M-quantile Chambers-Dunstan Untuk Pendugaan Area Kecil

Studi Kasus: Data Pengeluaran Rumah Tangga Per Kapita DI Yogyakarta Tahun 2018

(M-quantile Chambers-Dunstan Small Area Model with Case Per Capita Household Expenditure Data in Special Region Yogyakarta 2018)

Aldi Rochman Nulkarim^{1*}, Ika Yuni Wulansari²

^{1,2}Politeknik Statistika STIS

Jl. Otto Iskandardinata No.64C Jakarta 13330

E-mail: 211709522@stis.ac.id

ABSTRAK

Metode *Small Area Estimations* (SAE) digunakan sebagai pendekatan yang reliabel dalam mengatasi kendala ketidakcukupan sampel pada survei sampel. BPS mengkaji statistik area kecil menggunakan metode SAE populer seperti *Empirical Best Linear Unbiased Prediction* dalam model Fay-Herriot (EBLUP-FH). Metode EBLUP-FH sebagai pendekatan parametrik memerlukan asumsi normalitas dan terbebas dari *outliers* pada kedua komponen *random effect*-nya. Namun, hal tersebut sulit dipenuhi karena seringkali data di lapangan berperilaku ekstrim. Metode SAE *M-quantile Chambers-Dunstan* (CD) merelaksasi asumsi parametrik dan *robust* dalam inferensi terhadap *outliers*. Penelitian ini mengkaji metode *M-quantile CD* melalui penerapannya pada data riil untuk estimasi rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan di DI Yogyakarta tahun 2018. Penelitian ini menggunakan data Susenas 2018 dan Podes 2018. Hasil implementasi pada data riil menunjukkan model *M-quantile CD* berhasil memperbaiki presisi *EBLUP-FH*. Dengan mengimplementasikan *M-quantile CD* diharapkan estimasi data berperilaku ekstrim lebih akurat untuk pengambilan kebijakan di daerah.

Kata kunci: SAE, *M-quantile*, Pengeluaran

ABSTRACT

The *Small Area Estimations* (SAE) method is used as a reliable approach in overcoming sample insufficiency problem in sample surveys. BPS studies small area statistics using popular SAE methods such as *Empirical Best Linear Unbiased Prediction* in Fay-Herriot (EBLUP-FH) model. The EBLUP-FH method is a parametric approach that requires the assumption of normality and is free from outliers on the random effect. However, it is difficult to satisfy because real data often behave in extreme ways. The SAE *M-quantile Chambers-Dunstan* (CD) method relaxes parametric assumptions and is robust to outliers. This study examines *M-quantile CD* through its application to real data in estimating average household expenditure per capita at sub-district level in DI Yogyakarta 2018. This study uses Susenas and Podes data in 2018. The result shows that *M-quantile CD* succeeds in improving the precision of *EBLUP-FH*. By implementing *M-quantile CD*, it is expected that the estimation of extreme data is more accurate for local area policymaking.

Keywords: SAE, *M-quantile*, Expenditure

PENDAHULUAN

Lebih dari satu dekade terakhir telah tumbuh banyak permintaan statistik area kecil. Perubahan sistem pemerintahan Indonesia dari sentralisasi menjadi desentralisasi berdasarkan amanat UU No. 32 tahun 2004 telah mendorong meningkatnya pembentukan kebijakan pada level area lokal. Hal ini sangat tepat karena fenomena yang terjadi di area lokal bersifat heterogen dan memiliki karakteristik yang berbeda antara satu area dengan area lainnya. Oleh karena itu, kebijakan yang dibuat akan lebih akurat jika berlandaskan pada informasi statistik di level area lokal tersebut. Namun sayangnya, survei sampel tidak dapat memenuhi kebutuhan akan statistik area kecil. Kendala biaya dalam desain survei akan menyebabkan tidak terpenuhinya kecukupan sampel dalam pendugaan area kecil sehingga menghasilkan presisi pada tingkat yang tidak reliabel.

Salah satu pendekatan yang reliabel digunakan untuk mengatasi permasalahan ketidakcukupan sampel pada survei adalah dengan menggunakan metode *Small Area Estimation* (SAE). Metode ini mampu menghasilkan estimasi yang lebih efisien dibandingkan estimasi langsung (Tzavidis *et.al.*, 2018). Metode SAE merupakan bentuk dari estimasi tak langsung yang meminjam kekuatan (*borrow strength*) dari area lain, biasanya melalui model yang menghubungkan area-area yang ada dengan memanfaatkan informasi penyerta yang berasal dari sensus atau catatan administrasi (Rao, 2003).

Seringkali model yang paling populer digunakan dalam SAE adalah model *random effect*. Model ini memasukan *random area effect* untuk menjelaskan variasi antar area di luar yang dijelaskan oleh variabel penyerta (Rao, 2003). Salah satu contoh model *random effect* pada SAE seperti *Empirical Best Linear Unbiased Prediction* (EBLUP) dalam model Fay-Herriot (FH). Badan Pusat Statistik (2019) mulai memproduksi statistik area kecil dengan mengaplikasikan metode SAE populer seperti EBLUP-FH dalam publikasinya yang berjudul estimasi upah rata-rata per jam pekerja pada level kabupaten dengan menggunakan metode SAE.

Model semacam ini termasuk kedalam pendekatan parametrik dimana bagian *random*-nya sangat bergantung pada asumsi distribusi normal dan dapat dengan mudah terpengaruh oleh *outliers*. Pada praktiknya, hal demikian sulit untuk dipenuhi karena data di lapangan umumnya terpapar oleh *outliers*. Tidak terpenuhinya asumsi parametrik pada pemodelan SAE berbasis *random effect* akan menyebabkan presisi yang dihasilkan tidak optimal. Oleh karenanya, diperlukan pendekatan lain yang lebih fleksibel terhadap asumsi parametrik dan *robust* terhadap *outliers* sehingga pendugaan area kecil yang dilakukan tetap menghasilkan *presisi* yang optimal.

Chambers & Tzavidis (2006) memperkenalkan suatu pendekatan baru dalam SAE sebagai solusi untuk mengatasi kelemahan SAE berbasis model *random effect* dengan menggunakan *M-quantile small area model*. Berbeda dengan model SAE *random effect* pada umumnya, metode *M-quantile* tidak memerlukan asumsi distribusi dan *robust* dalam melakukan inferensi terhadap *outliers*. Hal ini karena *M-quantile small area model* tidak menggunakan bagian *random effect* untuk menjelaskan variasi antar area, akan tetapi menggunakan koefisien area-spesifik *M-quantile* untuk mengkarakterisasi perbedaan antar area. Selanjutnya, Tzavidis *et.al.* (2010) menjelaskan bahwa model *M-quantile* yang dikembangkan oleh Chambers & Tzavidis (2006) dapat menjadi bias apabila terdapat *error* yang asimetris atau heteroskedastis. Oleh karena itu, Tzavidis *et.al.* (2010) mengembangkan model *M-quantile Chambers-Dunstan* (CD) *small area model* dengan menambahkan *estimator Chambers-Dunstan* sebagai pengoreksi bias.

Pengeluaran rumah tangga per kapita adalah salah satu indikator kesejahteraan ekonomi masyarakat yang penting untuk diukur dalam skala mikro agar menjadi kompas bagi pembangunan di daerah. Oleh sebab itu, keakuratan dalam mengestimasi pengeluaran rumah tangga per kapita pada area kecil menjadi perhatian penting. Estimasi langsung untuk menduga rata-rata pengeluaran rumah tangga per kapita pada level area kecil akan menghasilkan estimasi dengan presisi yang tidak reliabel karena adanya masalah ketidakcukupan sampel. Metode SAE berbasis pemodelan EBLUP-FH dapat menjadi alternatif untuk digunakan dalam mengestimasi rata-rata pengeluaran rumah tangga per kapita pada level area kecil.

Namun, menurut Barigozzi *et.al.* (2009) sebaran data pengeluaran rumah tangga per kapita cenderung menceng ke kanan. Hal ini berarti sebaran data pengeluaran rumah tangga per kapita secara alami tidak mengikuti distribusi normal. Battistin *et.al.* (2009) menambahkan bahwa data pengeluaran rumah tangga per kapita biasanya memiliki *outliers*. Kedua karakteristik alami dari data pengeluaran rumah tangga per kapita ini akan menyulitkan terpenuhinya asumsi normalitas dan spesifikasi baku pada bagian *random* dari EBLUP-FH. Oleh karenanya penerapan *M-quantile CD small area model* dapat menjadi solusi karena mampu menghindari permasalahan terkait pemenuhan asumsi distribusi maupun spesifikasi baku pada bagian *random effect*. Adapun bagian *random effect* terdiri dari *random area effect* dan *sampling error* (*random individual effect*). Maka dari itu, pengaplikasian *M-quantile CD small area model* diharapkan dapat memperbaiki presisi dari EBLUP-FH dalam mengestimasi rata-rata pengeluaran per kapita pada area kecil. Penelitian sebelumnya mengenai pengeluaran konsumsi per kapita di Albania yang dilakukan oleh Tzavidis *et.al.* (2007) membuktikan bahwa tingkat akurasi dan presisi yang dihasilkan metode *M-quantile* secara rata-rata lebih tinggi dibandingkan metode EBLUP.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengkaji perbandingan performa antara metode *M-quantile CD* dan EBLUP-FH dengan mengaplikasikannya untuk mengestimasi rata – rata pengeluaran per kapita pada tingkat kecamatan di DI Yogyakarta. Dengan didapatkannya performa *M-quantile CD* dibandingkan EBLUP-FH, diharapkan dapat menjadi informasi mengenai solusi untuk pendugaan area kecil ketika data yang dimiliki berperilaku ekstrim.

METODE

Landasan Teori

Small Area Estimation (SAE)

Small area estimation merupakan prosedur statistik dalam menghasilkan pendugaan yang efisien (presisi) pada area kecil, dimana area kecil tersebut memiliki ukuran sampel yang kecil sehingga presisi yang dihasilkan

tidak reliabel atau tidak memiliki sampel sama sekali. Kemudian Rao (2003) membagi metode SAE kedalam dua model berdasarkan ketersediaan informasi penyerta, yaitu model *area-level* ketika informasi penyerta tersedia hingga level area dan model *unit-level* ketika informasi penyerta tersedia hingga level unit.

Empirical Best Linear Unbiased Prediction dalam model Fay-Herriot (EBLUP-FH)

Molina & Marhuenda (2015) menjelaskan model Fay-Herriot sebagai model *area-level* yang mengkombinasikan *sampling model* dan *linking model* kedalam *linear mixed model* sebagai berikut.

$$\hat{\theta}_j^{DIR} = \mathbf{x}_j^T \boldsymbol{\beta} + v_j + e_j \dots \dots \dots (1)$$

dimana

$$j = 1, \dots, d$$

$$v_j \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

$$e_j \stackrel{iid}{\sim} N(0, A_j)$$

$$\hat{\theta}_j^{DIR} = \text{direct estimator}$$

$\mathbf{x}$ = variabel penyerta

$\boldsymbol{\beta}$ = koefisien regresi sintetik

v = *random area effect*

e = *sampling error (random individual effect)*

Empirical Best Linear Unbiased Predictor (EBLUP) merupakan pengembangan dari *Best Linear Unbiased Predictor (BLUP)*. BLUP mengasumsikan varians dari *random area effect* telah diketahui (σ_v^2), akan tetapi pada praktiknya σ_v^2 tidak diketahui. Maka pada EBLUP dilakukan estimasi terhadap σ_v^2 , dengan salah satunya menggunakan metode REML sehingga dihasilkan estimator yang konsisten $\hat{\sigma}_v^2$. EBLUP dalam model Fay-Herriot dapat diekspresikan dalam bentuk kombinasi *direct estimator* dan *regression-synthetic estimator* yang dapat ditulis sebagai berikut.

$$\hat{\theta}_j^{EBLUP} = \hat{\gamma}_j \hat{\theta}_j^{DIR} + (1 - \hat{\gamma}_j) \mathbf{x}_j^T \hat{\boldsymbol{\beta}} \dots \dots \dots (2)$$

dimana

$$\hat{\gamma}_j = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + A_j)$$

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\sigma}_v^2) = \left[\sum_{j=1}^d (\hat{\sigma}_v^2 + A_j)^{-1} \mathbf{x}_j \mathbf{x}_j^T \right]^{-1} \sum_{j=1}^d \sum_{i=1}^d (\hat{\sigma}_v^2 + A_j)^{-1} \mathbf{x}_j \hat{\theta}_j^{DIR}$$

Pratesi & Giusti (2015) menunjukkan bahwa EBLUP-FH memiliki kelebihan yakni estimasi yang dihasilkan akan efisien jika asumsi normalitas pada *linear mixed model*-nya terpenuhi. Di sisi lain, EBLUP-FH memiliki beberapa kekurangan yaitu adanya asumsi normalitas baik pada *random area effect* dan *sampling error*-nya, modelnya tidak dalam *design-unbiased* sehingga estimasinya akan *biased* jika di bawah desain survei yang kompleks, dan inferensinya tidak *robust* terhadap *outliers*.

M-quantile Small Area Model

Sebuah pendekatan baru dalam pendugaan area kecil dirumuskan oleh Chambers & Tzavidis (2006) yang didasarkan pada regresi *M-quantile*. Model ini merelaksasi asumsi parametrik pada model *mixed effect*, lebih lanjut penggunaan *M-estimator* menjamin inferensi yang *robust* terhadap *outliers*. Regresi *M-quantile* memberikan generalisasi “*quantile-like*” pada regresi berbasis *influence function*. Regresi linear *M-quantile* dapat diekspresikan sebagai $Q_q(X; \psi) = X^T \hat{\boldsymbol{\beta}}_\psi(q)$, dimana estimasi $\hat{\boldsymbol{\beta}}$ diperoleh melalui prosedur *iterative weighted least square (IRLS)*.

Chambers & Tzavidis (2006) mengembangkan model regresi *M-quantile* dalam konteks pendugaan area kecil. Mereka mengamati bahwa jika perbedaan antar area adalah bagian yang signifikan dalam menjelaskan variabilitas populasi, unit populasi yang berada di area yang sama diharapkan memiliki koefisien *M-quantile* yang sama. Sebagai hasilnya, Chambers & Tzavidis (2006) mengembangkan prediktor rata – rata area kecil dengan mengasumsikan efek area dapat diekspresikan menggunakan nilai rata – rata dari nilai *M-quantile* pada unit di area tertentu. Nilai *M-quantile* tersebut meliputi *observed value* dan *unobserved value*. Persamaan *M-quantile* untuk area kecil dapat dituliskan sebagai berikut

$$\hat{y}_j^{MQ} = N_j^{-1} \left[\sum_{i \in S_j} y_{ij} + \sum_{i \in r_j} \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_\psi(\hat{q}_j) \right] \dots \dots \dots (3)$$

dimana

$$\hat{y}_j^{MQ} = \text{rata-rata variabel yang diteliti hasil estimasi } M\text{-quantile pada area ke-} j$$

$$N_j = \text{ukuran populasi pada area ke-} j$$

$$y_{ij} = \text{variabel yang diteliti pada area ke-} j \text{ dan unit ke-} i$$

β_ψ = koefisien regresi *M-quantile* dengan *influence function* ψ
 $\hat{q}_j$ = estimasi rata-rata koefisien *M-quantile* pada unit di area ke- j
 X_{ij} = variabel penyerta pada area ke- j dan unit ke- i
 s_j = sub populasi yang terpilih dalam sampel
 r_j = sub populasi yang tidak terpilih dalam sampel

Tzavidis et al (2010) menjelaskan bahwa model (3) dapat menjadi bias jika terdapat error yang heteroskedastis atau asimetris. Selanjutnya Tzavidis et al (2010) mengajukan pengembangan model *M-quantile* yaitu model *bias-adjusted M-quantile* yang dirumuskan sebagai berikut.

$$\hat{y}_j^{MQCD} = N_j^{-1} \left[\sum_{i \in s_j} y_{ij} + \sum_{i \in r_j} X_{ij}^T \hat{\beta}_\psi(\hat{q}_j) + \frac{N_j - n_j}{n_j} \sum_{i \in s_j} (y_{ij} - X_{ij}^T \hat{\beta}_\psi(\hat{q}_j)) \right] \dots\dots\dots (4)$$

dimana

n_j = ukuran sampel pada area ke- j

Model (4) dikenal dengan *M-quantile Chambers-Dunstan (CD)* karena berbasis estimator Chambers-Dunstan sebagai pengoreksi bias. Penggunaan estimator CD untuk mengoreksi bias akan menghasilkan estimator yang lebih efisien apabila terjadi kontaminasi *outliers*.

Tzavidis et al (2007) memformulasikan *Mean Square Error (MSE)* untuk *M-quantile small area model* dalam bentuk estimator MSE untuk area ke- j yang diekspresikan sebagai berikut.

$$\widehat{MSE}(\hat{y}_j^{MQ}) = \hat{V}(\hat{y}_j^{MQ}) + \left\{ \hat{B}(\hat{y}_j^{MQ}) \right\}^2 \dots\dots\dots (5)$$

dimana

$$\hat{V}(\hat{y}_j^{MQ}) = N_j^{-2} \sum_g \sum_{i \in s_g} \left[\left\{ (w_i - 1)^2 + \frac{N_j - n_j}{n_j - 1} \right\} I(g = j) + w_i^2 I(g \neq j) \right] \cdot \{ y_i - \mathbf{x}_i^T \hat{\beta}_\psi(\hat{\theta}_g) \}^2 \dots\dots\dots (6)$$

Selanjutnya Tzavidis et al (2007) menyatakan bahwa ketika $\widehat{MSE}(\hat{y}_j^{MQCD})$ maka estimator $\hat{B}(\hat{y}_j^{MQCD}) = 0$. Hal tersebut diakibatkan oleh penggunaan *CD estimator* yang dapat mengoreksi bias

Metode Pengumpulan Data

Seluruh data yang digunakan pada penelitian ini berasal dari data sekunder. Data yang digunakan berupa raw data yang mencakup data Survei Sosial Ekonomi Nasional (Susenas) 2018 sebagai sumber pembentukan variabel respon dan data Potensi Desa (Podes) 2018 sebagai sumber variabel penyerta. Lebih lanjut, variabel-variabel yang digunakan merupakan hasil sintesis dari penelitian terdahulu yang ditampilkan pada **Tabel 1**.

Tabel 1. Variabel hasil sintesis penelitian terkait

No (1)	Variabel (2)	Keterangan Variabel (3)	No (4)	Variabel (5)	Keterangan Variabel (6)
1	Y	Pengeluaran Rumah Tangga Per kapita	13	X12	Jumlah Puskesmas tanpa rawat inap
2	X1	Sumber utama sebagian besar penghasilan penduduk	14	X13	Jumlah Puskesmas pembantu
3	X2	Jumlah keluarga yang berlangganan telepon kabel	15	X14	Jumlah tempat praktik dokter
4	X3	Jumlah keluarga pengguna listrik PLN	16	X15	Jumlah tempat praktik bidan
5	X4	Jumlah warga peserta BPJS kesehatan Penerima Bantuan Iuran (PBI) dan Jamkesda	17	X16	Jumlah Polindes
6	X5	Jumlah surat miskin/SKTM yang dikeluarkan	18	X17	Jumlah posyandu
7	X6	Jumlah toko atau warung kelontong	19	X18	Jumlah minimarket / swalayan
8	X7	Jumlah bank umum	20	X19	Jumlah restoran/rumah makan
9	X8	Jumlah koperasi yang masih aktif/beroperasi	21	X20	Jumlah hotel dan penginapan
10	X9	Jumlah Lembaga Pendidikan (SD/MI, SMP/MTs, SMU/MA)	22	X21	Jumlah TKI
11	X10	Jumlah Poliklinik	23	X22	Jumlah Industri Mikro dan Kecil
12	X11	Jumlah Apotek	24	X23	Jumlah lokasi permukiman di bantaran sungai

Sumber: Hasil sintesis penelitian terkait

Metode Analisis

Analisis yang digunakan dalam penelitian ini meliputi analisis deskriptif dan analisis inferensia. Analisis deskriptif dalam penelitian ini bertujuan untuk mengetahui karakteristik data yang dilihat menggunakan grafik distribusi, grafik garis dan *boxplot* serta statistik deskriptif yang digunakan untuk merangkum karakteristik data. Sementara itu, analisis inferensia pada penelitian ini berfokus pada estimasi langsung dan estimasi tidak langsung berbasis pemodelan SAE. Terdapat 74 kecamatan di DI Yogyakarta yang akan dilakukan estimasi rata-rata pengeluaran rumah tangga per kapita. Pada estimasi langsung desain yang digunakan adalah desain Susenas. Sedangkan pada SAE metode yang digunakan meliputi dua pendekatan yaitu pendekatan parametrik dengan menggunakan metode *EBLUP-FH* dan pendekatan semiparametrik dengan menggunakan metode *M-quantile CD small area models*. Selanjutnya metode *M-quantile CD* dilakukan modifikasi menjadi *M-quantile CD pseudo area-level* dengan menggunakan *proxy* desa untuk level unit. Hal ini dilakukan karena adanya keterbatasan data penyerta yang hanya tersedia hingga level desa. Penggunaan *auxiliary variables* level area pada metode *M-quantile* pernah dilakukan dalam penelitian Fadhila (2019) dan Girinoto (2017). Selanjutnya seluruh prosedur analisis dilakukan dengan menggunakan bantuan *software R*. Adapun prosedur analisis dapat dijabarkan dalam tahapan berikut:

1. Melihat gambaran umum karakteristik data pengeluaran rumah tangga per kapita DI Yogyakarta tahun 2018
2. Menyiapkan variabel pengeluaran rumah tangga per kapita sebagai *variable of interest* dari *raw data* Susenas 2018
3. Melakukan estimasi langsung rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan DI Yogyakarta tahun 2018 menggunakan desain Susenas.
4. Memilih variabel penyerta dari kandidat variabel penyerta pada Tabel 1. Variabel penyerta yang dipilih adalah variabel penyerta yang memiliki hubungan yang kuat dengan variabel respon.
5. Melakukan pendugaan area kecil rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan DI Yogyakarta tahun 2018 menggunakan metode *EBLUP-FH*.
6. Melakukan uji *outliers Grubbs* untuk mendeteksi kontaminasi *outliers* pada *random area effect* dan *sampling error* hasil estimasi metode *EBLUP-FH*. Adapun kontaminasi *outliers* adalah persentase nilai *random effect* yang termasuk *outliers* dibandingkan seluruh nilai *random effect*. Semisal komponen *random area effect* memiliki tingkat kontaminasi *outliers* 15% dan jumlah seluruh *random area effect* sebanyak 70. Hal ini berarti terdapat $15\% \cdot 60 = 9$ *outliers*.
7. Melakukan uji normalitas Shapiro-Wilk pada *random area effect* dan *sampling error* hasil estimasi metode *EBLUP-FH*
8. Melakukan pendugaan area kecil rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan DI Yogyakarta tahun 2018 menggunakan metode *M-quantile CD small area model*. Dimana
9. Membandingkan RSE sebagai ukuran presisi pada masing-masing metode estimasi yang meliputi estimasi langsung, estimasi *EBLUP-FH*, dan estimasi *M-quantile CD small area model*.
10. Menetapkan metode estimasi terbaik berdasarkan tingkat presisi yang dihasilkan dalam mengestimasi rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan DI Yogyakarta tahun 2018.

HASIL DAN PEMBAHASAN

Pada penelitian ini, estimasi rata-rata pengeluaran per kapita rumah tangga dilakukan baik dengan menggunakan metode estimasi langsung dan *SAE*. Performa dari estimasi langsung digunakan sebagai tolok ukur peningkatan performa metode *SAE*. Metode *SAE* yang digunakan terdiri dari *EBLUP-FH* dan *M-quantile CD*.

Gambaran Umum Karakteristik Rumah Tangga Per Kapita di Provinsi DI Yogyakarta Tahun 2012

Berikut disajikan gambaran umum melalui visualisasi yang meliputi grafik distribusi dan *boxplot* untuk mengidentifikasi karakteristik pengeluaran rumah tangga per kapita di DI Yogyakarta pada Gambar 1 dan Gambar 2.

Grafik distribusi pada Gambar 1 dibawah menunjukkan bahwa distribusi data pengeluaran rumah tangga per kapita sangat menceng ke kanan. Adanya kemencengan ini mengakibatkan distribusi data menjadi tidak normal sehingga menyulitkan pemenuhan asumsi normalitas variabel respon dalam metode *SAE* konvensional seperti *EBLUP-FH*. Selain itu, nilai rata-rata yang dihasilkan terlihat memiliki jarak yang cukup jauh dari pusat data. Hal ini dapat diakibatkan karena rata-rata sangat terpengaruh oleh nilai-nilai *outliers*.

Sementara itu, *boxplot* pada Gambar 2 menunjukkan bahwa data pengeluaran rumah tangga per kapita terkontaminasi beberapa *outliers*. Kontaminasi *outliers* ini akan berdampak pada estimasi rata-rata pengeluaran rumah tangga tingkat kecamatan. Kontaminasi *outliers* dan distribusi yang tidak normal pada data pengeluaran rumah tangga per kapita diduga akan sangat mempengaruhi performa metode SAE konvensional terutama apabila komponen *random effect* yang dihasilkan juga terkontaminasi *outliers* atau melanggar asumsi normalitas

Sumber: Susenas 2018, diolah

Gambar 1. Grafik Distribusi Data Pengeluaran Rumah Tangga Per Kapita dalam Sebulan di Provinsi DI Yogyakarta Tahun 2018

Sumber: Susenas 2018, diolah

Gambar 2. *Boxplot* Data Pengeluaran Rumah Tangga Per Kapita dalam Sebulan di Provinsi DI Yogyakarta Tahun 2018

Pemilihan Variabel Penyerta

Sebelum dilakukan penghitungan rata-rata pengeluaran per kapita dengan menggunakan metode SAE, terlebih dahulu dilakukan pemilihan kandidat variabel penyerta. Kandidat variabel penyerta hasil sintesis dari penelitian terkait yang terdapat pada Tabel 1 selanjutnya dilakukan penghitungan korelasi pearson terhadap variabel pengeluaran rumah tangga per kapita.

Metode SAE yang berbasis pemodelan *linear M-quantile* memerlukan terpenuhinya asumsi hubungan linear antara variabel respon dan variabel penyerta dalam setiap *quantile* didalam distribusi bersyaratnya (Fabrizi et al 2014; Fabrizi et al 2014). Oleh sebab itu, untuk menjaga terpenuhinya asumsi tersebut, kandidat variabel penyerta dengan korelasi pearson lebih dari 0,5 dipilih sebagai variabel penyerta. Berdasarkan hasil penghitungan terdapat 6 kandidat variabel penyerta yang memiliki nilai korelasi pearson diatas 0,5 diantaranya variabel X1, X2, X7, X9, X13, dan X14. Penentuan variabel penyerta mana yang terbaik untuk dimasukkan kedalam model sangat bergantung pada kebaikan model SAE yang diukur melalui indikator RSE. Hal ini berarti, memasukan seluruh variabel penyerta kedalam model tidak serta merta akan mengakibatkan performa model menjadi paling optimal. Oleh karena itu, perlu memperhitungkan berbagai kombinasi dari keenam variabel penyerta terpilih agar didapatkan kombinasi yang paling mengoptimalkan RSE metode SAE. Setelah dilakukan penghitungan, didapatkan kombinasi variabel penyerta yang paling mengoptimalkan RSE adalah X9, X13, dan X14.

Komponen *Random Area Effect* dan *Sampling Error* pada Estimasi Rata-rata Pengeluaran Rumah Tangga Per Kapita Hasil Pemodelan SAE EBLUP-FH

Pada estimasi SAE konvensional seperti EBLUP-FH, kedua komponen *random effect* harus memenuhi asumsi normalitas (Rao & Molina, 2015). Oleh sebab itu, perlu dilakukan pemeriksaan kontaminasi *outliers* dan pemenuhan asumsi normalitas pada kedua komponen *random effect* yang dihasilkan oleh metode EBLUP-FH.

Untuk melihat komponen *random area effect* dan *sampling error* yang dihasilkan metode EBLUP-FH mengalami kontaminasi *outliers* atau tidak, dilakukan visualisai data menggunakan *boxplot* untuk menunjukan observasi yang terindikasi sebagai *outliers*.

Gambar 3. Boxplot random area effect

Gambar 4. Boxplot sampling error

Berdasarkan hasil pada Gambar 3 dan Gambar 4 menunjukkan bahwa baik komponen *random area effect* maupun *sampling error* terindikasi terkontaminasi *outliers*. Selanjutnya untuk menindaklanjuti observasi yang diduga *outliers* tersebut secara statistik adalah merupakan *outliers*, maka dilakukan uji *outliers* dengan uji grubbs dan dilakukan uji asumsi normalitas dengan menggunakan uji Shapiro-wilk. Hasil uji grubbs dan uji Shapiro-wilk disajikan pada Tabel 3 berikut.

Tabel 2. Hasil uji Grubbs dan uji Shapiro-Wilk

Komponen	Uji grubbs		Tingkat Kontaminasi	Uji Shapiro-wilk		Hasil
	Jumlah Observasi diduga <i>Outliers</i>	Jumlah Observasi Signifikan <i>Outliers</i>		Statistik	<i>p-value</i>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Random area effect</i>	1	1	1,35%	0,97	0,051	Mengikuti distribusi normal
<i>Sampling error</i>	12	1	14,86%	0,60	0,000	Tidak mengikuti distribusi normal

Keterangan: uji Grubbs dan Shapiro-wilk dilakukan pada taraf signifikansi 5%

Tabel 2 menunjukkan dengan taraf signifikansi 5%, komponen *random area effect* terkontaminasi *outliers* sebesar 1,35% dan memenuhi asumsi normalitas sementara *sampling error* terkontaminasi *outliers* sebesar 14,86% dan melanggar asumsi normalitas. Tidak terbebasnya komponen *random effect* yang dihasilkan EBLUP-FH dari *outliers* tentu akan sangat berdampak pada tingkat presisi yang dihasilkan. Tidak terpenuhinya asumsi normalitas pada *sampling error* tentunya dapat menyebabkan presisi yang dihasilkan EBLUP-FH tidak sebaik presisi yang dihasilkan saat kedua komponen *random effect* memenuhi asumsi normalitas.

Perbandingan Estimasi Langsung dan Estimasi Tidak Langsung Metode SAE Rata-rata Pengeluaran Rumah Tangga Per Kapita Tingkat Kecamatan DI Yogyakarta Tahun 2018

Estimasi langsung menggunakan desain Susenas sedangkan estimasi tidak langsung berbasis pemodelan SAE menggunakan metode EBLUP-FH dan *M-quantile CD* dalam mengestimasi rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan DI Yogyakarta tahun 2018 Selanjutnya ditunjukkan hasil perbandingan estimasi metode estimasi langsung dan estimasi tak langsung berbasis pemodelan SAE dalam bentuk statistik deskriptif pada Tabel 3 berikut.

Terdapat perbedaan identifikasi kecamatan dengan rata-rata pengeluaran rumah tangga per kapita tertinggi dan terendah antara hasil estimasi langsung dan tidak langsung. Kecamatan Depok yang terletak di Kabupaten Sleman diidentifikasi sebagai kecamatan dengan rata-rata pengeluaran rumah tangga per kapita tertinggi menurut EBLUP-FH sebesar Rp. 3.172.364 sedangkan berdasarkan *M-quantile CD* adalah Kecamatan Pakualaman di Kota Yogyakarta dengan rata – rata sebesar Rp. 2.821.211. Sementara berdasarkan estimasi langsung, kecamatan dengan rata – rata pengeluaran per kapita tertinggi adalah Kecamatan Ngaglik yaitu sebesar Rp. 3.060.216. Berbeda dengan estimasi tertinggi, pada estimasi terendah, metode EBLUP-FH mengidentifikasi Kecamatan Girimulyo yang berada di Kabupaten Kulon Progo sebagai kecamatan dengan

rata-rata pengeluaran rumah tangga per kapita terendah yaitu Rp. 505.629, Hasil tersebut sama dengan yang diidentifikasi oleh estimasi langsung dengan besar Rp. 477.237. Sementara itu, metode *M-quantile CD* mengidentifikasi Kecamatan Panggang yang terletak di Kabupaten Gunung Kidul sebagai kecamatan dengan rata-rata pengeluaran rumah tangga per kapita terendah dengan besar Rp. 2.821.211.

Tabel 3. Statistik deskriptif estimasi langsung dan tidak langsung rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan DI Yogyakarta tahun 2018

Statistik	Estimasi Langsung	Estimasi Tak Langsung	
		EBLUP-FH	<i>M-quantile CD</i>
(1)	(2)	(3)	(5)
Minimum	Rp. 477.237	Rp. 505.629	Rp. 471.259
Median	Rp. 1.069.362	Rp. 1.063.822	Rp. 1.030.034
Mean	Rp. 1.322.941	Rp. 1.223.023	Rp. 1.165.006
Maksimum	Rp. 3.060.216	Rp. 3.172.364	Rp. 2.821.211
Standar Deviasi	Rp. 640.864	Rp. 504.684	Rp. 505.196
Range	Rp. 2.582.979	Rp. 2.666.735	Rp. 2.349.952

Perbandingan Performa Metode Estimasi Langsung dan Estimasi Tidak Langsung Metode SAE dalam Menduga Rata-rata Pengeluaran Rumah Tangga Per Kapita Tingkat Kecamatan DI Yogyakarta Tahun 2018

Perbandingan performa metode estimasi dievaluasi dengan RSE sebagai gambaran tingkat presisi dan reliabilitas estimasi yang dihasilkan. Hasil RSE divisualisasikan pada **Gambar 5** dan **Gambar 6**.

Secara umum performa yang dihasilkan oleh estimasi tidak langsung metode SAE mampu memperbaiki tingkat presisi yang dihasilkan estimasi langsung. Selaras dengan hal tersebut, estimasi tidak langsung metode SAE dapat memperbaiki reliabilitas estimasi langsung. Rao & Molina (2015) menjelaskan bahwa dengan meminjam kekuatan dari area sekitar metode estimasi tidak langsung dapat memperbaiki reliabilitas estimasi langsung pada tingkat presisi yang reliabel.

Sementara itu, secara median metode *M-quantile CD* menghasilkan tingkat presisi tertinggi. Hal tersebut dapat dilihat dari *boxplot M-quantile CD* memiliki garis median yang terletak paling bawah. Meskipun demikian, tingkat presisi yang dihasilkan tidak lebih stabil jika dibandingkan EBLUP-FH. Hal ini dapat terjadi karena ukuran sampel yang terlalu kecil. Sebagaimana yang dijelaskan oleh Rao & Molina (2015) bahwa varians dari *M-quantile small area model* akan menjadi sangat tidak stabil apabila ukuran sampelnya sangat kecil. Ini disebabkan *estimator* varians (5) pada metode *M-quantile* hanya bergantung pada ukuran sampel domain ke- j , n_j . Namun demikian, metode *M-quantile CD* tetap dapat memperbaiki presisi EBLUP-FH untuk 49 kecamatan dari 74 kecamatan yang diestimasi.

Gambar 5. Grafik garis RSE estimasi langsung dan estimasi tidak langsung metode SAE

Pada kondisi ini, secara umum tingkat presisi yang dihasilkan oleh EBLUP-FH masih lebih baik dari estimasi langsung. Meskipun demikian, jika dilihat dengan teliti, tingkat presisi yang dihasilkan berada sangat dekat dengan tingkat presisi estimasi langsung di sepanjang domain. Hal ini dapat terlihat dari nilai RSE yang sangat berdekatan dan pola pergerakannya cenderung mengikuti estimasi langsung. Dengan demikian, beberapa temuan tersebut menunjukkan performa EBLUP-FH tidak optimal dalam mengestimasi rata-rata pengeluaran rumah tangga per kapita, akan tetapi model *M-quantile CD* berhasil menghasilkan estimasi yang paling efisien meskipun terdapat kontaminasi *outliers* maupun pelanggaran asumsi normalitas pada komponen *sampling error*. Hal ini dibuktikan dengan tingkat presisi yang dihasilkan secara median maupun rata-rata lebih tinggi dibandingkan tingkat presisi EBLUP-FH. Selain itu, model *M-quantile CD* mampu memperbaiki reliabilitas estimasi langsung pada tingkat yang lebih baik dibandingkan EBLUP-FH. Dengan demikian, metode yang paling baik untuk mengestimasi rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan di DI Yogyakarta tahun 2018 adalah metode *M-quantile CD*.

Gambar 6. Boxplot RSE estimasi langsung dan estimasi tidak langsung metode SAE

KESIMPULAN

Data pengeluaran rumah tangga per kapita DI Yogyakarta tahun 2018 terkontaminasi banyak *outliers* dan sebaran data yang dihasilkan menceng kanan. Hal ini menyebabkan komponen *sampling error* yang dihasilkan EBLUP-FH terkontaminasi *outliers* dan melanggar asumsi normalitas. Kemudian pemodelan SAE *M-quantile CD* menghasilkan tingkat presisi paling tinggi. Hal ini berarti penggunaan model *M-quantile CD* lebih efisien dalam mengestimasi rata-rata pengeluaran rumah tangga per kapita tingkat kecamatan di DI Yogyakarta.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2019). *Estimasi Upah Rata-rata Per jam Kerja Pada Level Kabupaten Kota dengan Menggunakan Metode SAE*. Badan Pusat Statistik.
- Barigozzi, M., Alessi, L., Capasso, M., & Fagiolo, G. (2009). The distribution of households consumption-expenditure budget shares. In *European Central Bank, Working Paper Series* (Vol. 1, Issue 1061).
- Battistin, E., Blundell, R., & Lewbel, A. (2009). Why is consumption more log normal than income? Gibrat's law revisited. *Journal of Political Economy*, 117(6), 1140–1154.
- Chambers, R., & Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, 93(2), 255–268.
- Fabrizi, E., Giusti, C., Salvati, N., & Tzavidis, N. (2014). Mapping average equivalized income using robust small area methods. *Papers in Regional Science*, 93(3), 685–701.
- Fabrizi, E., Salvati, N., Pratesi, M., & Tzavidis, N. (2014). Outlier robust model-assisted small area estimation. *Biometrical Journal*, 56(1), 157–175.
- Fadhila, Z. (2019). *Kajian Pendugaan Area Kecil Melalui Pemodelan Regresi M-Kuantil pada Data yang Mengandung Pencicilan*. Tesis Institut Pertanian Bogor. Bogor. 71 hlm.

- Girinoto. (2017). *Kajian Pendugaan Area Kecil untuk Indikator Kemiskinan Melalui Pendekatan Regresi Kekar M-Kuantil (Studi Kasus: Pendugaan Tingkat Kecamatan pada Kabupaten Bogor)*. Tesis Institute Pertanian Bogor. Bogor. 35 hlm.
- Molina, I., & Marhuenda, Y. (2015). Sae: An R package for small area estimation. *R Journal*, 7(1), 81–98.
- Pratesi, M., Giusti, C., Marchetti, S., & Salvati, N. (2015). Small Area Estimation models with outliers in covariates. *New Techniques and Technologies for Statistics*, 1–4.
- Pratesi, M., & Giusti, C. (2015). Small Area Estimation. *Proceeding Conference at 1st EMOS Spring School*. Pisa. https://ec.europa.eu/eurostat/cros/system/files/Lesson1_SAE%20EMOSSchool_last.pdf
- Rao, J. N. . (2003). *Small Area Estimation* (N. S. Robert M. Groves, Graham Kalton, J N. K. Rao & C. Skinner (eds.); 1st ed.).
- Rao, J. N. ., & Molina, I. (2015). *Small Area Estimation*. In Wiley Series in Survey Methodology (2nd ed.).
- RI (Republik Indonesia). (2004). Undang-Undang No. 32 Tahun 2004 tentang Pemerintahan Daerah. Lembaran Negara RI Tahun 2004, No. 4437. Sekretariat Negara. Jakarta.
- Tzavidis, N., Salvati, N., Pratesi, M., & Chambers, R. (2007). M-quantile models with application to poverty mapping. *Stat Meth Appl*, 17, 393–411.
- Tzavidis, N., Marchetti, S., & Chambers, R. (2010). Robust estimation of small-area means and quantiles. *Australian and New Zealand Journal of Statistics*, 52(2), 167–186.
- Tzavidis, N., Zhang, L. C., Luna, A., Schmid, T., & Rojas-Perilla, N. (2018). From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 181(4), 927–979.