

KAJIAN PENGARUH PENAMBAHAN INFORMASI GEROMBOL TERHADAP PREDIKSI AREA NIRCONTOH PADA DATA BINOMIAL*

Beny Trianjaya¹, Anang Kurnia^{2‡}, and Agus M Soleh³

¹Badan Pusat Statistik, Indonesia, benyt84@gmail.com

²Department of Statistics, IPB University, Indonesia, anangk@apps.ipb.ac.id

³Department of Statistics, IPB University, Indonesia, agusms@apps.ipb.ac.id

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 4 No 4 (2020), 566 - 578

Copyright © 2020 Beny Trianjaya, Anang Kurnia, and Agus M Soleh. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Employment data is one of the important indicators related to the development progress of a country. Labor conditions in the territory of Indonesia can only be compared between times through the Survei Angkatan Kerja Nasional (Sakernas) data. Data generated from Sakernas and published by BPS is the number of employed and unemployed. The obstacle in estimating the semester unemployment rate at the regency/municipality level is the lack of a number of examples. One of the indirect estimates currently developing is small area estimation (SAE). This study developed the generalized linear mixed model (GLMM) by adding cluster information and examines the development of modifications with several model scenarios. The purpose of this study was to develop a prediction model for basic GLMM on a small area approach by adding cluster information as a fixed effect or random effect. The simulation results show that Model-2, a model that adds a fixed effect k-cluster and also adds a mean from the estimated effect of random areas in the sample area, is the best model with the smallest relative bias (RB) and Relative root mean squares error (RRMSE). This model is better than the basic GLMM model (Model-0) and Model-1 (a model which only adds a mean from the estimated random effect area in the sample area). Model-2 is applied to estimate the proportion of unemployed sub-district level in Southeast Sulawesi Province. Estimating the proportion of unemployed with calibration Model-2 produced an estimated aggregation of the unemployment proportion of Southeast Sulawesi Province at 0.0272. These results are similar to BPS (0.0272). Thus, the results of the estimated proportion of unemployment at the sub-district level with a calibration Model-2 can be said to be feasible to use.

Keywords: binomial, clustering, glmm, small-area, unemployment.

* Received Jun 2019; Accepted Des 2020; Published online on Des 2020

1. Pendahuluan

Data ketenagakerjaan merupakan salah satu indikator penting terkait kemajuan pembangunan suatu negara. Kondisi ketenagakerjaan merupakan bentuk riil di masyarakat mengenai kemampuan negara dalam meningkatkan kesejahteraan rakyatnya, terutama dalam hal mencari nafkah demi pemenuhan kebutuhan hidup. Badan Pusat Statistik (BPS) adalah lembaga negara yang bertanggung jawab terhadap pengumpulan dan ketersediaan data nasional di Indonesia, termasuk memublikasikan data-data mengenai ketenagakerjaan. Sensus atau survei yang terkait dengan data ketenagakerjaan antara lain: Sensus Penduduk (SP), Survei Penduduk Antar Sensus (SUPAS), Survei Sosial Ekonomi Nasional (Susenas), dan Survei Angkatan Kerja Nasional (Sakernas). Kondisi ketenagakerjaan di wilayah Indonesia hanya dapat diperbandingkan antar waktu melalui data Sakernas.

Kegiatan Sakernas pertama kali dilaksanakan pada tahun 1976. Sakernas mengalami berbagai perubahan baik waktu pelaksanaan, level estimasi, cakupan, maupun metodologi ([BPS], 2016). Sejak tahun 2015, Sakernas dilaksanakan setiap semester, yaitu pada bulan Februari dan Agustus. Sakernas yang dilaksanakan pada bulan Februari dirancang hanya mampu melakukan pendugaan tingkat nasional dan provinsi. Sedangkan untuk dapat melakukan pendugaan tingkat kabupaten/kota dilaksanakan Sakernas pada bulan Agustus. Jumlah contoh pada Sakernas Agustus empat kali lebih banyak dibandingkan Sakernas Februari. Pada tahun 2016, penambahan contoh tidak dilakukan pada pelaksanaan Sakernas Agustus. Jumlah contoh yang lebih sedikit ini mengakibatkan pendugaan sampai tingkat kabupaten/kota pada tahun tersebut tidak dapat dilakukan. Dampak yang ditimbulkan akibat permasalahan ini adalah data ketenagakerjaan tahun 2016 tidak tersedia hingga tingkat kabupaten/kota di seluruh provinsi di Indonesia.

Provinsi Sulawesi Tenggara merupakan wilayah dengan Tingkat Pengangguran Terbuka (TPT) terkecil se-Sulawesi pada periode Agustus 2016, yaitu sebesar 2.72 persen. Nilai TPT ini merupakan nilai terendah se-pulau Sulawesi dan terendah keempat secara nasional.

Kendala dalam melakukan pendugaan angka pengangguran semesteran pada tingkat kabupaten/kota adalah kurangnya jumlah contoh. Menurut Sadik (2009), ukuran contoh yang sangat kecil akan mengakibatkan ragam penduga menjadi besar dan jika pada suatu area tidak terdapat unit contoh maka proses pendugaan parameter tidak dapat dilakukan. Untuk itu diperlukan suatu metode pendugaan tidak langsung yang mampu mengurangi ragam penduga.

Salah satu pendugaan tidak langsung yang sedang berkembang saat ini adalah pendugaan area kecil (*small area estimation/SAE*). Model pendugaan area kecil merupakan alternatif yang dapat digunakan untuk menduga parameter pada suatu area ketika ukuran contoh pada area tersebut terlalu kecil untuk mendapatkan presisi yang memadai dibandingkan ketika dilakukan pendugaan secara langsung (Rao & Molina, 2015). Metode ini memanfaatkan informasi dari luar dan dalam area itu sendiri ataupun dari luar survei untuk menduga parameter pada area yang relatif kecil dalam percontohan survei (Kurnia, 2009). Informasi tersebut berupa peubah penyerta (*auxiliary variable*) yang mempunyai hubungan dengan peubah yang menjadi tujuan amatan. Peubah tersebut dapat diambil dari sumber pendataan lain seperti sensus,

pencatatan administrasi atau survei lain.

Data yang dihasilkan dari Sakernas dan dipublikasikan oleh BPS adalah jumlah yang bekerja dan menganggur (dua kategori). Kejadian yang terdiri dari dua kemungkinan disebut dengan biner. Respon biner dengan peluang berhasil, π , dan ulangan (n) pada umumnya mengikuti sebaran Binomial (n, π), selanjutnya angka pengangguran diasumsikan mengikuti sebaran Binomial. Penelitian tentang data binomial dengan pendekatan area kecil telah dilakukan sebelumnya oleh Yudistira et al. (2018).

Pendugaan area kecil dapat digunakan untuk menduga area yang tidak terkena contoh. Anisa et al., (2014) melakukan kajian pendugaan area kecil dengan melibatkan pengaruh informasi gerombol terhadap prediksi area nircontoh. Metode yang digunakan adalah empirical best linear unbiased prediction (EBLUP) dengan beberapa skenario model penambahan informasi gerombol. Sundara et al. (2017) melakukan hal yang serupa, namun dengan menggunakan metode Empirical Bayes terhadap penelitiannya. Kedua penelitian menyimpulkan bahwa penambahan informasi gerombol menghasilkan prediksi yang lebih baik untuk area yang tidak terdapat contoh (nircontoh).

Penelitian ini mengembangkan model linier campuran terampat (MLCT)/generalized linear mixed model (GLMM) dengan menambahkan informasi gerombol dan mengkaji pengembangan dengan beberapa skenario model. Beberapa pengembangan model yang dikembangkan digunakan untuk menduga angka pengangguran tingkat kabupaten/kota di Provinsi Sulawesi Tenggara. Pendekatan yang dilakukan adalah metode pendugaan area kecil dengan melibatkan unit yang tersurvei maupun yang tidak tersurvei (nircontoh).

Tujuan yang ingin dicapai dari penelitian ini yaitu:

1. Mengembangkan model prediksi MLCT dasar pendekatan area kecil dengan menambahkan informasi gerombol sebagai pengaruh tetap maupun pengaruh acak.
2. Menduga angka pengangguran tingkat kecamatan di Provinsi Sulawesi Tenggara menggunakan model dasar dan model yang dikembangkan dengan melibatkan unit nircontoh.
3. Mengkaji kebaikan model yang dikembangkan dibandingkan dengan model dasar yaitu MLCT dasar.

2. Metodologi

2.1 Pengembangan Model

Model dasar yang digunakan pada penelitian ini yaitu Model Linear Campuran Terampat (MLCT) yang selanjutnya disebut sebagai Model-0 atau model dasar yang didefinisikan sebagai berikut:

a) Model untuk populasi:

$$\text{logit}(\pi_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i \quad (1)$$

b) Model prediksi untuk area contoh:

$$\text{logit}(\tilde{\pi}_{ij}) = \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}} + \hat{v}_i \quad (2)$$

c) Model prediksi untuk area nircontoh:

$$\text{logit}(\tilde{\pi}_{i^*j^*}) = \mathbf{x}'_{i^*j^*}\hat{\boldsymbol{\beta}} \quad (3)$$

Dalam hal ini π_{ij} merupakan proporsi dari peubah yang menjadi perhatian, y_{ij} , pada

area ke- i dan unit ke- j yang terkena contoh, x_{ij} merupakan vektor peubah penyerta berukuran $p \times 1$, p adalah banyaknya peubah penyerta, $\hat{\beta}$ adalah koefisien model pengaruh tetap area, dan v_i merupakan pengaruh acak area. Penduga nilai yang menjadi perhatian pada area nircontoh ke- i^* dan unit nircontoh ke- j ($\tilde{\pi}_{i^*j^*}$) diperoleh dari informasi peubah penyerta yang terdapat pada unit nircontoh. Model ini mengasumsikan bahwa $v_i \sim iid N(0, \sigma_v^2)$.

Pengembangan model dilakukan dengan menambahkan informasi gerombol ke- k pada Model-0 (Anisa *et al.*, 2014). Informasi gerombol yang ditambahkan kepada model dasar menghasilkan pengembangan dua model yang selanjutnya diusulkan pada penelitian ini. Kedua model tersebut adalah:

1. Model-1, yaitu modifikasi Model-0 dengan menambahkan nilai tengah dari pendugaan pengaruh acak area pada masing-masing gerombol pada model prediksi untuk area yang tidak terdapat contoh. Penambahan tersebut dinyatakan sebagai $\tilde{v}_{(k)} = \frac{1}{m_k} \sum_{i=1}^{m_k} \hat{v}_i$, dengan m_k adalah banyaknya area contoh pada gerombol ke- k .

a) Model untuk populasi:

$$\text{logit}(\pi_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i \tag{4}$$

b) Model prediksi untuk area contoh:

$$\text{logit}(\tilde{\pi}_{ij(k)}) = \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}} + \hat{v}_i \tag{5}$$

c) Model prediksi untuk area nircontoh:

$$\text{logit}(\tilde{\pi}_{i^*j^*(k)}) = \mathbf{x}'_{i^*j^*}\hat{\boldsymbol{\beta}} + \tilde{v}_{(k)} \tag{6}$$

2. Model-2, yaitu modifikasi Model-1 dengan menambahkan pengaruh tetap gerombol ke- k . Pengaruh tetap gerombol dinyatakan sebagai \hat{C}_k yang merupakan bentuk sederhana dari notasi penduga koefisien peubah boneka untuk setiap gerombol. Sehingga untuk sejumlah k gerombol dapat diuraikan menjadi $\hat{C}_k = \hat{a}_1d_1 + \dots + \hat{a}_{k-1}d_{k-1}$, dengan d_1, \dots, d_{k-1} merupakan peubah boneka untuk gerombol dan $\hat{a}_1, \dots, \hat{a}_{k-1}$ merupakan penduga koefisien bagi peubah boneka.

a) Model untuk populasi:

$$\text{logit}(\pi_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + C_k + v_i \tag{7}$$

b) Model prediksi untuk area contoh:

$$\text{logit}(\tilde{\pi}_{ij(k)}) = \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}} + \hat{C}_k + \hat{v}_i \tag{8}$$

c) Model prediksi untuk area nircontoh:

$$\text{logit}(\tilde{\pi}_{i^*j^*(k)}) = \mathbf{x}'_{i^*j^*}\hat{\boldsymbol{\beta}} + \hat{C}_k + \tilde{v}_{(k)} \tag{9}$$

2.2 Kajian Simulasi

Data simulasi pada penelitian ini menggambarkan populasi buatan dari data yang mengandung distribusi binomial. Data tersebut diusahakan agar memiliki kemiripan dengan data sesungguhnya. Oleh karena itu, data simulasi dibangkitkan berdasarkan statistik yang menggambarkan data pengangguran dan data pendukung lainnya di Provinsi Sulawesi Tenggara.

Simulasi bertujuan untuk mengevaluasi kebaikan model dasar dan beberapa model yang dikembangkan. Proses simulasi dilakukan berdasarkan langkah-langkah berikut:

1. Menetapkan jumlah area populasi sebanyak 209 area, M, yaitu banyaknya kecamatan di Provinsi Sulawesi Tenggara.

2. Menetapkan jumlah area contoh sebanyak 69 area, m, yaitu banyaknya kecamatan yang menjadi contoh.
3. Menetapkan banyaknya unit populasi sebanyak 2272 unit, N, yaitu banyaknya desa di Provinsi Sulawesi Tenggara yang menyebar di seluruh area.
4. Menetapkan banyaknya unit yang menjadi contoh sebanyak 98 unit, n, yaitu banyaknya desa yang menjadi contoh yang terbagi di masing-masing area yang menjadi contoh. Banyaknya unit yang menjadi contoh tiap area mengikuti data Sakernas Agustus 2016 Provinsi Sulawesi Tenggara.
5. Menentukan jumlah gerombol sebanyak tiga gerombol, berdasarkan penggerombolan data terapan.
6. Menetapkan parameter pengaruh efek tetap yang diambil berdasarkan pemodelan dengan MLCT baku (Model-0). Parameter ini memiliki nilai yang berbeda untuk setiap gerombol ke- k , β_k . Nilai β_k untuk setiap gerombol dapat dilihat pada Tabel 1.

Tabel 1: Titik potong dan kemiringan pada populasi untuk gerombol ke- k .

Gerombol	β_{0k}	β_{1k}
1	-4.3	0.1
2	-3.3	1.6
3	-2.3	3.1

7. Membangkitkan peubah penyerta, X_1 , yang berasal dari sebaran seragam (0,1) sebanyak jumlah desa populasi (2272). Peubah penyerta yang digunakan mengacu kepada data proporsi desa dengan listrik PLN.
8. Membangkitkan jumlah angkatan kerja contoh, AKs, yang menyebar binomial dengan peluang=0.67 dan banyak ulangan=30. Angka ini berdasarkan nilai Angkatan Kerja (AK) pada data contoh Sakernas Agustus 2016 di Provinsi Sulawesi Tenggara.
9. Membangkitkan pengaruh acak area, v , sebanyak jumlah area. Terdapat dua skenario untuk melihat pengaruh penggerombolan.
 - a. Skenario 1, populasi dengan gerombol terpisah secara sempurna. Pengaruh acak area dibangkitkan berdasarkan sebaran normal $v_i \sim iid N(0,0.33)$. Skenario ini dilakukan untuk melihat kemampuan model pada kondisi populasi yang ideal, yaitu karakteristik antar gerombol mampu dibedakan dengan baik.
 - b. Skenario 2, populasi dengan gerombol tidak terpisah secara sempurna. Pengaruh acak area dibangkitkan berdasarkan sebaran normal $v_i \sim iid N(0,3.3)$. Skenario ini dilakukan untuk melihat kemampuan model pada kondisi populasi yang tidak ideal, yaitu saat terjadi penggerombolan yang beririsan.
10. Menghitung nilai θ_{ij} untuk masing-masing unit berdasarkan persamaan (8).
11. Menghitung nilai π_{ij} untuk masing-masing unit. Nilai π_{ij} diperoleh dengan menggunakan persamaan:

$$\pi_{ij} = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})} \quad (10)$$

12. Membangkitkan nilai peubah yang menjadi perhatian (jumlah pengangguran) untuk setiap unit contoh, y_{ij} , yang menyebar binomial dengan peluang= π_{ij} dan ulangan=angkatan kerja contoh (AKs).
13. Membangkitkan jumlah angkatan kerja populasi, AKp , yang menyebar normal dengan nilai tengah 492 dan simpangan baku 531. Jumlah ini berdasarkan angkatan kerja di Provinsi Sulawesi Tenggara tahun 2016.
14. Melakukan pendugaan parameter untuk area ke- i dan unit ke- j , $\tilde{\theta}_{ij}$, untuk masing-masing skenario dan model. Sehingga diperoleh $\tilde{\theta}_{ij}$ untuk setiap model pendugaan.
 - a. Dugaan proporsi Model-0 diperoleh dengan mengalikan peubah penyerta dan parameter pengaruh tetap (β_0).
 - b. Dugaan proporsi Model-1 diperoleh dengan menambahkan hasil dugaan pada Model-0 dengan pengaruh acak yang diperoleh dari Model-1.
 - c. Dugaan proporsi Model-2 diperoleh dengan cara mengalikan peubah penyerta dan parameter pengaruh tetap yang didapatkan dari Model-2, kemudian menambahkannya dengan pengaruh acak yang didapatkan dari Model-2.
15. Menghitung nilai $\tilde{\pi}_{ij}$ untuk masing-masing skenario dan model yang merupakan proporsi hasil transformasi balik dari $\tilde{\theta}_{ij}$.

$$\tilde{\pi}_{ij} = \frac{\exp(\tilde{\theta}_{ij})}{1 + \exp(\tilde{\theta}_{ij})} \tag{11}$$

16. Menghitung nilai \tilde{y}_{ij} untuk masing-masing skenario dan model.

$$\tilde{y}_{ij} = AKp_{ij} \times \tilde{\pi}_{ij} \tag{12}$$

17. Melakukan agregasi proporsi pengangguran pada tingkat area, $\tilde{\pi}_i$, untuk masing-masing skenario dan model

$$\tilde{\pi}_i = \frac{\sum_{\forall j} \tilde{y}_{ij}}{\sum_{\forall j} AKp_{ij}} \tag{13}$$

18. Mengulangi proses pada langkah (7) sampai (17) sebanyak $B=1000$ kali sehingga dapat dihitung *relative bias* (RB) dan *relative root mean squares error* (RRMSE) dari hasil pendugaan parameter setiap area dengan rumus sebagai berikut:

$$RB_i = \frac{1}{B} \sum_{b=1}^B \left(\frac{\hat{\pi}_{ib} - \pi_i}{\pi_i} \right) \times 100\% \tag{14}$$

$$RRMSE_i = \frac{1}{\pi_i} \sqrt{\frac{1}{B} \sum_{b=1}^B \left(\frac{\hat{\pi}_{ib} - \pi_i}{\pi_i} \right)^2} \times 100\% \tag{15}$$

19. Menghitung rata-rata RB dan RRMSE.

$$RB = \frac{1}{m} RB_i \tag{16}$$

$$RRMSE = \frac{1}{m} RRMSE_i \tag{17}$$

20. Mengevaluasi model berdasarkan nilai RB dan RRMSE.

2.3 Kajian Aplikasi

Bahan dan Data

Data yang digunakan dalam penelitian ini bersumber dari data Sakernas Agustus 2016 untuk peubah yang menjadi perhatian dan Podes 2014 untuk peubah penyerta atau *auxiliary variable* (Tabel 2). Jumlah unit penelitian adalah sebanyak 2272 desa/kelurahan di Provinsi Sulawesi Tenggara, terdiri dari 98 (j) desa/ kelurahan yang tersurvei dan 2174 (j^*) desa/kelurahan tidak tersurvei pada Sakernas Agustus 2016. Parameter yang akan diduga (\tilde{y}_{ij}) adalah jumlah pengangguran di 2272 desa/kelurahan, dengan i adalah indeks untuk kecamatan, $i = 1, \dots, 69$. Notasi j merupakan indeks untuk setiap desa, $j = 1, \dots, 98$. Berdasarkan dugaan jumlah pengangguran tersebut, kemudian dihitung proporsi pengangguran di 209 kecamatan di Provinsi Sulawesi Tenggara.

Tabel 2: Peubah-peubah yang digunakan dalam penelitian.

	Nama Peubah	Sumber
Y	Jumlah pengangguran	Sakernas 2016
X ₁	Persentase penduduk laki-laki	Podes 2014
X ₂	Persentase keluarga pertanian	Podes 2014
X ₃	Rasio sarana pendidikan per 100 penduduk	Podes 2014
X ₄	Proporsi rumah tangga pengguna listrik PLN	Podes 2014
X ₅	Proporsi surat keterangan miskin yang dikeluarkan pemerintah desa	Podes 2014
X ₆	Rasio industri mikro dan kecil per 100 penduduk	Podes 2014
X ₇	Rasio pasar (bangunan permanen dan semi permanen) per jumlah penduduk	Podes 2014
X ₈	Rasio kedai makanan per jumlah penduduk	Podes 2014
X ₉	Rasio toko per jumlah penduduk	Podes 2014
X ₁₀	Rasio hotel/penginapan per jumlah penduduk	Podes 2014

Metode Analisis

Langkah-langkah yang dilakukan dalam kajian aplikasi ini adalah sebagai berikut :

1. Melakukan eksplorasi data, memeriksa peubah-peubah yang berkorelasi terhadap peubah yang menjadi perhatian, serta melakukan pemilihan peubah yang mampu membedakan karakteristik setiap gerombol.
2. Menggerombolkan area berdasarkan peubah-peubah yang dipilih dari data PODES 2014.
3. Melakukan pemodelan dengan model dasar MLCT dan dua pengembangan model.
4. Menduga nilai proporsi pengangguran setiap desa, baik terkena contoh maupun tidak, untuk semua model (3 model).
 - a. Menduga parameter model
 - b. Menghitung dugaan logit proporsi ($\tilde{\theta}_{ij}$)
 - c. Menghitung dugaan proporsi pengangguran tingkat desa ($\tilde{\pi}_{ij}$) dengan melakukan transformasi balik terhadap hasil langkah b.

$$\tilde{\pi}_{ij} = \frac{\exp(\tilde{\theta}_{ij})}{1 + \exp(\tilde{\theta}_{ij})} = \frac{\exp(\text{logit}(\tilde{\pi}_{ij}))}{1 + \exp(\text{logit}(\tilde{\pi}_{ij}))}. \quad (18)$$

5. Melakukan agregasi untuk mendapatkan nilai proporsi pengangguran kecamatan ($\tilde{\pi}_i$) dan provinsi.
6. Mengevaluasi hasil pendugaan dari model-model yang dibentuk dengan membandingkan *root mean squares error* (RMSE) dugaan kecamatan yang dihasilkan dari masing-masing model.

$$RMSE((\tilde{\pi}_i) = \sqrt{MSE(\tilde{\pi}_i)}. \quad (19)$$

7. Membandingkan nilai agregasi tingkat provinsi dengan angka proporsi pengangguran yang dikeluarkan oleh BPS. Jika hasilnya berbeda jauh, maka lakukan kalibrasi pembeda (*difference*) dengan formula (Ghaniswati, 2019):

$$\tilde{\pi}_{ij}^{DB} = \tilde{\pi}_{ij} + \left(\sum_{p=1}^M \sum_{q=1}^N W_{pq} \hat{\pi}_{pq} - \sum_{p=1}^M \sum_{q=1}^N W_{pq} \tilde{\pi}_{pq} \right). \quad (20)$$

Nilai M adalah banyaknya kecamatan di Provinsi Sulawesi Tenggara, nilai N adalah banyaknya desa di Provinsi Sulawesi Tenggara, $\sum_{p=1}^M \sum_{q=1}^N W_{pq} \hat{\pi}_{pq}$ adalah dugaan langsung tingkat provinsi, W_{pq} adalah rasio antara penduduk angkatan kerja kecamatan ke- p desa ke- q dengan total angkatan kerja tingkat provinsi, $\tilde{\pi}_{ij}$ adalah dugaan proporsi pengangguran kecamatan ke- i desa ke- j dan $\tilde{\pi}_{ij}^{DB}$ merupakan dugaan proporsi pengangguran kecamatan ke- i desa ke- j hasil kalibrasi pembeda.

3. Hasil dan Pembahasan

3.1 Kajian Simulasi

Data yang dibangkitkan dalam simulasi pada penelitian ini terdiri dari dua skenario. Skenario 1 adalah populasi dengan gerombol terpisah secara sempurna. Skenario ini dilakukan untuk melihat kemampuan model pada kondisi populasi yang ideal, yaitu karakteristik antar gerombol mampu dibedakan dengan baik. Skenario 2 merupakan populasi dengan gerombol tidak terpisah secara sempurna. Skenario ini dilakukan untuk melihat kemampuan model pada kondisi populasi yang tidak ideal, yaitu saat terjadi penggerombolan yang beririsan.

Data populasi dibedakan menjadi dua, yaitu data contoh dan data nircontoh. Data contoh digunakan untuk membangun model prediksi, model dasar (Model-0) dan model yang dikembangkan (Model-1 dan Model-2). Paramater yang diamati adalah nilai tengah dari seluruh area, baik itu area contoh (i) maupun area nircontoh (i^*). Prediksi menggunakan Model-0 atau model dasar untuk area nircontoh mengabaikan pengaruh area. Model ini menghasilkan prediksi yang bersifat umum. Pengembangan model pertama (Model-1), yaitu pengembangan Model-0 dengan menambahkan nilai tengah dari pendugaan pengaruh acak area pada masing-masing gerombol pada model prediksi untuk area yang tidak terdapat contoh. Model-2 adalah pengembangan Model-1 dengan menambahkan pengaruh tetap gerombol ke- k . Model ini juga menambahkan nilai tengah dari pendugaan pengaruh acak area pada area nircontoh.

Relative Bias (RB)

Relative Bias (RB) adalah salah satu nilai yang digunakan untuk melakukan evaluasi kebaikan model. Nilai tengah dari *relative bias* prediksi proporsi pengangguran untuk Skenario 1 dan 2 pada ketiga model tersaji pada Tabel 3.

Nilai tengah dari *relative bias* Model-2 lebih kecil dibandingkan kedua model lain untuk kedua skenario. Skenario 1 memiliki nilai RB yang lebih kecil daripada Skenario 2 untuk ketiga model. Artinya penambahan pengaruh tetap gerombol pada area contoh dan penambahan nilai tengah dari pendugaan pengaruh acak area pada area nircontoh mampu meningkatkan prediksi. Hal ini berlaku pada kondisi ideal dengan penggerombolan terpisah sempurna.

Tabel 3: Nilai tengah dari *Relative Bias* (RB) (%).

Skena	Model-0	Model-1	Model-2
1	58.4914	40.7588	18.4581
2	340.5075	627.5818	552.808

Relative Root Mean Squares Error (RRMSE)

Indikator lain yang digunakan untuk mengevaluasi kebaikan model selain *relative bias* (RB) adalah *relative root mean squares error* (RRMSE). Model dengan RRMSE paling kecil dapat disimpulkan sebagai model terbaik. Nilai tengah dari RRMSE untuk ketiga model pada kedua kondisi skenario dapat dilihat pada Tabel 4.

Tabel 4: Nilai tengah dari RRMSE (%).

Skena	Model-0	Model-1	Model-2
1	103.5418	81.9220	59.0016
2	404.5762	449.8271	392.9396

Nilai tengah dari RRMSE Model-2 lebih kecil dibandingkan kedua model lain untuk kedua skenario. Sementara untuk ketiga model, Skenario 1 memiliki nilai RB yang lebih kecil daripada Skenario 2. Hal ini menunjukkan bahwa secara umum penambahan informasi gerombol mampu meningkatkan prediksi parameter area contoh dan nircontoh pada kondisi penggerombolan yang terpisah secara sempurna.

Hasil dari RRMSE ini serupa dengan hasil pada RB. Simulasi yang dilakukan mengungkapkan bahwa model yang dikembangkan secara umum memiliki akurasi prediksi yang lebih baik dibandingkan dengan model linier campuran terampat (MLCT) baku untuk semua skenario.

3.2 Kajian Aplikasi

Deskripsi Data

Sulawesi Tenggara merupakan salah satu provinsi di Sulawesi. Secara astronomis terletak di bagian selatan Garis Khatulistiwa, memanjang dari utara ke selatan di

antara 02°45'-06°15' Lintang Selatan dan membentang dari barat ke timur di antara 120°45'-124°45' Bujur Timur. Provinsi ini memiliki luas 38067.7 km². Sebelah utara berbatasan dengan Provinsi Sulawesi Selatan dan Provinsi Sulawesi Tengah, bagian selatannya berbatasan dengan Provinsi NTT di Laut Flores. Sebelah barat berbatasan dengan Provinsi Sulawesi Selatan di Laut Bone. Sebelah timur dengan Provinsi Maluku di Laut Banda. Pada tahun 2016 Provinsi Sulawesi Tenggara memiliki 15 kabupaten, 2 kota, 222 kecamatan, 1914 desa, dan 752 kelurahan.

Jumlah contoh terpilih Sakernas Agustus 2016 di Sulawesi Tenggara sebanyak 100 Blok Sensus (BS). Jumlah tersebut merupakan 2 persen dari total BS contoh secara nasional (5000 BS). Setiap Blok Sensus terpilih pada Sakernas 2016 akan dipilih 10 rumah tangga sebagai unit survei, sehingga jumlah keseluruhan rumah tangga terpilih adalah 1000 rumah tangga. BS terpilih terbanyak berada di Kota Kendari, yaitu sebanyak 10 BS dan paling sedikit di Kabupaten Konawe Kepulauan (3 BS). Dari 2272 jumlah desa/kelurahan di Provinsi Sulawesi Tenggara, terpilih sebanyak 98 desa/kelurahan yang menjadi contoh dan 2174 desa/kelurahan tidak menjadi contoh.

Pada tahun 2016 Provinsi Sulawesi Tenggara memiliki 15 kabupaten, 2 kota, 222 kecamatan, 1 914 desa, dan 752 kelurahan. Berdasarkan publikasi data ketenagakerjaan BPS hasil Sakernas Agustus 2016 Tingkat Partisipasi Angkatan Kerja (TPAK) Provinsi Sulawesi Tenggara tahun 2016 sebesar 73.47. Hal ini menunjukkan bahwa dari total penduduk usia kerja (15 tahun ke atas), sekitar 73.47 persen termasuk dalam angkatan kerja. Angkatan kerja merupakan penduduk usia 15 tahun dan lebih yang bekerja, mempunyai pekerjaan namun sementara tidak bekerja, dan pengangguran. Penduduk yang termasuk bukan angkatan kerja adalah penduduk usia kerja (15 tahun atau lebih) yang masih sekolah, mengurus rumah tangga, atau melaksanakan kegiatan lainnya selain kegiatan pribadi ([BPS], 2017).

TPAK Provinsi Sulawesi Tenggara pada Agustus tahun 2016 meningkat sebesar 5.11 poin dari tahun sebelumnya. TPAK Agustus 2015 yang berada di angka 68.36 persen menjadi 73.47 persen pada tahun 2016. Peningkatan TPAK ini merupakan indikasi meningkatnya kecenderungan penduduk ekonomi aktif untuk mencari atau melakukan kegiatan ekonomi.

Dari tahun 2013 hingga tahun 2017 terlihat bahwa TPT di kabupaten/kota di Provinsi Sulawesi Tenggara mengalami fluktuasi. Setiap tahunnya BPS mengeluarkan publikasi mengenai TPT sampai dengan tingkat kabupaten/kota. Namun, pada tahun 2016 publikasi tersebut tidak dapat dikeluarkan. Hal ini dikarenakan penambahan contoh pada Sakernas Agustus tahun 2016 tidak dilakukan. Jumlah contoh Sakernas Agustus 2016 sama dengan jumlah contoh Sakernas Februari 2016. Oleh karena itu, pada tahun 2016 publikasi TPT yang dikeluarkan oleh BPS hanya sampai tingkat provinsi.

Analisis Gerombol

Pemodelan area kecil yang dilakukan mencoba memasukkan informasi gerombol ke dalam model. Kecamatan-kecamatan yang ada di Provinsi Sulawesi Tenggara terlebih dahulu digerombolkan.. Peubah-peubah yang digunakan dalam penggerombolan terdiri dari Persentase penduduk laki-laki (X1), Persentase keluarga pertanian (X2), Rasio sarana pendidikan per 100 penduduk (X3), Proporsi rumah tangga pengguna

listrik PLN (X4), Proporsi surat keterangan miskin yang dikeluarkan pemerintah desa (X5), Rasio industri mikro dan kecil per 100 penduduk (X6), Rasio pasar (bangunan permanen dan semi permanen) per jumlah penduduk (X7), Rasio kedai makanan per jumlah penduduk (X8), Rasio toko per jumlah penduduk (X9), Rasio hotel/penginapan per jumlah penduduk (X10).

Analisis gerombol yang dilakukan pada penelitian ini adalah dengan menggunakan metode *Ward*. Ukuran kemiripan dan ketidakmiripan yang digunakan untuk penggerombolan adalah jarak *Euclid*. Penentuan jumlah gerombol yang akan digunakan ditentukan dari hasil evaluasi pembentukan gerombol. Hasil evaluasi menunjukkan bahwa dengan peubah-peubah PODES 2014, metode gerombol yang cocok adalah hierarki dengan jumlah gerombol sebanyak tiga. Selain hasil evaluasi juga digunakan dendrogram sebagai bahan pertimbangan pembentukan gerombol. Kecamatan di Provinsi Sulawesi Tenggara pada penelitian ini terbagi menjadi tiga gerombol yang tersaji pada Tabel 5.

Tabel 5: Penggerombolan kecamatan di Provinsi Sulawesi Tenggara.

Gerombol	Kecamatan contoh	Kecamatan nircontoh	Banyaknya kecamatan
1	48	113	161
2	19	15	34
3	2	12	14

Pemodelan Pendugaan Area Kecil

Model area kecil dibangun berdasarkan regresi linier campuran terampat. Peubah yang menjadi perhatian (Y) adalah proporsi pengangguran di Provinsi Sulawesi Tenggara. Peubah penyerta (X) yang digunakan sebanyak dua peubah, yaitu: (1) proporsi rumah tangga pengguna listrik PLN dan (2) proporsi surat keterangan miskin yang dikeluarkan pemerintah desa. Model yang digunakan adalah model level unit dengan desa/kelurahan sebagai level unit dan kecamatan sebagai I evel area. Pemodelan area kecil dilakukan dengan menggunakan model linier campuran terampat baku (Model-0) dan dua pengembangan model (Model-1 dan Model-2).

Tabel 6: Nilai tengah dugaan proporsi pengangguran dan RMSE hasil dugaan proporsi pengangguran.

Nilai Tengah	Model-0	Model-1	Model-2
Dugaan	0.026400	0.025200	0.024200
RMSE	0.000535	0.000573	0.000475

Dugaan proporsi pengangguran kecamatan memiliki nilai tengah yang relatif sama untuk ketiga model (Tabel 6). Namun, Model-2 memiliki nilai tengah dugaan proporsi pengangguran yang lebih rendah daripada kedua model lain. Nilai tengah RMSE menunjukkan bahwa Model-2 memiliki nilai tengah yang lebih kecil dibandingkan dengan kedua model lain. Hal ini sejalan dengan hasil yang diperoleh pada kajian

simulasi. Dengan kata lain memperlihatkan bahwa Model-2 sudah cukup baik untuk menduga indikator ketenagakerjaan, dalam hal ini proporsi pengangguran, di Provinsi Sulawesi Tenggara.

Pendugaan Proporsi Pengangguran

Pendugaan proporsi pengangguran tingkat desa menggunakan Model-2 dengan menambahkan informasi gerombol pada efek tetap dan juga efek acak. Model-2 ini digunakan karena dianggap terbaik daripada dua model yang lain, dilihat dari RMSE yang dihasilkan. Setelah mendapatkan hasil dugaan proporsi pengangguran untuk desa contoh dan nircontoh, kemudian dilakukan agregasi pada tingkat kecamatan, sesuai tujuan yang ingin dicapai. Selanjutnya, dugaan proporsi yang telah dihasilkan juga dilakukan agregasi untuk tingkat yang lebih tinggi, yaitu provinsi, untuk melihat kelayakan hasil dugaan yang telah diperoleh sebelumnya.

Agregasi dilakukan dengan menggunakan rataan tertimbang dan kemudian dilakukan kalibrasi. Kalibrasi yang dilakukan menggunakan teknik perbedaan (*difference*). Kalibrasi dengan menggunakan teknik ini menghasilkan dugaan proporsi pengangguran tingkat provinsi yang sama dengan yang dikeluarkan oleh BPS untuk Provinsi Sulawesi Tenggara pada tahun 2016 (0.0272). Selang kepercayaan yang terbentuk untuk agregasi dugaan proporsi pengangguran berkisar antara -0.00085 sampai dengan 0.04145. Terlihat bahwa nilai dugaan terletak di dalam selang kepercayaan yang terbentuk. Dengan demikian, hasil pendugaan area kecil dengan kalibrasi layak digunakan untuk menduga proporsi pengangguran tingkat kecamatan di Provinsi Sulawesi Tenggara.

Hasil dugaan proporsi pengangguran tingkat kecamatan yang telah dilakukan kalibrasi berkisar antara 0.0079 sampai dengan 0.0898.

4. Simpulan

Penggunaan area kecil dengan menambahkan informasi gerombol memberikan pengaruh terhadap akurasi dalam pendugaan proporsi pengangguran. Hasil perbandingan antara MLCT baku dan model yang dikembangkan memperlihatkan bahwa Model-2 merupakan model yang lebih baik untuk menduga proporsi pengangguran. Model-2 merupakan pengembangan MLCT baku dengan menambahkan pengaruh tetap gerombol ke- k . Model ini juga menambahkan nilai tengah dari pendugaan pengaruh acak area pada area nircontoh. Pendugaan dengan menggunakan Model-2 menghasilkan dugaan proporsi pengangguran kalibrasi di Provinsi Sulawesi Tenggara tahun 2016 sebesar 2.72 persen. Angka tersebut sama dengan hasil yang dikeluarkan oleh BPS (2.72 persen). Dari hasil RMSE terlihat bahwa Model-2 memiliki nilai yang paling kecil dibandingkan dua model lain. Hasil ini sejalan dengan kajian simulasi dan memperlihatkan bahwa dugaan proporsi pengangguran dengan menggunakan informasi gerombol (Model-2) layak digunakan.

Daftar Pustaka

- Anisa, R., Kurnia, A., & Indahwati. (2014). Cluster information of non-sampled area in small area estimation. *IOSR J Math*, 10(1): 15–19.
- [BPS] Badan Pusat Statistik. (2016). *Pedoman Pencacahan Sakernas Semesteran 2016*. Jakarta (ID): BPS.
- [BPS] Badan Pusat Statistik Provinsi Sulawesi Tenggara. (2017). *Keadaan Ketenagakerjaan Provinsi Sulawesi Tenggara 2016*. Kendari (ID): BPS Provinsi Sulawesi Tenggara.
- Ghaniswati, R. (2019). *Model pendugaan area kecil untuk proporsi penduduk miskin level desa (studi kasus proporsi penduduk miskin di Kabupaten Lembata) [tesis]*. Bogor (ID): IPB University.
- Kurnia, A. (2009). *Prediksi terbaik empirik untuk model transformasi logaritma di dalam pendugaan area kecil dengan penerapan pada data susenas [disertasi]*. Bogor (ID): IPB University.
- Rao, J., & Molina, I. (2015). *Small area estimation*. New York (US): John Wiley and Sons.
- Sadik, K. (2009). *Metode prediksi tak-bias linear terbaik dan bayes berhierarki untuk pendugaan area kecil berdasarkan model state space [disertasi]*. Bogor (ID): IPB University.
- Sundara, V. Y., Sadik, K., & Kurnia, A. (2017). Cluster information of non-sampled area in small area estimation of poverty indicators using Empirical Bayes. *AIP Conference Proceedings*, 1827(1), 020026. AIP Publishing LLC.
- Yudistira, Y., Kurnia, A., & Soleh, A. M. (2018). Binomial regression in small area estimation method for estimate proportion of cultural indicator. *Indonesian Journal of Statistics and Its Applications*, 2(2): 56–62.