pendugaan langsung Persentase Penduduk Miskin di Jawa Timur, 2019

| Statistik | RSE Pendugaan Langsung |
|------------------------|------------------------|
| $RSE \leq 0$ | 31 |
| $0 < RSE \leq 25\%$ | 116 |
| $25\% < RSE \leq 50\%$ | 312 |
| $RSE > 50\%$ | 173 |
| $RSE = NA$ | 25 |

Selain pada hasil pendugaan langsung, eksplorasi data juga dilakukan pada *auxiliary variable* yang akan digunakan. Didapatkan 10 *auxiliary variable* dari hasil studi literatur dan regresi *stepwise*. Selanjutnya, pendugaan parameter dilakukan menggunakan 4 variabel yang berpengaruh signifikan dalam model, yaitu proporsi desa yang terdapat akses internet (X1), rasio SMP terhadap jumlah keluarga (X2), rasio kasus gizi buruk terhadap jumlah keluarga (X3), dan IKG 2019 (X4)

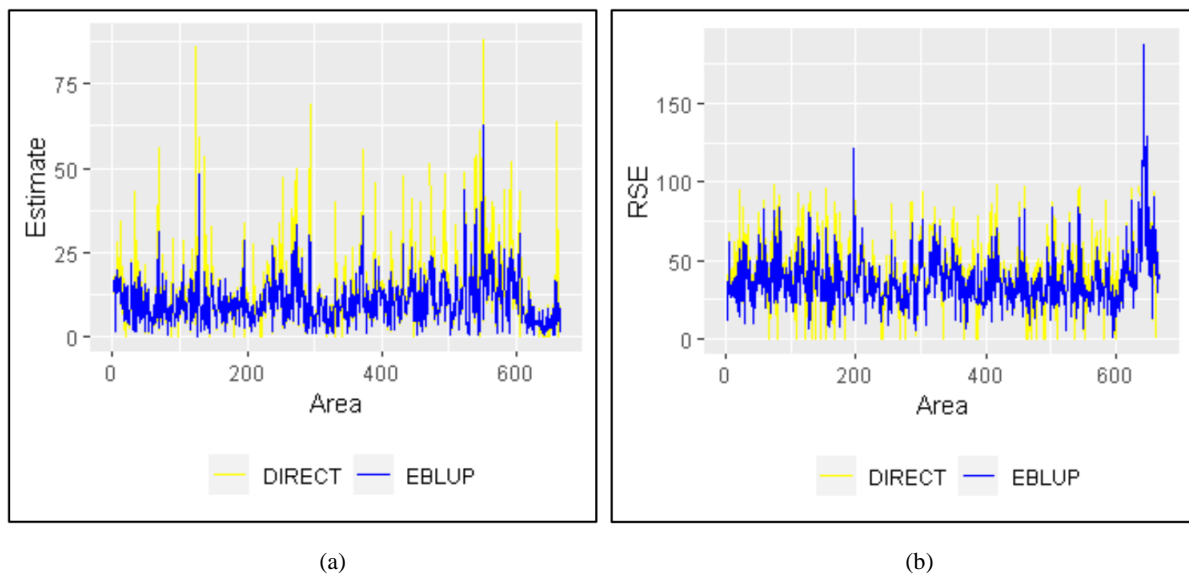
Pendugaan Kemiskinan dengan Metode EBLUP

Pendugaan parameter dilakukan menggunakan empat *auxiliary variable* yang berpengaruh secara signifikan terhadap persentase penduduk miskin. Estimasi *random effect variance* didapatkan sebesar 50,73943, sedangkan hasil estimasi nilai beta dan *standard error*-nya dapat dilihat pada Tabel 4.

TABEL 4. Hasil Estimasi Nilai Beta dan Standard Error Metode EBLUP

| Variabel | Beta | Standar Error |
|-----------|----------|---------------|
| Intercept | -0,77782 | 2,375068 |
| X1 | -3,83891 | 1,159809 |
| X2 | 3139,483 | 856,7852 |
| X3 | 636,3389 | 308,6022 |
| X4 | 0,356245 | 0,060634 |

Hasil estimasi langsung dan EBLUP untuk 666 kecamatan di Provinsi Jawa Timur dapat dilihat pada Gambar 2a. Terlihat bahwa hasil estimasi untuk kedua metode memiliki pola yang sama. Dengan memanfaatkan informasi *cluster*, didapatkan pula hasil estimasi untuk daerah-daerah tidak ter sampel dan area dengan estimasi langsung bernilai nol.



GAMBAR 2. Perbandingan Estimasi dan RSE persentase Penduduk Miskin Provinsi Jawa Timur Menggunakan Metode Direct dan EBLUP Provinsi Jawa Timur, 2019

Selain itu, dari Gambar 2b dapat disimpulkan pula bahwa metode EBLUP mampu menghasilkan nilai RSE lebih rendah dari hasil estimasi langsung untuk 544 dari 657 area ter sampel. Dengan metode SAE yang digunakan, didapatkan pula RSE hasil estimasi untuk daerah-daerah tidak ter sampel dan area dengan RSE estimasi langsung bernilai nol. Nilai RSE cenderung tinggi untuk beberapa area tidak ter sampel. Untuk area dengan RSE estimasi langsung bernilai nol, didapatkan nilai RSE positif sehingga pada plot terlihat RSE EBLUP lebih dari RSE hasil estimasi langsung. Secara garis besar, dapat dikatakan metode EBLUP dapat memperbaiki hasil estimasi langsung.

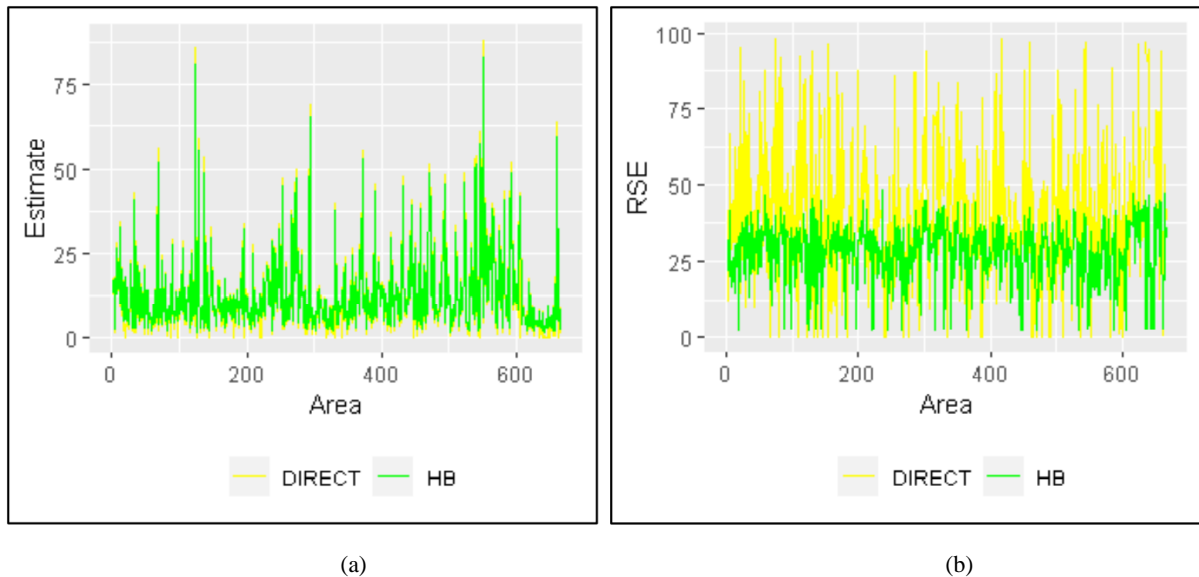
Pendugaan Kemiskinan dengan Metode HB

Pendugaan parameter menggunakan metode HB juga dilakukan menggunakan empat *auxiliary variable*. Dari hasil iterasi pada proses MCMC, didapatkan estimasi *area random effect* sebesar 0,7420516 dan estimasi parameter beta yang dilihat pada Tabel 5.

TABEL 5. Hasil Estimasi Nilai Mean Beta dan Standar Deviasi Metode HB

| Variabel | Beta | Standar Error |
|-----------|----------|---------------|
| Intercept | -3.57652 | 0.034892 |
| X1 | -0.5524 | 0.034989 |
| X2 | 97.7791 | 19.97572 |
| X3 | 51.78822 | 12.40957 |
| X4 | 0.050532 | 0.001013 |

Perbandingan hasil estimasi langsung dan metode HB beserta RSE-nya untuk 666 kecamatan di Provinsi Jawa Timur dapat dilihat pada Gambar 3a. Tidak berbeda dengan hasil estimasi EBLUP, hasil estimasi HB juga memiliki pola yang sama dengan hasil estimasi langsung. Selain itu, didapatkan pula hasil estimasi untuk daerah-daerah tidak ter sampel dan area dengan nilai estimasi langsung nol.



GAMBAR 3. Perbandingan Estimasi dan RSE persentase Penduduk Miskin Provinsi Jawa Timur Menggunakan Metode Direct dan HB Provinsi Jawa Timur, 2019.

Berdasarkan plot RSEnya pada Gambar 3b, terlihat bahwa metode HB mampu menghasilkan nilai RSE lebih rendah dari hasil estimasi langsung untuk 459 dari 657 area ter sampel. Didapatkan pula RSE hasil estimasi untuk area tidak ter sampel dan area dengan RSE estimasi langsung bernilai nol, dengan nilai yang tidak jauh berbeda dengan nilai RSE area ter sampel lainnya. Untuk area dengan RSE estimasi langsung bernilai nol, didapatkan nilai RSE positif sehingga pada plot terlihat RSE HB lebih dari RSE estimasi langsung. Secara garis besar, dapat dikatakan metode HB juga dapat memperbaiki hasil estimasi langsung.

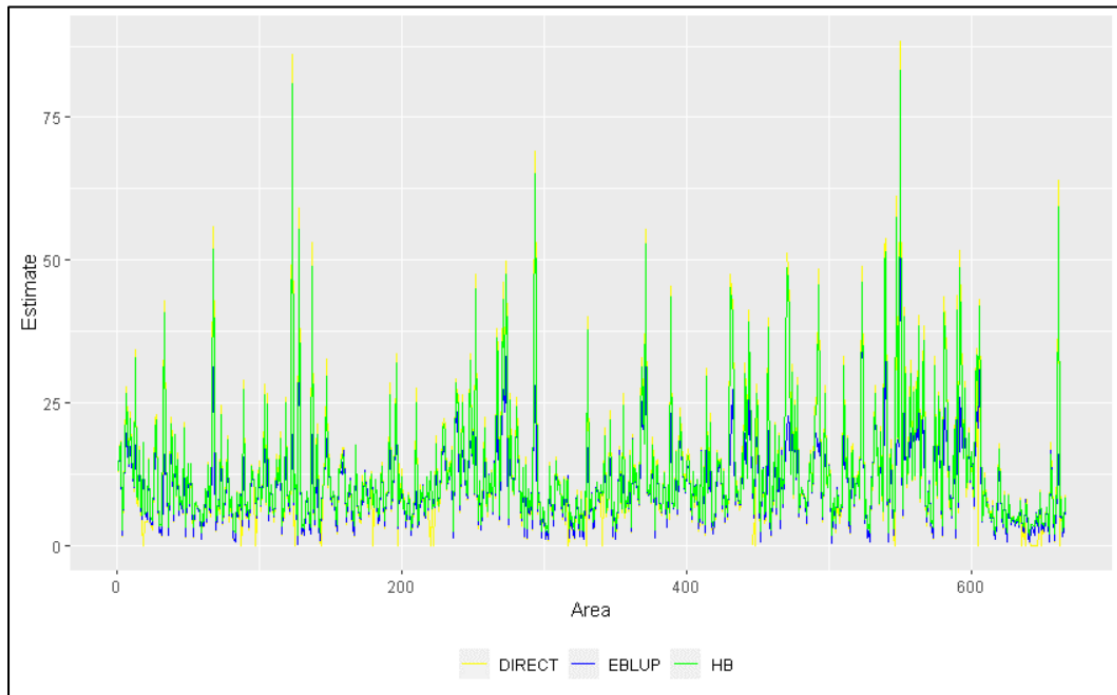
Perbandingan Hasil Pendugaan Metode EBLUP dan HB

Baik metode SAE EBLUP maupun HB telah mampu mengestimasi persentase penduduk miskin pada seluruh area, baik area ter sampel maupun tidak ter sampel. Dilihat dari nilai tengah dan rata-ratanya pada Tabel 6, penggunaan metode HB mampu menghasilkan hasil estimasi yang lebih mendekati nilai estimasi langsung. Nilai maksimal pada hasil estimasi metode HB juga tidak jauh berbeda dengan estimasi langsung. Hal ini berbeda dengan estimasi EBLUP yang memiliki nilai maksimum jauh lebih rendah dari hasil estimasi langsung, hal ini menunjukkan bahwa hasil estimasi EBLUP cenderung tertarik menuju rata-ratanya.

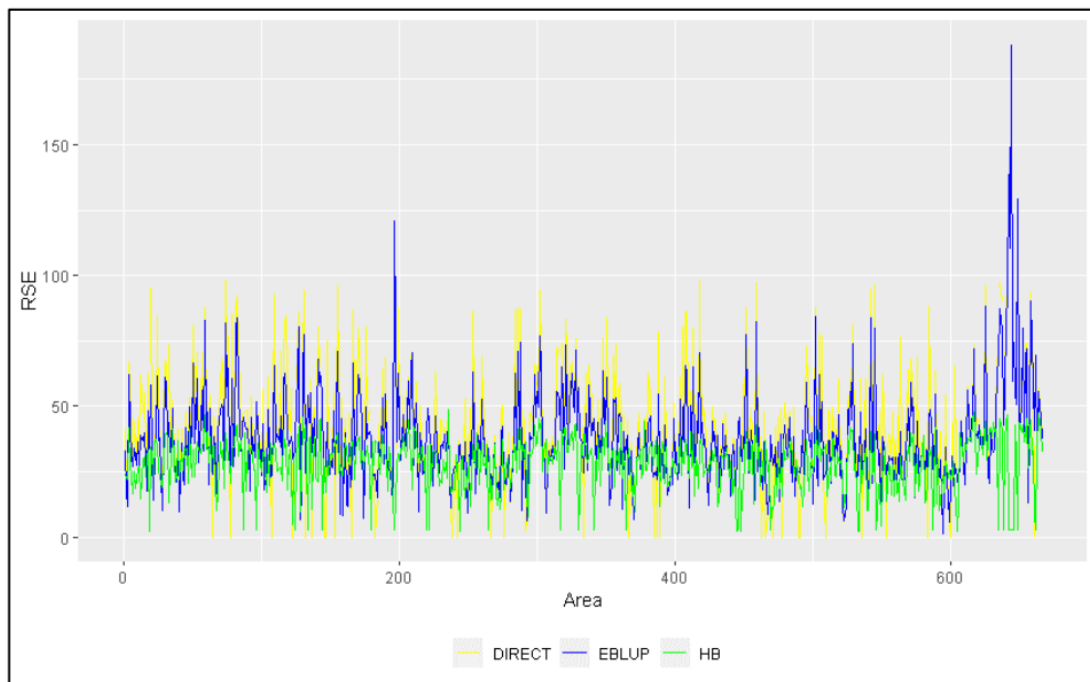
TABEL 6. Ringkasan Statistik Hasil Estimasi dan RSE persentase Penduduk Miskin Provinsi Jawa Timur menurut Metode yang Digunakan, 2019

| Statistik | Hasil Estimasi | | | RSE | | |
|-----------|--------------------|-------|-------|--------------------|--------|-------|
| | Pendugaan Langsung | EBLUP | HB | Pendugaan Langsung | EBLUP | HB |
| Minimal | 0 | 0,25 | 1,71 | 0 | 1,14 | 2,28 |
| Median | 8,73 | 9,135 | 9,05 | 38,515 | 35,45 | 29,69 |
| Mean | 12,31 | 10,47 | 12,69 | 40,61 | 37,52 | 27,94 |
| Maksimal | 88,28 | 62,97 | 83,17 | 98,37 | 187,99 | 48,73 |
| NA | 9 | 0 | 0 | 34 | 0 | 0 |

Dilihat dari nilai RSE, metode EBLUP dan HB mampu menurunkan nilai rata-rata dan median dari RSE hasil estimasi jika dibandingkan dengan estimasi langsung. Akan tetapi, nilai maksimum dari RSE metode EBLUP jauh di atas nilai maksimum RSE estimasi langsung yang berasal dari estimasi RSE area tidak ter sampel. Berdasarkan Gambar 4, diketahui bahwa hasil estimasi menggunakan HB lebih mendekati hasil estimasi langsung daripada hasil estimasi EBLUP. Selanjutnya, dilihat dari plot perbandingan RSE pada Gambar 5 dapat disimpulkan bahwa metode HB cenderung menghasilkan nilai RSE yang lebih rendah dari estimasi langsung dan SAE EBLUP.



GAMBAR 4. Perbandingan Hasil Estimasi persentase Penduduk Miskin Provinsi Jawa Timur Menggunakan Metode Direct, EBLUP, dan HB Provinsi Jawa Timur, 2019



GAMBAR 5. Perbandingan RSE Hasil Estimasi persentase Penduduk Miskin Provinsi Jawa Timur Menggunakan Metode Direct, EBLUP, dan HB Provinsi Jawa Timur, 2019

Hasil RSE metode HB secara rata-rata berada di bawah metode EBLUP. Dari Tabel 7, terlihat bahwa dibandingkan dengan metode EBLUP, metode HB mampu menghasilkan nilai estimasi yang lebih akurat dengan RSE seluruh sampel di bawah 50%. Nilai RSE yang lebih rendah menunjukkan bahwa hasil estimasi yang dihasilkan lebih akurat. Sehingga dapat dikatakan metode HB mampu memperbaiki hasil estimasi langsung lebih baik dari metode EBLUP.

TABEL 7. Jumlah Kecamatan menurut Kelompok RSE dan Metode Estimasi persentase Penduduk Miskin Provinsi Jawa Timur, 2019

| Kelompok RSE | Pendugaan Langsung | EBLUP | HB |
|----------------------|---------------------------|--------------|-----------|
| RSE \leq 0 | 31 | - | - |
| 0 < RSE \leq 25% | 116 | 159 | 220 |
| 25% < RSE \leq 50% | 312 | 391 | 446 |
| RSE > 50% | 173 | 116 | - |
| RSE = NA | 25 | - | - |

KESIMPULAN DAN SARAN

Penggunaan SAE EBLUP dan HB pada pendugaan persentase penduduk miskin level kecamatan Provinsi Jawa Timur tahun 2019 terbukti dapat meningkatkan keakuratan estimasi dibandingkan pendugaan langsung. Hal ini terlihat dari berkurangnya jumlah kecamatan yang memiliki RSE >50%. Selain itu, metode EBLUP dan HB dapat mengatasi pendugaan pada area yang tidak ter sampel dengan memanfaatkan informasi cluster. Meskipun demikian, masih ditemukan nilai maksimum dari RSE metode EBLUP yang jauh di atas nilai maksimum RSE pendugaan langsung yang diduga disebabkan oleh estimasi RSE area tidak ter sampel. Di sisi lain, metode HB mampu menghasilkan nilai estimasi yang lebih akurat dengan RSE seluruh sampel di bawah 50%. Nilai RSE yang lebih rendah menunjukkan bahwa hasil estimasi yang dihasilkan lebih akurat. Sehingga dapat disimpulkan bahwa metode HB mampu memperbaiki hasil estimasi langsung lebih baik dari metode EBLUP pada pendugaan persentase penduduk miskin level kecamatan Provinsi Jawa Timur tahun 2019. Kedepannya, perlu dieksplor lagi pemilihan variabel penjelas yang digunakan dalam model karena akan mempengaruhi kinerja model dalam menurunkan RSE. Selain itu, perlu juga untuk melihat agregasi dengan level yang lebih tinggi terkait dengan konsistensi hasil pendugaan. Hal itu dapat dilakukan dengan SAE Benchmarking. Salah satu alternatif yang dapat digunakan untuk mengatasi banyaknya nilai nol pada variabel respons yaitu model zero inflated.

REFERENSI

- Bappenas, 2019. Analisis Wilayah dengan Kemiskinan Tinggi. Jakarta: Bappenas.
- BPS, 2021. Penduduk di bawah Garis Kemiskinan. Diakses dari: <https://sirusa.bps.go.id/sirusa/index.php/indikator/196>
- BPS, 2019. Penghitungan dan Analisis Kemiskinan Makro Tahun 2019. Jakarta: BPS.
- Ikhsan, E., Hidayat, C. A., Nurizza, W. A., 2018. Efisiensi Metode Eblup pada Small Area Estimation. Studi Kasus: Estimasi Persentase Penduduk Miskin di Provinsi Nusa Tenggara Timur Tahun 2017. Jurnal Aplikasi Statistika & Komputasi Statistik V.10.2.2018, ISSN 2086-4132.
- Molina, J. R. 2015. Small Area Estimation: Second Edition. New Jersey: John Wiley & Sons, Inc.
- Notodiputro, K. A. & Kurnia, A., 2010. Small Area Statistic in Indonesia.
- Noviyanti, R. A. & Zain, I. 2014. Pendekatan Small Area Estimation pada Scan Statistic untuk Pendeteksian Kantong Kemiskinan. Prosiding Seminar Nasional Matematika, universitas Jember, 19 November 2014.
- Permatasari, N. & Ubaidillah, A., 2021. msae: An R Package of Multivariate Fay-Herriot Models for Small Area Estimation. The R Journal, 13(2), pp. 111-122.

- Pfeffermann, D., 2013. New Important Developments in Small Area Estimation. *Journal of Statistical Science*, 2013, Vol. 28, No. 1, 40–68. DOI: 10.1214/12-STS395.
- r-project, 2020. sae: Small Area Estimation. Diakses dari: <https://cran.r-project.org/web/packages/sae/>
- Rao, J., 2003. *Small Area Estimation*. New Jersey: John Wiley & Sons, Inc.
- Ubaidillah, A., 2014. *Small Area Estimation dengan Pendekatan Hierarchical Bayesian Neural Network untuk Pemetaan Kemiskinan di Kota Jambi [Thesis]*. Surabaya: Institute Teknologi Sepuluh November.
- Ubaidillah, A., Notodiputro, K. A., Kurnia, A. & Mangku, I. W., 2019. Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia. *Journal of Applied Statistics*, 46(15), pp. 2845-2861.
- Wardani, A., 2008. *Perbandingan Metode Empirical Bayes (EB) dan Empirical Best Linear Unbiased Prediction (EBLUP) pada Pendugaan Area Kecil (Studi Kasus Pendugaan Pengeluaran Per Kapita di Kota Bogor)* [Skripsi]. Bogor: Institute Pertanian Bogor.
- Yuliani, J. R. B., Maiyastri & Diana, R., 2019. Pendugaan Angka Kemiskinan di Kabupaten Padang Pariaman Menggunakan Small Area Estimation dengan Pendekatan Hierarchical Bayesian (HB) Loglogistik. *Jurnal Matematika Unand*, Volume 8, pp. 76-83.