PENERAPAN MODEL REGRESI *THRESHOLD* UNTUK PENINGKATAN EFISIENSI DALAM PENDUGAAN AREA KECIL

Studi Simulasi dan Penerapannya Pada Data Pengeluaran Perkapita Makanan Rumahtangga di Provinsi Jawa Tengah

(Application of Threshold Regression Models for Increasing Efficiency in Small Area Estimation)

Azka Ubaidillah¹

Politeknik Statistika STIS¹

Jalan Otto Iskandardinata No. 64 C Jakarta Timur

E-mail: azka@stis.ac.id

ABSTRAK

Pendugaan area kecil (SAE) dewasa ini berkembang cukup pesat seiring dengan meningkatnya kebutuhan atas penyediaan statistik yang terpercaya di area kecil, yaitu area dengan jumlah contoh (sampel) yang sedikit atau tidak mencukupi untuk dilakukan pendugaan secara langsung. Metode SAE dapat meningkatkan efektivitas contoh dengan "meminjam" kekuatan dari informasi area yang bertetangga dan pengaruh peubah penyertanya. Dalam aplikasinya, model Fay-Herriot menggunakan pendekatan *Empirical Best Linear Prediction* (EBLUP) banyak dilakukan karena sifat modelnya yang sederhana. Salah satu sifat sederhana dari model EBLUP adalah penggunaan hubungan linier antara peubah yang diamati dengan peubah penyertanya. Namun sering dijumpai bahwa hubungan linier tersebut belum cukup untuk meningkatkan efisiensi model SAE sebagai akibat pola yang terbentuk antara peubah amatan dan peubah penyertanya tidak linier. Paper ini menjelaskan salah satu alternatif cara untuk meningkatkan efisiensi model SAE dengan menerapkan model regresi *threshold*. Dari hasil simulasi dan aplikasi data pengeluaran perkapita makanan tingkat kabupaten/kota di Jawa Tengah tahun 2015 diperoleh keterangan bahwa model regresi *threshold* menghasilkan pendugaan dengan RMSE (*root mean square error*) dan RSE (*relative standard error*) yang lebih kecil dibandingkan model EBLUP. Hal ini menunjukkan bahwa penerapan model regresi *threshold* mampu untuk meningkatkan efisiensi dalam pendugaan area kecil.

Kata kunci: EBLUP, efisiensi, SAE, dan *threshold* regression.

ABSTRACT

Abstract Small Area Estimation (SAE) is rapidly developing along with the increasing need for providing reliable statistics in small areas, areas with an insufficient number of samples for direct estimation. The SAE method can increase the effectiveness of the sample by "borrowing" the strength of neighboring area information and the influence of its auxiliary variables. In application, the Fay-Herriot model with the Empirical Best Linear Prediction (EBLUP) approach, is mostly employed due to the simple nature of the model. One of the simple characteristics of the EBLUP model is the use of a linear relationship between the variables of interest and the auxiliary variables. However, it is often found that the linear relationship is not enough to increase the efficiency of the SAE model because there is a pattern formed between the observed variables and the accompanying variables which are not linear. This paper offers one alternative way to improve the efficiency of the SAE model by applying the threshold regression model. From the simulation study and the application to per capita expenditure of food estimation at the district level in Central Java in 2015, it was obtained that the threshold regression model produced estimates with RMSE (root mean square error) and RSE (relative standard error) which are smaller than the EBLUP model. This shows that the application of the threshold regression model is appropriate to improve efficiency in SAE method.

Keywords: EBLUP, efficiency, SAE, and *threshold* regression

PENDAHULUAN

Selama beberapa dekade terakhir ini, survei telah mengambil alih peranan sensus atau pencacahan lengkap sebagai cara yang lebih hemat untuk mendapatkan informasi tentang topik menarik yang lebih luas. Survei dapat digunakan untuk memperoleh estimasi yang andal untuk area yang luas (Cochran, 1977). Namun, penduga langsung untuk wilayah kecil, berdasarkan sampel data dari survei, cenderung tidak efisien karena memiliki kesalahan baku yang besar. Akibatnya hasil pendugaan langsung untuk area kecil menjadi tidak terpercaya. Untuk itu kemudian dikembangkan metode *Small Area Estimation* (SAE) untuk meningkatkan efektivitas ukuran sampel di area kecil dengan meminjam kekuatan dari area yang bersesuaian dan hubungan antara peubah penyerta (*auxiliary variable*) dan peubah amatan (Rao dan Molina, 2015).

Berdasarkan ketersediaan peubah penyerta, SAE dapat diklasifikasikan menjadi dua jenis yaitu (Rao dan Molina, 2015): model level area dan model level unit. Model level area adalah model yang digunakan apabila peubah penyerta hanya tersedia pada level area, misalnya kabupaten, kecamatan, desa atau kelurahan. Sedangkan model level unit adalah model yang digunakan apabila peubah penyerta tersedia sampai level unit, misalnya rumahtangga. Model level area lebih sering digunakan dalam analisis karena tidak memerlukan ketersediaan peubah penyerta sampai level unit. Model SAE Level area pertama kali diaplikasikan oleh Fay dan Herriot (1979) menggunakan penduga *Empirical Best Linear Unlimited Prediction* (EBLUP) untuk menduga log pendapatan per kapita di area kecil di Amerika Serikat tahun 1972 (Rao dan Molina, 2015).

Model EBLUP adalah model sederhana dimana salah satu asumsi yang digunakan adalah adanya hubungan linier antara peubah yang menjadi amatan dengan koefisien parameter dari peubah penyerta. Namun seringkali dijumpai bahwa hubungan linier tersebut tidak cukup untuk meningkatkan efisiensi hasil pendugaan model SAE. Salah satu alternatif yang bisa digunakan untuk meningkatkan efisiensi pendugaan model SAE adalah dengan memanfaatkan informasi tambahan dari peubah penyerta berupa parameter *threshold*. Menurut Fong dkk (2017), model regresi *threshold* adalah model regresi nonlinier yang bergantung pada titik perubahan atau *threshold* (ambang batas). Model ini menggunakan cara yang sederhana namun elegan dan dapat ditafsirkan untuk memodelkan jenis hubungan nonlinear tertentu antara peubah amatan dan peubah penyertanya.

Beberapa analisis yang menggunakan aplikasi model regresi *threshold* diantaranya adalah kajian pengaruh dari inovasi teknologi regional pada pembangunan berkelanjutan yang ditulis oleh Yi dan Li (2018) dengan menggunakan data panel. Meskipun begitu, menurut Dagenais (1967) dan Hansen (1999), model regresi *threshold* dapat juga diaplikasikan untuk data *cross section* misalkan kegiatan rumahtangga dalam pembelian barang dan kegiatan perukaan dalam meningkatkan kapasitas usahanya. Dalam penelitian ini, aplikasi model regresi *threshold* akan digunakan untuk meningkatkan efisiensi pendugaan SAE pada data pengeluaran per kapita makanan rumahtangga pada level kabupaten/kota di Provinsi Jawa Tengah tahun 2015. Data pendugaan langsung diperoleh dari Survei Sosial Ekonomi Nasional (Susenas) tahun 2015 di Provinsi Jawa Tengah.

Struktur penulisan paper ini dibagi menjadi empat bab, yaitu pendahuluan, metode, hasil dan pembahasan serta kesimpulan. Bab pendahuluan menjelaskan latar belakang penelitian meliputi alasan menggunakan metode SAE dengan penambahan pengaruh dari model regresi *threshold* serta aplikasinya pada data pengeluaran per kapita makanan rumahtangga. Bab metode menjelaskan metode SAE dan model regresi *threshold* serta daftar peubah yang digunakan. Bab hasil dan pembahasan menjelaskan hasil simulasi dan aplikasi model pada *real data*. Bab kesimpulan menjelaskan kesimpulan yang diperoleh dari penelitian ini.

METODE

Metode SAE adalah model pendugaan tidak langsung dengan memanfaatkan informasi tambahan dari peubah penyerta yang mempunyai hubungan kuat dengan peubah amatan. Peubah penyerta diasumsikan tidak memiliki kesalahan pengukuran (*measurement error*) yang bersumber dari data sensus atau catatan administratif. Berdasarkan ketersediaan peubah penyertanya, metode SAE dibagi menjadi dua bentuk model, yaitu model level area dan model level unit. Salah satu

kendala yang dialami oleh model level unit adalah penyediaan peubah penyerta sampai level sampling unit terkecil yaitu rumahtangga. Oleh karena itu dalam penelitian ini akan digunakan model SAE level area.

Pada model SAE level area, peubah yang diamati diasumsikan merupakan fungsi dari rata-rata peubah respon, $\theta_i = g(\bar{Y}_i)$ untuk $g(\cdot)$ tertentu yang berkaitan dengan peubah penyerta $\mathbf{x}_i = (x_{i1}, \dots, x_{iK})^T$ yang tersedia sampai level area kecil dan mengikuti model linier sebagai berikut:

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + z_i u_i, \quad i = 1, \dots, m \text{ (area kecil)} \dots \dots \dots (1)$$

Dengan T menunjukkan notasi *transpose* matriks, z_i adalah konstanta bernilai positif yang diketahui dan $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)^T$ adalah koefisien regresi. Sedangkan u_i adalah pengaruh acak area kecil yang diasumsikan memiliki sebaran (seringkali sebaran normal) yang identik dan saling bebas dengan nilai harapan $E(u_i) = 0$ dan ragam $\text{Var}(u_i) = \sigma_u^2$. Untuk melakukan inferensi tentang rata-rata area kecil $\bar{Y}_i$ pada persamaan (1), dimisalkan suatu penduga langsung $\hat{Y}_i$ diketahui yaitu:

$$\hat{\theta}_i = g(\hat{Y}_i) = \theta_i + e_i \dots \dots \dots (2)$$

dimana $e_i \sim \text{ind}(0, \psi_i)$ adalah sampling error dan ψ_i diketahui. Dengan menggabungkan persamaan (1) dan (2) maka diperoleh model sebagai berikut:

$$\hat{\theta}_i = \mathbf{x}_i^T \boldsymbol{\beta} + z_i u_i + e_i \dots \dots \dots (3)$$

dengan u_i dan e_i saling bebas. Model (3) merupakan bentuk model campuran linier atau *linear mixed models* (LMM) dan dikenal sebagai model Fay-Herriot.

Penduga *Best Linear Unbiased Prediction* (BLUP) untuk θ_i dirumuskan sebagai berikut (Rao dan Molina, 2015):

$$\tilde{\theta}_i^H = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \gamma_i (\hat{\theta}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}) = \gamma_i \hat{\theta}_i + (1 - \gamma_i) \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} \dots \dots \dots (4)$$

dimana

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_u^2) = [\sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T / (\sigma_u^2 z_i^2 + \psi_i)]^{-1} [\sum_{i=1}^m \mathbf{x}_i \hat{\theta}_i / (\sigma_u^2 z_i^2 + \psi_i)] \dots \dots \dots (5)$$

adalah penduga BLUE untuk $\boldsymbol{\beta}$, dan

$$\gamma_i = \sigma_u^2 z_i^2 / (\sigma_u^2 z_i^2 + \psi_i) \dots \dots \dots (6)$$

Penduga BLUP (4) mengandung nilai σ_u^2 yang tidak diketahui. Untuk memperoleh penduga σ_u^2 , beberapa metode yang bisa digunakan diantaranya adalah metode moment, ML dan REML (Rao dan Molina, 2015). Dengan mengganti σ_u^2 yang tidak diketahui pada persamaan (4) menjadi $\hat{\sigma}_u^2$, maka diperoleh penduga EBLUP $\hat{\theta}_i^H$ sebagai berikut:

$$\hat{\theta}_i^H = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \dots \dots \dots (7)$$

dimana $\hat{\boldsymbol{\beta}}$ dan $\hat{\gamma}_i$ adalah nilai dari $\tilde{\boldsymbol{\beta}}$ dan γ_i pada persamaan (5) dan (6) ketika σ_u^2 diganti dengan $\hat{\sigma}_u^2$. Persamaan (7) menggambarkan bahwa penduga EBLUP ($\hat{\theta}_i^H$) adalah nilai rata-rata tertimbang (dengan penimbang $\hat{\gamma}_i$) dari penduga langsung ($\hat{\theta}_i$) dan model sintesisnya ($\mathbf{x}_i^T \hat{\boldsymbol{\beta}}$).

Mean squared error (MSE) untuk $\hat{\theta}_i^H$ dengan asumsi u_i dan e_i menyebar normal adalah sebagai berikut (Rao dan Molina, 2015):

$$\text{MSE}(\hat{\theta}_i^H) \approx g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2) \dots \dots \dots (8)$$

dimana

$$g_{1i}(\sigma_u^2) = \gamma_i \psi_i,$$

$$g_{2i}(\sigma_u^2) = (1 - \gamma_i)^2 \mathbf{x}_i^T [\sum_{t=1}^m \mathbf{x}_t \mathbf{x}_t^T / (\sigma_u^2 z_t^2 + \psi_t)]^{-1} \mathbf{x}_i \text{ dan}$$

$$g_{3i}(\sigma_u^2) = E(\hat{\theta}_i^H - \tilde{\theta}_i^H)^2 = \psi_i^2 z_i^4 (\sigma_u^2 z_i^2 + \psi_i)^{-3} \bar{V}(\hat{\sigma}_u^2) \dots \dots \dots (9)$$

dimana $\bar{V}(\hat{\sigma}_u^2)$ adalah penduga ragam asimtotik untuk σ_u^2 dari metode ML atau REML yang besarnya sama untuk $m \rightarrow \infty$ dan dirumuskan:

$$\bar{V}(\hat{\sigma}_u^2)_{ML} = \bar{V}(\hat{\sigma}_u^2)_{RE} = [J(\sigma_u^2)]^{-1} = 2 \left[\sum_{i=1}^m z_i^4 / (\sigma_u^2 z_i^2 + \psi_i)^2 \right]^{-1}, \dots\dots\dots (10)$$

dimana $J(\sigma_u^2)$ adalah notasi matriks informasi Fisher. Adapun penduga tak bias untuk $MSE(\hat{\theta}_i^H)$ dengan asumsi u_i dan e_i menyebar normal adalah sebagai berikut:

$$mse(\hat{\theta}_i^H) = g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2) + 2g_{3i}(\hat{\sigma}_u^2) \dots\dots\dots (11)$$

Diantara beberapa kendala dalam pemodelan EBLUP adalah kesulitan mendapatkan peubah penyerta yang mampu meningkatkan efisiensi hasil pendugaan. Salah satu penyebabnya adalah hubungan yang tidak linier antara peubah amatan dengan peubah penyerta. Untuk itu dibutuhkan pemodelan non linier, salah satunya menggunakan pendekatan pemodelan regresi *threshold*. Menurut Fong dkk (2017), model regresi *threshold* menggunakan cara yang sederhana namun elegan dan dapat ditafsirkan untuk memodelkan jenis hubungan nonlinear tertentu antara peubah amatan dan peubah penyertanya.

Menurut Fong dkk (2017), ada beberapa tipe model regresi *threshold*, diantaranya adalah segmented model, hinge model, upper hinge model, step model dan segmented model. Ilustrasi dari beberapa jenis model regresi *threshold* dapat dilihat pada Gambar 1 berikut ini:

Gambar 1. Tipe Model Regresi *Threshold*

Pada penelitian ini akan digunakan salah satu tipe model regresi *Threshold*, yaitu tipe segmented model dengan rumus sebagai berikut (Fong dkk, 2017):

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha(X'_i - \tau)_+ + \emptyset X'_i + e_i, \dots\dots\dots (12)$$

dimana notasi τ menunjukkan parameter *threshold*, X' menunjukkan peubah penjelas dengan pengaruh *threshold*, α menunjukkan parameter segmentasi pada saat $X' \geq \tau$, $\emptyset$ menunjukkan parameter peubah X' dan x menunjukkan peubah penjelas yang tidak ada pengaruh *threshold*. Nilai $(X'_i - \tau)_+ = X'_i - \tau$, jika $X' \geq \tau$ dan $(X'_i - \tau)_+ = 0$ jika $X' < \tau$. Keterangan lebih lanjut dapat dilihat di Fong (2018), Fong dkk (2017) dan Hansen (1999). Apabila nilai $\alpha = 0$, maka persamaan (12) akan menjadi model regresi linier berganda.

Aplikasi model digunakan untuk menduga pengeluaran perkapita makanan rumah tangga di Jawa Tengah pada tingkat kabupaten/kota tahun 2015. Data tersebut bersumber pada hasil Survei Sosial Ekonomi Nasional (Susenas) bulan September 2015 dimana survei dirancang untuk level pendugaan tingkat provinsi. Untuk itu metode SAE (EBLUP) akan diterapkan supaya diperoleh hasil pendugaan yang lebih reliabel. Peubah penyerta yang digunakan dalam pemodelan mengikuti penelitian Ubaidillah dkk (2019) bersumber pada publikasi Provinsi Jawa Tengah dalam Angka 2016 (BPS, 2016) meliputi jumlah keluarga pra sejahtera (X_1), upah minimum kabupaten (X_2), PDRB atas

dasar harga konstan (X_3) dan proporsi jumlah sekolah menengah atas (SMA) per 10.000 penduduk (X_4).

HASIL DAN PEMBAHASAN

Hasil penelitian meliputi hasil dari kajian simulasi dan aplikasi *real data*. Kajian simulasi mencakup dua simulasi yaitu simulasi 1 untuk mengevaluasi bias dan kinerja model regresi *threshold* dibandingkan model regresi linier berganda. Simulasi 2 untuk mengevaluasi perbandingan hasil pemodelan SAE dengan metode EBLUP dan EBLUP₊ (EBLUP dengan penambahan informasi *threshold*). Sedangkan aplikasi *real data* untuk membandingkan hasil pendugaan area kecil antara metode pendugaan langsung (*direct estimation*), metode EBLUP dan metode EBLUP₊.

Simulasi 1 dilakukan dengan langkah-langkah sebagai berikut:

- Gunakan persamaan (12) dan gunakan parameter $\beta_0 = 2$, $\beta_1 = 5$, $\alpha = -5$, $\phi = 5$.
- Gunakan m (notasi untuk observasi atau area) yaitu m1=30, m2=50 dan m3=100.
- Untuk masing-masing m, bangkitkan
 - $X_i \sim Unif(0,1)$, $X'_i \sim Unif(10,14)$ dan parameter *threshold* $\tau = \bar{X}'$.
 - Lakukan pengulangan sebanyak K=100 kali untuk
 - Bangkitkan $e_i^{[k]} \sim N(0,4)$; $k = 1, \dots, K$, kemudian hitung $Y_i^{[k]}$,
 - Lakukan pemodelan dengan regresi linier berganda dan regresi *threshold*. Model regresi berganda mengasumsikan X'_i sebagai peubah penjelas yang tidak terpengaruh oleh *threshold* atau $\alpha = 0$. Dari masing-masing kedua model tersebut kemudian hitung nilai Absolut Bias, Relatif Bias, R_{adj}^2 dan ResSE (*Residual standard error*).
- Hitung rata-rata Bias, rata-rata Relatif Bias, rata-rata R_{adj}^2 dan rata-rata ResSE

Tabel 4. Nilai rata-rata absolut bias, rata-rata relatif bias dari model regresi linier berganda dan model regresi *threshold*

m	Penduga Parameter	Regresi Linier Berganda		Regresi <i>Threshold</i>	
		Absolut Bias	Relatif Bias	Absolut Bias	Relatif Bias
m1=30	$\hat{\beta}_0$	26,012	13,006	6,893	3,446
	$\hat{\beta}_1$	2,411	0,482	0,624	0,125
	$\hat{\phi}$	1,465	0,293	1,121	0,224
m2=50	$\hat{\beta}_0$	25,037	12,518	4,057	2,028
	$\hat{\beta}_1$	2,270	0,454	0,361	0,072
	$\hat{\phi}$	1,002	0,200	0,912	0,182
m3=100	$\hat{\beta}_0$	27,130	13,565	3,505	1,753
	$\hat{\beta}_1$	2,460	0,492	0,313	0,063
	$\hat{\phi}$	0,595	0,119	0,595	0,119

Tabel 2. Nilai rata-rata R_{adj}^2 dan rata-rata residual standard error (ResSE) model regresi linier berganda dan model regresi *threshold*

m	Regresi Linier Berganda		Regresi <i>Threshold</i>	
	R_{adj}^2	ResSE	R_{adj}^2	ResSE
m1=30	0,598	2,476	0,740	1,981
m2=50	0,697	2,470	0,804	1,984
m2=100	0,654	2,436	0,773	1,974

Hasil simulasi 1 ditampilkan dalam Tabel 1 dan Tabel 2. Dari Tabel 1 terlihat bahwa model regresi *threshold* menghasilkan nilai absolut bias dan relatif bias yang lebih kecil dibandingkan model

linier berganda. Sedangkan dari Tabel 2 terlihat bahwa model regresi *threshold* menghasilkan nilai R^2_{adj} yang lebih besar dibandingkan dengan R^2_{adj} yang dihasilkan oleh regresi linier berganda. Hal ini mengindikasikan bahwa untuk data yang pola hubungan antara peubah amatan dan peubah penjelasnya tidak berupa hubungan linier, maka pendekatan regresi linier berganda cenderung bias dan tidak efisien.

Selanjutnya dilakukan simulasi 2 untuk mengevaluasi perbandingan hasil pemodelan SAE dengan metode EBLUP dan EBLUP₊ (EBLUP dengan penambahan informasi *threshold*). Langkah-langkah simulasi 2 hampir sama dengan simulasi 1 hanya saja menambahkan komponen efek acak area ke dalam model regresi. Model persamaan yang digunakan adalah model (3) dikombinasi dengan model (12) sebagai berikut:

1. Gunakan parameter $\beta_0 = 2$, $\beta_1 = 5$, $\alpha = -5$, $\emptyset = 5$.
2. Gunakan m (notasi untuk observasi atau area) yaitu m1=30, m2=50 dan m3=100.
3. Untuk masing-masing m, bangkitkan
 - a. $X_i \sim Unif(0,1)$, $X'_i \sim Unif(10,14)$ dan parameter *threshold* $\tau = \bar{X}'$.
 - b. Lakukan pengulangan sebanyak K=100 kali untuk
 - (i) Bangkitkan $e_i^{\{k\}} \sim N(0,4)$ dan $u_i^{\{k\}} \sim N(0,9)$; kemudian dan parameter θ_i ,
 - (ii) Hitung parameter θ_i dan penduga langsung $\hat{\theta}_i$,
 - (ii) Hitung penduga EBLUP $\hat{\theta}_i^H$ dengan model EBLUP dan model EBLUP₊. Model EBLUP mengasumsikan X'_i sebagai peubah penyerta yang tidak terpengaruh oleh *threshold* atau $\alpha = 0$. Dari masing-masing kedua model tersebut kemudian hitung nilai *relative mean squared error* (RMSE) dan *relative standard error* (RSE).
4. Hitung rata-rata RMSE, rata-rata RSE dari hasil model EBLUP dan EBLUP₊.

Tabel 3 menunjukkan hasil simulasi 2 dimana terlihat bahwa model regresi EBLUP₊ menghasilkan nilai rata-rata RMSE dan rata-rata RSE yang lebih kecil dibandingkan model penduga langsung dan model EBLUP. Hal ini mengindikasikan bahwa untuk data yang pola hubungan antara peubah amatan dan peubah penjelasnya yang tidak linier, maka pendekatan EBLUP₊ lebih efisien dibandingkan dengan model EBLUP.

Tabel 3. Perbandingan nilai rata-rata $\hat{\gamma}$, rata-rata RMSE dan rata-rata RSE antara model pendugaan langsung, model EBLUP dan EBLUP₊

m (Area)	Kriteria	Model		
		Penduga langsung	EBLUP	EBLUP ₊
m1=30	Rata-rata $\hat{\gamma}$	1.000	0.733	0.657
	Rata-rata RMSE	2.000	1.782	1.728
	Rata-rata RSE	3.209	2.855	2.769
m2=50	Rata-rata $\hat{\gamma}$	1.000	0.733	0.678
	Rata-rata RMSE	2.000	1.754	1.707
	Rata-rata RSE	3.251	2.847	2.771
m3=100	Rata-rata $\hat{\gamma}$	1.000	0.728	0.673
	Rata-rata RMSE	2.000	1.728	1.672
	Rata-rata RSE	3.262	2.814	2.722

Gambar 2 menunjukkan boxplot nilai RMSE hasil simulasi 2 dengan m=100 sebanyak 100 kali iterasi dari model pendugaan langsung, model EBLUP dan model EBLUP₊. Dari boxplot tersebut terlihat model SAE menggunakan EBLUP dan EBLUP₊ mampu menurunkan nilai RMSE dari hasil pendugaan langsung secara signifikan. Hal ini menunjukkan bahwa pendugaan model SAE lebih *reliable* (dapat dipercaya) dibandingkan hasil pendugaan langsung. Tingkat reliabilitas suatu penduga dapat dilihat dari besarnya nilai RMSE. Semakin kecil nilai RMSE maka suatu penduga akan

semakin reliable karena selang kepercayaan (*confident interval*) yang dihasilkan oleh suatu penduga akan semakin kecil.

Gambar 2 juga menunjukkan bahwa model EBLUP₊ menghasilkan nilai RMSE yang lebih kecil dibandingkan dengan RMSE hasil pendugaan EBLUP. Hal ini mengindikasikan bahwa model EBLUP₊ lebih dapat diandalkan dibandingkan model EBLUP untuk kasus hubungan tidak linier antara peubah amatan dengan peubah penyertanya.

Gambar 2. Boxplot RMSE Hasil Simulasi dengan $m=100$ sebanyak 100 kali iterasi dari Model Pendugaan Langsung, Model EBLUP dan Model EBLUP₊

Aplikasi *real data* diterapkan untuk data pengeluaran per kapita makanan rumah tangga pada level kabupaten/kota di Provinsi Jawa Tengah tahun 2015. Formula penduga langsung (*direct estimator*) dan peubah penyertanya mengikuti formula pada penelitian yang dilakukan oleh Benavent dan Morales (2016) dan Ubaidillah dkk (2019). Hasil penerapan model (EBLUP dan EBLUP₊) pada *real data* dapat dilihat pada Tabel 4:

Tabel 4. Perbandingan nilai $\hat{y}$, rata-rata RMSE dan rata-rata RSE antara model pendugaan langsung, model EBLUP dan EBLUP₊

Kriteria	Pendugaan langsung	EBLUP	EBLUP ₊
$\hat{y}$	1.00000	0.87930	0.87345
Rata-rata RMSE	0.04336	0.04124	0.04122
Rata-rata RSE	0.33720	0.32082	0.32061

Gambar 3. Plot RMSE Model Penduga Langsung dan Model EBLUP+

Tabel 4 menunjukkan bahwa model EBLUP₊ mempunyai nilai $\hat{\gamma}$ yang paling kecil dibandingkan model pendugaan langsung dan model EBLUP. Artinya pengaruh peubah penyerta pada model EBLUP₊ lebih besar dibanding model EBLUP. Selain itu, nilai rata-rata RMSE dan nilai rata-rata RSE yang dihasilkan oleh model EBLUP₊ juga lebih kecil dibanding model pendugaan langsung dan model EBLUP. Hal ini membuktikan bahwa model EBLUP₊ lebih efisien dibandingkan model pendugaan langsung dan model EBLUP.

Secara visual, Gambar 3 terlihat plot RMSE dari hasil penduga EBLUP₊ yang lebih kecil dibandingkan RMSE penduga langsung pada semua area kecil (kabupaten/kota) di Provinsi Jawa Tengah. Hal ini mengindikasikan bahwa penduga EBLUP₊ lebih *reliable* (dapat dipercaya) dibandingkan dengan pendugaan langsung.

KESIMPULAN

Dari hasil simulasi dan aplikasi *real data* terlihat bahwa penerapan model regresi *threshold* pada SAE mampu memberikan pengaruh yaitu meningkatkan efisiensi hasil pendugaan area kecil. Untuk pola hubungan antara peubah amatan dan peubah penyerta yang tidak linier, model EBLUP dapat diperbaiki kinerjanya dengan memanfaatkan parameter *threshold* dari peubah penyertanya.

DAFTAR PUSTAKA

- Benavent, R. and D. Morales. (2016). *Multivariate Fay-Herriot models for small area estimation*. *Comput. Stat. Data Anal*, 94, 372–390.
- BPS (Badan Pusat Statistik Provinsi Jawa Tengah). (2016). *Provinsi Jawa Tengah Dalam Angka 2016*. Badan Pusat Statistik Provinsi Jawa Tengah.
- Cochran, W.G. (1977). *Sampling techniques*, 3rd ed., John Wiley & Son, New York.
- Dagenais, M.G. (1967). *A Threshold Regression Model*. *Econometrica*. Vol 37, No.2.
- Fay, R.E. and R.A. Herriot. (1979). *Estimates of income for small places: an application of James-Stein procedures to census data*. *J. Amer. Stat. Ass.* 74: 269–277.
- Fong, Y. (2018). *Fast Bootstrap Confidence Intervals for Continuous Threshold Linear Regression*. *Journal of Computational and Graphical Statistics*, in press.
- Fong, Y., Y. Huang, P. Gilbert and S. Permar. (2017). *chngpt: threshold regression model estimation and inference*. *BMC Bioinformatics*, 18: 454–460.

- Hansen, B.E. (1999). *Threshold effects in non-dynamic panels: Estimation, testing, and inference*, *Journal of Econometrics* 93: 345-368.
- Rao, J.N.K. and I. Molina. (2015). *Small Area Estimation, 2nd Ed.*, John Wiley and Sons Inc., New York.
- Ubaidillah, A., K.A. Notodiputro, A. Kurnia and I.W. Mangku. (2019). *Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia*. *Journal of Applied Statistics*. <https://doi.org/10.1080/02664763.2019.1615420>
- Yi, S. and Li, A.X. (2018). *Application of threshold regression analysis to study the impact of regional technological innovation level on sustainable development*. *Renewable and Sustainable Energy Reviews*. 89: 27-32.