

PREDIKSI TERBAIK EMPIRIK
UNTUK MODEL TRANSFORMASI LOGARITMA
DI DALAM PENDUGAAN AREA KECIL
DENGAN PENERAPAN PADA DATA SUSENAS

ANANG KURNIA



SEKOLAH PASCASARJANA
INSTITUT PERTANIAN BOGOR
2009

PERNYATAAAN MENGENAI DISERTASI DAN SUMBER INFORMASI

Dengan ini saya menyatakan bahwa disertasi dengan judul *Prediksi Terbaik Empirik untuk Model Transformasi Logaritma di dalam Pendugaan Area Kecil dengan Penerapan pada Data Susenas* adalah karya saya sendiri dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan atau tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir disertasi ini.

Bogor, November 2009

Anang Kurnia
G161050041

ABSTRACT

ANANG KURNIA. An Empirical Best Prediction Method for Logarithmic Transformation Model in Small Area Estimation with Particular Application to Susenas Data. Supervised by KHAIRIL ANWAR NOTODIPUTRO, ASEP SAEFUDDIN, and I WAYAN MANGKU.

Currently statistician has given attention seriously to small area statistics. Fay and Herriot (1979) are the first researchers who developed small area estimation (SAE) which is based on linear mixed model. The model, known as Fay-Herriot model, has become a reference in the developing of further research on SAE.

The dissertation focuses on developing robust SAE method to overcome large ratio variation between small areas and total variation. The developed method is also robust with respect to misspecification model, in a sense that the model fails to capture the correct trend. There are some reasons why these topics are selected as a focus of this dissertation. Firstly, based on some preliminary studies in an application of SAE methods using BPS' data, some improvements on precision of estimation has been shown but it was still unsatisfactory. Secondly, the fact that the pattern of social and economic data is difficult to fulfill linear assumption which is required in standard SAE method. The transformation on response variable, independent variable or both can be used to linearized the relationship.

The lognormal empirical best prediction (EBP) model which are suggested in this research give better results than standard SAE models in terms of the smallest relative root mean square error (RRMSE). However, the relative bias is still large enough (approximately 8 %). On the other hand, the evaluation model using mean square prediction error (MSPE) is underestimated which was indicated by ignoring estimation effect on error variance ($\hat{\sigma}_i^2$) and bias of lognormal EBP. In general, the suggested model improves the efficiency of estimation.

In this research, the efforts to reduce the bias on mean estimation of model lognormal EBP are still not explored yet. Further research on these will improve performance the suggested model, including the correction factor on estimation of non-sampled unit. Meanwhile, the MSPE could be decreased analytically by modeling the effect of variance estimation through Taylor expansion or resampling technique such as jackknife and bootstrap.

Keywords : small area estimation, logarithmic transformation, empirical best prediction

RINGKASAN

ANANG KURNIA. Prediksi Terbaik Empirik untuk Model Transformasi Logaritma di dalam Pendugaan Area Kecil dengan Penerapan pada Data Susenas. Dibimbing oleh KHAIRIL ANWAR NOTODIPUTRO, ASEP SAEFUDDIN, dan I WAYAN MANGKU.

Statistik area kecil (*small area statistics*) saat ini telah menjadi perhatian para statistisi dunia secara sangat serius. Fay dan Herriot (1979) merupakan peneliti pertama yang mengembangkan pendugaan area kecil (*small area estimation, SAE*) berdasarkan model linier campuran. Model yang dikembangkan kemudian menjadi rujukan dalam pengembangan penelitian SAE lebih lanjut.

Ruang lingkup penulisan disertasi ini tertuju untuk melakukan pengembangan metode SAE yang bersifat kekar terutama dalam mengatasi besarnya rasio keragaman antar area kecil terhadap keragaman total dan kemungkinan terjadinya ketidak cocokan model. Pemilihan fokus penelitian ini didasari oleh: (i) berdasarkan berbagai kajian awal, diindikasikan bahwa penerapan berbagai metode SAE dalam kasus data BPS walaupun memberikan perbaikan presisi hasil pendugaan namun masih belum memuaskan, dan (ii) kenyataan bahwa pola hubungan data sosial dan ekonomi sulit memenuhi hubungan linier dalam skala pengukuran asli, dan baru dipenuhi setelah dilakukan transformasi pada peubah respon, peubah bebas atau pada kedua-duanya.

Dalam pendugaan area kecil, khususnya pendugaan berbasis model, konsep model linier menjadi inti dari analisis yang dilakukan. Keterpenuhan asumsi linier tentu menjadi syarat penting dalam melakukan analisis lebih lanjut. Berdasarkan eksplorasi metode, model aditif dan nonparametrik mampu memberikan perbaikan dalam mengurangi pengaruh ketidaktepatan pemodelan linier yang digunakan. Namun demikian, dalam kasus data Susenas, walaupun keduanya memberikan perbaikan presisi pendugaan namun masih belum memuaskan. Salah satu kelemahan yang diperlihatkan dalam kasus ini adalah sempitnya selang pengukuran peubah penyerta.

Penggunaan EBLUP baku dalam konteks pendugaan pengeluaran per kapita atau data sosial secara umum menunjukkan hasil yang juga tidak memuaskan. Evaluasi sisaan dari model pendugaan pengeluaran per kapita memberikan petunjuk perlunya dilakukan upaya transformasi. Lebih lanjut, upaya perbaikan metode SAE berbasis model dengan terlebih dahulu melakukan transformasi logaritma terhadap data respon sebelum diterapkan teknik SAE menjadi fokus utama dalam disertasi ini. Akibat yang dihadapi dari melakukan transformasi ini adalah diperolehnya penduga parameter dari transformasi-balik yang bersifat berbias serta kesulitan memperoleh bentuk eksplisit dari kuadrat tengah

galatnya. Pembahasan dua hal tersebut disajikan dalam Bab V disertasi ini yang juga disertai dengan evaluasi sifat-sifat statistiknya melalui simulasi serta contoh aplikasi untuk data BPS (Susenas dan Podes) di Kabupaten dan Kota Bogor.

Model yang diajukan, lognormal EBP, memberikan hasil yang lebih baik seperti yang diperlihatkan oleh RRMSE yang paling kecil, walaupun bias relatifnya masih cukup besar. Sifat *underestimate* MSPE dari lognormal EBP diindikasikan setidaknya dipengaruhi oleh dua hal yaitu karena (1) mengabaikan pengaruh pendugaan σ_i^2 dan (2) pengaruh bias pendugaan. Namun demikian, secara umum model yang diajukan mampu meningkatkan efisiensi pendugaan. Kelebihan lain yang diperoleh dari model lognormal EBP adalah mampu mengurangi pengaruh langsung jika terjadi rasio keragaman antar area kecil terhadap keragaman total yang cukup besar.

Dalam penelitian ini belum dilakukan upaya-upaya untuk memperbaiki bias pada pendugaan nilai tengah dari model lognormal EBP maupun bias pada MSPE-nya. Kajian lebih lanjut untuk kedua hal tersebut akan meningkatkan performa dari model yang dikembangkan. Konsep dengan menambahkan faktor koreksi pada komponen dugaan nilai tengah anggota populasi yang bukan anggota contoh dapat dipertimbangkan untuk hal ini. Sedangkan untuk memperbaiki MSPE bisa dilakukan secara analitik dengan memformulasikan pengaruh pendugaan σ_i^2 melalui penguraian deret Taylor atau dengan teknik *resampling* seperti jackknife dan bootstrap.

Kata Kunci: pendugaan area kecil, transformasi logaritma, prediksi terbaik empirik

@ Hak Cipta Milik Institut Pertanian Bogor (IPB), Tahun 2009
Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

PREDIKSI TERBAIK EMPIRIK
UNTUK MODEL TRANSFORMASI LOGARITMA
DI DALAM PENDUGAAN AREA KECIL
DENGAN PENERAPAN PADA DATA SUSENAS

ANANG KURNIA

Disertasi
sebagai salah satu syarat untuk memperoleh gelar Doktor
pada Program Studi Statistika

SEKOLAH PASCASARJANA
INSTITUT PERTANIAN BOGOR
2009

Penguji Luar Komisi pada Ujian Tertutup, 3 November 2009 :

1. Prof. Dr. Ir. A. A. Mattjik, MSc (Dosen Dept. Statistika IPB)
2. Dr. Ir. Putu Purnaba, DEA (Dosen Dept. Matematika IPB)

Penguji Luar Komisi pada Ujian Terbuka, 24 November 2009 :

1. Prof. Dr. Ir. Aunuddin, MSc (Dosen Dept. Statistika IPB)
2. Dr. Slamet Sutomo (Deputi Bidang Neraca dan Analisis Statistik - BPS)

Judul Disertasi : Prediksi Terbaik Empirik untuk Model
Transformasi Logaritma di dalam Pendugaan
Area Kecil dengan Penerapan pada Data Susenas
Nama Mahasiswa : Anang Kurnia
N R P : G161050041
Program Studi : Statistika

Disetujui:
Komisi Pembimbing

Prof. Dr. Ir. Khairil Anwar Notodiputro, MS
Ketua

Dr. Ir. Asep Saefuddin, MSc
Anggota

Dr. Ir. I Wayan Mangku, MSc
Anggota

Diketahui:

Ketua Program Studi Statistika

Dekan Sekolah Pascasarjana

Dr. Ir. Aji Hamim Wigena, MSc

Prof. Dr. Ir. Khairil Anwar Notodiputro, MS

Tanggal Ujian Terbuka : 24 November 2009

Tanggal Lulus :

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala karunia serta rahmat-Nya sehingga disertasi ini berhasil diselesaikan dengan baik. Tema utama penulisan disertasi ini adalah pengembangan model prediksi terbaik empirik untuk model transformasi (logaritma) pada pendugaan area kecil. Penelitian ini merupakan bagian dari payung penelitian Hibah Penelitian Tim Pascasarjana IPB tahun 2006-2007 kerjasama antara Departemen Statistika FMIPA IPB dengan Badan Pusat Statistik.

Disertasi ini merupakan rangkuman pekerjaan penelitian yang dilakukan penulis dalam bidang pendugaan area kecil. Beberapa bagian dari disertasi ini juga sudah dipublikasikan baik dalam seminar nasional maupun internasional serta jurnal ilmiah. Oleh karena itu, penulisan disertasi ini secara umum dibagi dalam tiga bagian besar yang membahas perkembangan pendugaan area kecil di dunia, beberapa contoh penerapan dan permasalahan yang dihadapi dalam kasus data BPS (Susenas dan Podes), serta pengembangan model transformasi logaritma untuk mengatasi pola hubungan peubah penyerta yang diharapkan mampu mereduksi pengaruh ketidaktepatan pemodelan serta rasio keragaman.

Penulis menyadari bahwa karya ilmiah ini dapat diselesaikan hanya atas pertolongan Allah SWT serta bantuan dan jasa baik dari berbagai pihak. Oleh karena itu, selain rasa syukur dan pujian pada Allah SWT, pada kesempatan ini penulis juga mengucapkan rasa terima kasih dengan penuh rasa hormat atas arahan, dukungan, motivasi serta bimbingannya kepada Prof. Dr. Ir. Khairil Anwar Notodiputro, MS selaku ketua komisi pembimbing, Dr. Ir. Asep Saefuddin, MSc dan Dr. Ir. I Wayan Mangku, MSc selaku anggota komisi pembimbing.

Penghargaan khusus penulis sampaikan kepada Prof. Dr. Ray Chambers dan Prof. Dr. David Steel atas segala masukan dan diskusinya, juga seluruh teman sejawat di Departemen Statistika IPB serta rekan-rekan di PS Statistika Sekolah Pascasarjana IPB. Ungkapan terima kasih yang tulus juga penulis sampaikan kepada kedua orangtua, bapak dan ibu mertua serta kakak dan adik-adik semua. Kepada istri tercinta Dian Handayani dan anak tersayang M. Irfan Hanifiandi Kurnia, penulis sampaikan terima kasih atas motivasi, dukungan, doa, kesabaran serta kasih sayangnya selama ini.

Semoga Allah SWT senantiasa memberikan kebaikan untuk kita semua.

Bogor, November 2009

Anang Kurnia

RIWAYAT HIDUP

Penulis dilahirkan di Cineam (Tasikmalaya), Jawa Barat pada tanggal 24 Agustus 1973 dari pasangan Bapak Hadian Supratman (alm) dan Ibu Masnah Suryati. Penulis merupakan anak kedua dari dua bersaudara. Pendidikan Sarjana ditempuh di Jurusan Statistika, FMIPA IPB pada tahun 1992 – 1996 dengan pembimbing skripsi Prof. Dr. Andi Hakim Nasoetion (alm) dan Dr. Ir. Aji Hamim Wigena, MSc. Pada tahun 2000, penulis memperoleh gelar Magister Sains dari Program Studi Statistika Sekolah Pascasarjana IPB dengan pembimbing thesis Dr. Ir. Asep Saefuddin, MSc dan Dr. Ir. Bambang Juanda, MS. Sejak tahun 2005 penulis menempuh Program Doktor pada Program Studi Statistika Sekolah Pascasarjana IPB.

Dalam rangka penelitian disertasi, penulis berkesempatan untuk mengikuti program sandwich di Centre for Statistical and Survey Methodology, University of Wollongong, Australia pada bulan Oktober 2008 sampai dengan Januari 2009 dengan pembimbing Prof. Dr. Ray Chambers dan Prof. Dr. David Steel.

Selama mengikuti pendidikan Program Doktor, penulis telah menghasilkan beberapa karya ilmiah yang telah dipublikasikan dalam seminar nasional maupun internasional serta jurnal ilmiah. Karya-karya ilmiah tersebut merupakan bagian dari Program Doktor penulis. Karya-karya ilmiah tersebut adalah :

1. Kurnia A, Notodiputro KA, Saefuddin A dan Mangku IW. 2009. Penduga Terbaik Empirik untuk Model Transformasi Logaritma pada Pendugaan Area Kecil. Prosiding pada Seminar Nasional Statistika IX, 7 Nopember 2009, ITS – Surabaya.
2. Notodiputro KA dan Kurnia A. 2009. Small Area Statistic in Indonesia. Proceeding at the International Conference on Small Area Estimation 2009, 29 June – 1 July 2009, Elche – Spain.
3. Kurnia A, Notodiputro KA dan Chambers R. 2009. An Empirical Study of EBLUP Log-Transformation for Small Areas. Proceeding at

the International Conference on Small Area Estimation 2009, 29 June – 1 July 2009, Elche – Spain.

4. Kurnia A, Notodiputro KA, Saefuddin A dan Mangku IW. 2009. Indonesia Case Small Area Estimation. Proceeding at The First International Seminar on Science and Technology, 24-25 January 2009. Yogyakarta – Indonesia.
5. Kurnia A dan Notodiputro KA. 2008. Generalized Additive Mixed Models for Small Area Estimation. Mathematics Journal Universiti Teknologi Malaysia, 385-391.
6. Notodiputro KA, Kurnia A dan Sadik K. 2008. Statistical Models for Small Area Estimation. Proceeding at The 3rd International Conference on Mathematics and Statistics, 5 – 6 August 2008. MSMSSEA & IPB – Indonesia.
7. Kurnia A, Notodiputro KA, Saefuddin A dan Mangku IW. 2008. Generalized Regression dalam Pendugaan Area Kecil. Prosiding pada Seminar Nasional Mahasiswa S3 Matematika Indonesia, 31 Mei 2008, UGM - Yogyakarta.
8. Kurnia A dan Notodiputro KA. 2007. Pendekatan Generalized Additive Mixed Models dalam Pendugaan Parameter pada Kasus Small Area Estimation. Jurnal Sains MIPA (ISSN 1978-1873) 13(3), 145 - 151.
9. Kurnia A, Notodiputro KA, dan Ibrahim NA. 2007. A Nonparametric Approach in Small Area Estimation. Proceeding at The 9th Islamic Countries Conference on Statistical Sciences, ISOSS - University of Malaya, 12-14 December 2007. Shah Alam, Malaysia.
10. Kurnia A, Sartono B dan Wulandari R. 2007. Pengaruh Misspesifikasi Desain Survey pada Pendugaan Area Kecil dengan Pendekatan Generalized Regression. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika, 24 November 2007. Universitas Negeri Yogyakarta.
11. Notodiputro KA dan Kurnia A. 2007. Development of Small Area Estimation Research in Indonesia. Proceeding at the 2nd International Conference on Mathematical Sciences 2007 (ICoMS-2007), 28 - 29 May 2007. MSMSSEA - Universiti Teknologi Malaysia, Johor Bahru – Malaysia.
12. Notodiputro KA dan Kurnia A. 2007. Small Area Estimation Research Activity in Bogor Agricultural University. Paper has presented at The Seminar on Use, Analysis and Application of Small-area Statistics, Cooperation between BPS and JICA, March 7, 2007. BPS – Jakarta.
13. Kurnia A dan Notodiputro KA. 2006. Pengaruh Pendugaan Ragam Penarikan Contoh pada Small Area Estimation. JMAP Edisi Khusus, Universitas Negeri Jakarta.

14. Handayani D dan Kurnia A. 2006. Pendekatan Empirical Bayes pada Pendugaan Nilai Tengah Populasi Terhingga dalam Kasus Small Area Estimation. JMAP Edisi Khusus, Universitas Negeri Jakarta.
15. Kurnia A dan Notodiputro KA. 2006. EB-EBLUP MSE Estimator on Small Area Estimation with Application to BPS'Data. *Proceeding at The First International Conference on Mathematics and Statistics (ICoMS-1)*, June 19-21, 2006. Bandung.
16. Kismiantini, Kurnia A dan Notodiputro KA. 2006. Risk of Dengue Haemorrhagic Fever in Bekasi Municipality with Small Area Approach. *Proceeding at The First International Conference on Mathematics and Statistics (ICoMS-1)*, June 19-21, 2006. Bandung.
17. Kurnia A dan Notodiputro KA. 2006. Penerapan Metode Jackknife dalam Pendugaan Area Kecil. *Forum Statistika dan Komputasi*, ISSN 0853-8115 Vol. 11 No.1.
18. Kurnia A dan Notodiputro KA. 2005. Aplikasi Metode Bayes pada Small Area Estimation. *Prosiding Seminar Nasional Statistika VII*, 26 November 2005. ITS Surabaya.
19. Kurnia A dan Notodiputro KA. 2005. Pendekatan Generalized Linear Mixed Model pada Small Area Estimations. *Forum Statistika dan Komputasi*, ISSN 0853-8115 Vol. 10 No.2.

DAFTAR ISTILAH

ACR	: <i>Average of Coverage Rates</i> (Rataan Tingkat Ketercakupan)
ARB	: <i>Average of Relative Bias</i> (Rataan Bias Relatif)
ARBMSE	: <i>Average of Relative Bias Mean Square Error</i> (Rataan Bias Kuadrat Tengah Galat Relatif)
ARRMSE	: <i>Average of Relative Root Mean Square Error</i> (Rataan Akar Kuadrat Tengah Galat Relatif)
BLUP	: <i>Best Linear Unbiased Prediction</i> (Prediksi Takbias Terbaik Linear)
BP	: <i>Best Prediction</i> (Prediksi Terbaik)
EB	: <i>Empirical Bayes</i>
EBLUP	: <i>Empirical Best Linear Unbiased Prediction</i> (Prediksi Takbias Terbaik Linear Empirik)
EBP	: <i>Empirical Best Prediction</i> (Prediksi Terbaik Empirik)
GAM	: Generalized Additive Model (Model Aditif Terampat)
GAMM	: Generalized Additive Mixed Model (Model Aditif Terampat Campuran)
GREG	: <i>General Regression</i>
GREG-linier	: Penduga GREG yang pendugaan parameter regresinya berdasarkan model regresi linier
GREG-logistik	: Penduga GREG yang pendugaan parameter regresinya berdasarkan model regresi logistik
GWLS	: <i>Generalized Weighted Least Square</i> (Metode Kuadrat Terkecil Terboboti)
HB	: <i>Hierarchical Bayes</i>
IRLS	: <i>Iterative Reweighted Least Squares</i>
M-GREG	: GREG yang pendugaan koefisien regresinya berdasarkan metode M-estimator
MKT-GREG	: GREG yang pendugaan koefisien regresinya berdasarkan metode kuadrat terkecil
MBDE	: <i>Model Based Design Estimator</i> (Penduga Langsung Berbasis Model)
MLE	: <i>Maximum Likelihood Estimator</i> (Penduga Kemungkinan Maksimum)
MSE	: <i>Mean Square Error</i> (Kuadrat Tengah Galat)
MSPE	: <i>Mean Square Prediction Error</i> (Kuadrat Tengah Galat Prediksi)
REML	: <i>Restricted/residual Maximum Likelihood</i> (Penduga Kemungkinan Maksimum Terkendala)

RMSE	: <i>Root Mean Square Error</i> (Akar Kuadrat Tengah Galat)
RRMSE	: <i>Relative Root Mean Square Error</i> (Akar Kuadrat Tengah Galat Relatif)
SAE	: <i>Small Area Estimation</i> (Pendugaan Area Kecil)
SRS	: <i>Simple Random Sampling</i> (Penarikan Contoh Acak Sederhana)
TSCS	: <i>Two Stage Cluster Sampling</i> (Penarikan Contoh Acak Gerombol Dua Tahap)
<i>Direct Estimation</i>	: Pendugaan Langsung (Pendugaan parameter yang dilakukan hanya berdasarkan data survei pada masing-masing area)
<i>Indirect Estimation</i>	: Pendugaan Tidak Langsung (Pendugaan parameter yang dilakukan dengan melibatkan informasi tambahan baik dari dalam area yang menjadi perhatian, area lain maupun survei lain)

DAFTAR ISI

	halaman
DAFTAR TABEL	xvi
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xviii
 I. PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Tujuan Penelitian	6
1.3. Ruang Lingkup Penelitian	7
1.4. Kebaruan Penelitian	8
1.5. Sistematika Penulisan Disertasi	8
 II. METODE PENDUGAAN AREA KECIL	
2.1. Model Dasar Pendugaan Area Kecil	10
2.2. Konsep Pendugaan Berbasis Disain	15
2.3. Konsep dan Sifat Statistik Penduga EBLUP	21
2.4. Konsep dan Sifat Statistik Penduga Bayes	26
 III. METODE PENDUGAAN LANGSUNG	
3.1. Latar Belakang	35
3.2. GREG pada Penarikan Contoh Acak Sederhana dan Gerombol Dua Tahap	36
3.3. Penerapan pada Data Susenas	39
3.4. Kajian Simulasi Regresi Kekar pada GREG	45
3.5. Kesimpulan	46
 IV. METODE PENDUGAAN TIDAK LANGSUNG	
4.1. Latar Belakang	48
4.2. Pendekatan Generalized Aditif Mixed Model	48

4.3. Pendekatan Nonparametrik P-Spline	50
4.4. Penerapan Model Aditif dan Nonparametrik pada Data Susenas	53
4.5. Pendekatan EBLUP Baku untuk Pendugaan Pengeluaran per Kapita di Kabupaten dan Kota Bogor	56
4.6. Kesimpulan	60
V. PREDIKSI TERBAIK EMPIRIK UNTUK MODEL TRANSFORMASI LOGARITMA	
5.1. Latar Belakang	61
5.2. Model Transformasi Logaritma	62
5.3. Evaluasi Sifat Statistik Model EBP Berdasarkan Simulasi ...	67
5.4. Penerapan pada Data Susenas	69
5.5. Kesimpulan	71
VI. PEMBAHASAN	72
VII. KESIMPULAN DAN SARAN	
7.1. Kesimpulan	76
7.2. Saran	76
DAFTAR PUSTAKA	77
LAMPIRAN	82

DAFTAR TABEL

halaman

3.1. Formula penduga langsung berdasarkan metode SRS dan TSCS	38
3.2. Formula penduga GREG berdasarkan metode SRS	38
3.3. Formula penduga GREG berdasarkan metode TSCS	38
3.4. RRMSE penduga langsung dan GREG untuk pendugaan pengeluaran per kapita berdasarkan metode SRS dan TSCS di Kota Bogor berdasarkan data Susenas tahun 2005	42
4.1. Pendugaan tingkat pengangguran dalam persen di Kota Bogor berdasarkan data Susenas tahun 2005	54
4.2. Sebaran contoh desa/kelurahan pada Susenas 2005 di Kabupaten dan Kota Bogor	57
5.1. Ringkasan hasil kajian simulasi	68

DAFTAR GAMBAR

halaman

2.1. Plot kurva linier dan logistik	19
3.1. Boxplot nilai RRMSE berdasarkan metode SRS	44
3.2. Boxplot nilai RRMSE berdasarkan metode TSCS	44
3.3. RRMSE MKT-GREG vs M-GREG berdasarkan proporsi pencilan	46
4.1. Diagram pencar peubah penyerta	55
4.2. Boxplot RRMSE penduga langsung dan EBLUP	57
4.3. Plot sisaan baku vs prediksi pada model dasar	58
4.4. Boxplot sisaan baku pada model dasar	58
4.5. Histogram sisaan baku pada model dasar	59
4.6. Plot QQ – sisaan baku pada model dasar	59
4.7. Histogram data asal (peubah respon)	59
5.1. Boxplot penduga pengeluaran per kapita desa/kelurahan di Kabupaten dan Kota Bogor berdasarkan empat metode pendekatan	69
5.2. Boxplot penduga MSE relatif (%) pada pendugaan pengeluaran per kapita desa/kelurahan di Kabupaten dan Kota Bogor berdasarkan empat metode pendekatan	69
5.3. Histogram sisaan baku pada model transformasi	70
5.4. Plot sisaan baku vs prediksi pada model transformasi	70

DAFTAR LAMPIRAN

	halaman
1. Sifat statistik transformasi-balik logaritma dari respon yang menyebar lognormal	82
2. Pemrograman GAMM untuk SAE	83
3. Pemrograman P-Spline untuk SAE	86
4. Penduga pengeluaran per kapita dalam rupiah di Kabupaten dan Kota Bogor serta penduga RMSE relatifnya menggunakan metode pendugaan langsung dan EBLUP berdasarkan data Susenas tahun 2005	88
5. Pemrograman SAS untuk simulasi EBP	91
6. Ringkasan kajian simulasi untuk model EBP (bias relatif dan RRMSE)	95
7. Ringkasan kajian simulasi untuk model EBP (bias relatif MSE Dan tingkat ketercakupan)	98
8. Pemrograman SAS pendugaan pendapatan per kapita di Kabupaten dan Kota Bogor berdasarkan data Susenas 2005 dan Podes 2005	101
9. Penduga pengeluaran per kapita dalam rupiah di Kabupaten dan Kota Bogor dan RMSE relatifnya berdasarkan data Susenas tahun 2005	104

BAB I

PENDAHULUAN

1.1. Latar Belakang

Statistik area kecil (*small area statistics*) saat ini telah menjadi perhatian para statistisi dunia secara sangat serius. Telah banyak penelitian yang dikembangkan baik untuk perbaikan teknik dan pengembangan metode maupun aplikasi dalam berbagai kasus dan persoalan nyata yang dihadapi. Fay dan Herriot (1979) merupakan peneliti pertama yang mengembangkan pendugaan area kecil (*small area estimation, SAE*) berbasis model. Model yang dikembangkannya kemudian menjadi rujukan dalam pengembangan penelitian pendugaan area kecil lebih lanjut sampai dengan saat ini.

Perhatian yang besar ini terjadi seiring dengan meningkatnya kebutuhan pemerintah dan para pengguna statistik (termasuk dunia bisnis) terhadap informasi yang lebih rinci, cepat, dan handal, tidak saja untuk lingkup superpopulasi seperti negara tetapi pada lingkup yang lebih kecil (sub-populasi) seperti kabupaten, kecamatan dan desa/kelurahan atau sub-populasi lain yang dibangun oleh karakteristik jenis kelamin, status sosial ekonomi, pendidikan, ras dan yang lainnya. Bagi kita di Indonesia, pentingnya statistik area kecil semakin dirasakan seiring dengan era otonomi daerah dimana sistem ketatanegaraan bergeser dari sistem sentralisasi ke sistem desentralisasi. Pada sistem desentralisasi, pemerintah daerah memiliki kewenangan yang lebih besar untuk mengatur dirinya sendiri, khususnya pada level pemerintah kabupaten/kota. Dengan demikian kebutuhan statistik sampai pada level desa/kelurahan menjadi suatu kebutuhan dasar sebagai landasan bagi pemerintah daerah kabupaten/kota untuk menyusun sistem perencanaan, pemantauan dan penilaian pembangunan daerah atau kebijakan penting lainnya.

Pendugaan area kecil merupakan konsep terpenting dalam pendugaan parameter di suatu area yang relatif kecil dalam percontohan survei (*survey sampling*). Metode pendugaan area kecil digunakan untuk menduga karakteristik dari sub-populasi (domain yang lebih kecil). Pendugaan langsung (*direct estimation*) pada sub-populasi tidak memiliki presisi yang memadai karena kecilnya jumlah contoh yang digunakan untuk memperoleh dugaan tersebut. Alternatif metode lain adalah dengan cara menambahkan informasi pada area tersebut dari area lain melalui pembentukan model yang tepat. Pendugaan parameter area kecil dengan pendekatan seperti ini disebut pendugaan tidak langsung (*indirect estimation*) dalam arti bahwa dugaan tersebut mencakup data dari domain yang lain. Chand dan Alexander (1995) menyebutkan bahwa prosedur pendugaan area kecil pada dasarnya memanfaatkan kekuatan informasi area sekitarnya (*neighbouring areas*) dan sumber data di luar area yang statistiknya ingin diperoleh melalui pembentukan model yang tepat untuk meningkatkan efektifitas ukuran contoh. Secara umum pendugaan area kecil dapat dikatakan sebagai suatu metode untuk menduga parameter pada suatu area yang relatif kecil dalam percontohan survei dengan memanfaatkan informasi dari luar area, dari dalam area itu sendiri dan dari luar survei.

Perkembangan penelitian dalam pendugaan area kecil saat ini menunjukkan kemajuan yang pesat. Berbagai persoalan di dalamnya dicoba untuk dicarikan penyelesaiannya oleh banyak peneliti. Namun demikian, dalam aplikasinya SAE masih menemui berbagai kendala seperti yang juga ditemukan penulis pada saat melakukan tahapan eksplorasi metode. Kendala-kendala yang ditemukan dalam SAE baik yang ditemukan penulis maupun dikemukakan oleh peneliti-peneliti lain diantaranya :

1. rasio keragaman antar area kecil (komponen ragam) dibandingkan dengan keragaman total masih cukup besar,

2. ketidakcocokan model dalam model yang melibatkan peubah penyerta (*auxiliary variable*) masih sering terjadi,
3. penduga keragaman antar area kecil dengan metode kemungkinan maksimum terkendala (*restricted/residual maximum likelihood*, REML) masih bisa diperoleh nol (Lahiri, 2009),
4. permasalahan pola hubungan antara data survei dengan informasi spasial, dimana informasi spasial biasanya terbentuk dari seluruh area dalam populasi sedangkan area dalam survei mungkin tidak mencakup seluruh populasi (Chandra, Salvati dan Chambers, 2007), dan
5. sering ditemukannya data-data pencilan sehingga mengganggu akurasi pendugaan parameter (Chambers dan Tzavidis, 2006).

Pembahasan lebih lanjut dalam disertasi ini dibatasi untuk dua kendala pertama yaitu (i) masalah rasio keragaman antar area kecil terhadap keragaman total yang masih cukup besar serta (ii) masih sering terjadinya ketidakcocokan model dalam model yang melibatkan peubah penyerta.

Kendala pertama mengakibatkan penduga tidak langsung untuk model baku *empirical best linear unbiased prediction* (EBLUP) atau *empirical Bayes* (EB) yang dihasilkan selalu akan lebih kuat dipengaruhi oleh nilai penduga langsungnya. Hal ini dapat dijelaskan sebagai berikut. Berdasarkan model dasar pendugaan area kecil Fay dan Herriot (1979), penduga terbaik (*best prediction*, BP) bagi parameter yang menjadi perhatian pada area ke- i (θ_i) dengan meminimumkan kuadrat tengah galat (*mean square error*, MSE) adalah

$$\hat{\theta}_i^{BP} = E(\theta_i | y_i, \beta, \tau^2) = \gamma_i y_i + (1 - \gamma_i) x_i^T \beta$$

dengan $\gamma_i = \tau^2 / (\tau^2 + \sigma_i^2)$ dan τ^2 adalah keragaman antar area kecil serta σ_i^2 adalah keragaman karena galat contoh (*sampling error*) di dalam setiap area kecil. Dengan demikian, jika τ^2 sangat besar dibandingkan $(\tau^2 + \sigma_i^2)$ maka γ_i akan mendekati satu dan $\hat{\theta}_i^{BP}$ cenderung akan sama dengan penduga langsungnya (y_i).

Selanjutnya kendala kedua mengakibatkan lemahnya pengaruh peubah penyerta dalam memberikan koreksi terhadap penduga langsung. Hal ini disebabkan ketidaktepatan model dalam menggambarkan pola hubungan antara peubah yang menjadi perhatian dengan satu atau lebih peubah penyerta. Seperti diketahui, model dasar pendugaan area kecil menggunakan model linier, padahal dapat terjadi sesungguhnya data tidak mengikuti pola hubungan linier.

Beberapa penelitian yang terkait dengan masalah di atas diantaranya pengembangan *model-based design estimator* (MBDE), *general regression* (GREG), serta *calibration model*. Chambers dan Chandra (2006) menyampaikan bahwa MBDE memberikan indikasi yang lebih baik (*robust*) dibandingkan dengan EBLUP jika terjadi ketidakcocokan pada model area kecil yang sedang diteliti. Lebih lanjut Chandra, Salvati dan Chambers (2007) menyampaikan hasil kajiannya tentang kemampuan MBDE dalam mengatasi pengaruh spasial dalam model linier SAE. Namun demikian, masih ada beberapa masalah yang dihadapi seperti:

- (i) kemungkinan pembobot yang dihasilkan untuk MBDE bernilai negatif sehingga tidak memungkinkan untuk digunakan dalam pendugaan. Pertanyaannya adalah bagaimana cara memodifikasi pembobot tersebut sehingga memiliki sifat *strictly positive*,

- (ii) kemungkinan hubungan yang tidak linier antara data survei dengan peubah penyerta, sehingga diperlukannya upaya transformasi atau teknik lainnya, dan
- (iii) Chandra, Salvati dan Chambers (2007) dalam papernya masih mengasumsikan seluruh area memiliki contoh (*sample units*) walaupun disadari hal ini dalam prakteknya sulit untuk terpenuhi.

Beberapa penelitian lain yang sejalan dengan kendala yang dihadapi penulis adalah yang dilakukan oleh Zheng dan Little (2004) yang menggunakan pendekatan P-spline untuk menyelesaikan pengaruh acak dalam penarikan contoh acak gerombol dua tahap. Sementara itu, Kurnia dan Notodiputro (2008) telah melakukan kajian awal tentang penggunaan *generalized additive mixed models* (GAMM) serta Kurnia, Notodiputro dan Ibrahim (2007) menggunakan pendekatan nonparametrik P-spline untuk mengurangi pengaruh ketidakcocokan model. Hasil kajian awal ini menunjukkan bahwa pendekatan yang dilakukan mampu memberikan perbaikan akurasi dan presisi jika dibandingkan dengan pendekatan baku EBLUP maupun *empirical Bayes* (EB). Pendekatan P-spline juga cukup memberikan perbaikan yang memuaskan jika data mengandung pengaruh spasial seperti yang disampaikan Opsomer, et.al (2008). Namun demikian, seperti yang dilaporkan dalam Kurnia dan Notodiputro (2008) maupun Kurnia, Notodiputro dan Ibrahim (2007), karena masalah utama yang dihadapi adalah rasio keragaman antar area kecil dibandingkan dengan keragaman total yang cukup besar, maka metode yang digunakan walaupun menunjukkan perbaikan tetapi belum memuaskan karena pengaruhnya yang relatif kecil.

Dua masalah yang menjadi fokus pada disertasi ini sudah dicoba oleh beberapa peneliti untuk dicarikan solusinya dari berbagai cara pandang seperti yang kemukakan oleh Chambers dan Chandra (2006) melalui pendekatan MBDE, Chambers dan Tzavidis (2006) melalui pendekatan

regresi *M-Quantile*, Li (2006) mengembangkan pendekatan *Automated* GREG, Kurnia dan Notodiputro (2008) mengembangkan model aditif, serta Zheng dan Little (2004), Kurnia, Notodiputro dan Ibrahim (2007) dan Opsomer et.al (2008) mengembangkan model pendekatan nonparametrik. Namun demikian, persoalan yang dihadapi tersebut belum mampu dijawab secara tuntas. Dengan kekhasan kasus data yang dihadapi, yaitu bersifat *contextual covariate* (peubah penyerta hanya tersedia pada taraf area) serta pola sebaran peubah yang menjadi perhatian tidak simetrik atau pola hubungan peubah yang menjadi perhatian dengan peubah penyerta tidak linier, dalam penelitian ini **dikembangkan solusi alternatif melalui pendekatan transformasi logaritma pada peubah yang menjadi perhatian serta teknik pendugaan parameternya menggunakan metode prediksi terbaik empirik.**

Berdasarkan uraian di atas, penulis mengajukan tiga pokok pertanyaan penelitian yang akan dijawab dalam disertasi ini. Pokok-pokok pertanyaan tersebut adalah :

1. Apakah dengan melakukan transformasi logaritma pada peubah yang menjadi perhatian mampu mendapatkan penduga yang lebih baik ?
2. Bagaimana bentuk kuadrat tengah galat dari penduga yang diperoleh ?
3. Bagaimana sifat statistik dan performa dari penduga yang diperoleh ?

1.2. Tujuan Penelitian

1. Memperbaiki pendugaan area kecil ketika ditemukan masalah rasio keragaman antar area kecil terhadap keragaman total cukup besar serta pola hubungan peubah yang menjadi perhatian dengan peubah penyerta tidak linier atau peubah yang menjadi perhatian tidak mengikuti pola sebaran normal.

2. Mengembangkan pendugaan area kecil berdasarkan model transformasi logaritma.
3. Menjabarkan kuadrat tengah galat dari penduga yang dikembangkan.
4. Mengevaluasi performa dari penduga yang diperoleh melalui kajian simulasi.

1.3. Ruang Lingkup Penelitian

Ruang lingkup penelitian secara umum adalah melakukan pengembangan metode pendugaan area kecil yang bersifat kekar terutama dalam mengatasi dua masalah yang dihadapi, yaitu (i) besarnya rasio keragaman antar area kecil terhadap keragaman total dan (ii) kemungkinan terjadinya ketidakcocokan model.

Secara spesifik, penelitian ini fokus pada pengembangan prediksi terbaik empirik (*empirical best prediction*, EBP) dalam SAE untuk model transformasi logaritma pada peubah yang menjadi perhatian dengan peubah penyerta bersifat *contextual covariate*. Pemilihan fokus penelitian ini didasari oleh :

1. berdasarkan berbagai kajian awal (eksplorasi metode) seperti yang dilaporkan pada Bab III dan Bab IV dalam disertasi ini, diindikasikan bahwa penerapan berbagai metode baku SAE dalam kasus data BPS (Susenas dan Podes) walaupun memberikan perbaikan akurasi dan presisi hasil pendugaan namun masih belum memuaskan, dan
2. kenyataan bahwa pola hubungan data sosial dan ekonomi sulit memenuhi hubungan linier dalam skala pengukuran asli, dan baru dipenuhi setelah dilakukan transformasi baik pada peubah respon, peubah bebas atau pada kedua-duanya. Walaupun masalah transformasi sudah banyak dikaji dalam penelitian-penelitian statistika, namun sepengetahuan penulis masih belum banyak penelitian sejenis dalam bidang SAE.

1.4. Kebaruan Disertasi

Penelitian ini berupaya untuk menemukan alternatif penyelesaian pendugaan nilai tengah pada pendugaan area kecil melalui pendekatan transformasi logaritma pada peubah yang menjadi perhatian. Metode yang digunakan di dalam disertasi ini adalah metode prediksi terbaik empirik. Pada penelitian ini juga dijabarkan bentuk kuadrat tengah galat dari penduga yang dikembangkan.

Penelitian sejenis dilakukan oleh Fabrizi dan Trivisano (2009) dengan mencoba pendekatan Bayes berhierarki. Kendala utama yang dilaporkan adalah *complicated*-nya pengukuran keragaman dari penduga yang diperoleh serta masalah penetapan sebaran prior dari parameter-parameter model.

1.5. Sistematika Penulisan Disertasi

Disertasi ini disusun dari serangkaian penelitian yang dilakukan penulis dalam bidang pendugaan area kecil. Beberapa bagian dari disertasi ini juga sudah dipublikasikan baik dalam seminar nasional maupun internasional serta jurnal ilmiah. Oleh karena itu, sistematika penulisan disertasi ini secara umum dibagi dalam tiga bagian besar yang membahas perkembangan pendugaan area kecil (*state of the art*), beberapa contoh penerapan dan permasalahan yang dihadapi dalam kasus data BPS (Susenas dan Podes), serta pengembangan model transformasi logaritma pada peubah yang menjadi perhatian yang diharapkan juga mampu mereduksi pengaruh ketidaktepatan pemodelan dan besarnya rasio keragaman antar area kecil terhadap keragaman total.

Pada Bab II, disajikan review perkembangan pendugaan area kecil (*state of the art*) sebagai pengantar untuk memahami permasalahan serta konsep dasar pendugaan parameter dalam SAE. Dalam bab ini juga dibahas berbagai teknik pendugaan baku dalam pendugaan area

kecil meliputi konsep pendugaan berbasis disain survei, penduga BLUP/EBLUP, dan penduga Bayes.

Eksplorasi beberapa metode baku SAE dan penerapannya untuk kasus data BPS disajikan pada Bab III dan Bab IV. Pada Bab III dibahas secara mendalam teknik pendugaan area kecil berbasis disain survei. Dalam bab ini, penulis mencoba mengatasi masalah penelitian dengan pendekatan pendugaan area kecil berbasis disain survei, khususnya GREG dan modifikasinya. Alasan pendekatan ini diajukan karena secara formula tidak dipengaruhi langsung oleh rasio keragaman antar area kecil terhadap keragaman total. Adapun pada Bab IV, penulis membahas teknik pendugaan area kecil berbasis model. Indikasi yang diperoleh dari eksplorasi metode pada Bab IV ini mendasari pengembangan pendugaan statistik area kecil untuk model transformasi.

Pengembangan prediksi terbaik empirik untuk model transformasi logaritma dalam pendugaan area kecil merupakan inti dari disertasi ini. Kajian lengkap untuk permasalahan tersebut dibahas pada Bab V. Bentuk transformasi logaritma pada peubah yang menjadi perhatian dipilih karena eksplorasi pada data asli (data sebelum dilakukan transformasi) serta sisaan modelnya menunjukkan pola tidak normal. Adapun model prediksi terbaik empirik digunakan karena pendekatan ini secara formula juga tidak dipengaruhi langsung oleh rasio keragaman antar area kecil terhadap keragaman total.

Bab VI merupakan pembahasan umum yang menyajikan keterkaitan antar bab dalam disertasi. Kelebihan serta kekurangan metode-metode yang telah dijabarkan dalam bab-bab sebelumnya dibahas secara menyeluruh dalam bab ini. Sebagai penutup dari disertasi ini disampaikan kesimpulan dan saran pada Bab VII.

BAB II

METODE PENDUGAAN AREA KECIL

Disertasi ini membahas pengembangan teknik pendugaan area kecil khususnya pengembangan prediksi terbaik empirik untuk model transformasi. Berikut disajikan kajian ulang metode-metode baku dalam pendugaan area kecil sebagai pengantar untuk memahami permasalahan serta konsep dasar pendugaan parameter dalam SAE. Dalam bab ini dibahas model dasar pendugaan area kecil, konsep pendugaan berbasis disain survei, penduga BLUP/EBLUP, dan penduga Bayes.

2.1. Model Dasar Pendugaan Area Kecil

Penduga parameter yang bersifat kekar untuk suatu area kecil, saat ini merupakan tujuan penting bagi banyak badan dan penelitian dalam SAE. Area kecil tersebut didefinisikan sebagai himpunan bagian dari populasi dimana suatu peubah menjadi perhatian. Pendekatan klasik untuk menduga parameter area kecil ke- i (θ_i) didasarkan pada aplikasi model disain penarikan contoh (*design-based*), pendugaan tersebut kemudian disebut pendugaan langsung (*direct estimation*). Metode pendugaan tersebut menimbulkan dua permasalahan penting. *Pertama*, penduga yang dihasilkan merupakan penduga tak bias tetapi memiliki ragam yang besar karena diperoleh dari ukuran contoh yang kecil. *Kedua*, apabila pada suatu area kecil ke- i tidak terwakili di dalam survei, maka tidak memungkinkan dilakukan pendugaan secara langsung.

Adakalanya kita memiliki informasi tambahan yang dapat digunakan untuk pendugaan pada area kecil. Dalam beberapa kasus kita bisa memperoleh informasi tentang parameter yang menjadi perhatian dari area kecil lain yang memiliki karakteristik serupa, atau nilai pada waktu yang lalu, atau nilai dari peubah yang memiliki hubungan

dengan peubah yang sedang diamati. Pendugaan paramater dan inferensinya yang menggunakan informasi tambahan tersebut, dinamakan pendugaan tidak langsung (*indirect estimation* atau *model-based estimation*). Metode dengan memanfaatkan informasi tambahan tersebut secara statistik memiliki sifat "meminjam kekuatan" (*borrowing strength*) informasi dari hubungan antara peubah respon dengan informasi yang ditambahkan. Metode ini memiliki sejarah yang panjang tetapi baru mendapat perhatian dalam beberapa dekade terakhir untuk digunakan sebagai pendekatan pada pendugaan parameter area kecil. Dalam hal ini, dua ide utama digunakan untuk mengembangkan model pendugaan area kecil yaitu (1) asumsi bahwa keragaman di dalam area kecil, peubah respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada informasi tambahan, kemudian disebut model pengaruh tetap (*fixed effect models*), dan (2) asumsi keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan dan merupakan pengaruh acak area kecil (*random effect*).

Model pengaruh tetap menerangkan seluruh keragaman peubah respon di dalam area kecil oleh keragaman faktor-faktor yang diketahui. Pendugaan karakteristik area kecil berdasarkan model pengaruh tetap merujuk pada *synthetic estimator* (Levy dan French, 1977), *composite estimator* (Schaible et. al., 1977) dan *prediction estimator* (Holt et. al., 1979; Sarndal, 1984; Marker, 1999). Secara lengkap pembahasan penduga-penduga tersebut disajikan dalam Rao (2003).

Fay dan Herriot (1979) secara umum menggunakan model pengaruh campuran (*linear mixed model*) dengan pengaruh acak yang hanya mengandung intersep, dengan kata lain model hanya meliputi pengaruh acak area, untuk menduga rata-rata pendapatan sub-populasi (<1000) menggunakan data sensus 1970 di Amerika Serikat. Model Fay-Herriot tersebut merupakan model dasar bagi

pengembangan pemodelan area kecil yaitu $y_i = \theta_i + e_i$; $\theta_i = x_i^T \beta + v_i$, dimana e_i dan v_i saling bebas dengan $E(e_i) = E(v_i) = 0$ serta $\text{Var}(e_i) = \sigma_i^2$ dan $\text{Var}(v_i) = \tau^2$ ($i = 1, 2, 3, \dots, m$). y_i adalah penduga langsung bagi area ke- i dan diperoleh dari data survei yang bersesuaian, θ_i merupakan parameter yang menjadi perhatian, e_i adalah galat contoh, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ adalah peubah penyerta, dan v_i adalah pengaruh acak area.

Russo et. al. (2005) menjabarkan lebih lanjut model area kecil dengan memperjelas pengaruh acak sub-populasi di dalam model sebagai berikut :

1. $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$
vektor data pendukung (peubah penyerta)
2. $\theta_i = x_i^T \beta + v_i$ untuk $i = 1, 2, \dots, m$
merupakan parameter yang menjadi perhatian dan diasumsikan memiliki hubungan dengan peubah penyerta pada (1) sedang v_i pengaruh acak dengan nilai tengah nol dan ragam τ^2
3. $y_i = \theta_i + e_i$
penduga langsung untuk sub-populasi ke- i yang merupakan fungsi linier dari parameter yang menjadi perhatian dan galat contoh e_i
4. $y_i = x_i^T \beta + v_i + e_i$ untuk $i = 1, 2, \dots, m$
model tersebut terdiri dari pengaruh acak dan pengaruh tetap sehingga merupakan bentuk khusus dari model linier campuran dengan struktur peragam yang diagonal.

Model regresi merupakan upaya untuk membentuk model umum dan memanfaatkan kekuatan dan keakuratan pendugaan pada level populasi, sedangkan deviasi sub-populasi untuk menangkap kekhasan

yang terjadi pada setiap sub-populasi dan bersifat acak. Dengan demikian jika kita hanya akan memanfaatkan informasi umum maka $\theta_i = x_i^T \beta$, dan jika pengaruh umum dan lokal kita adopsi, diperoleh $\theta_i = x_i^T \beta + v_i$.

Secara statistika model pada point (4) di atas melibatkan pengaruh acak akibat disain penarikan contoh (e_i) dan pengaruh acak pemodelan sub-populasi (v_i). Model tersebut merupakan bentuk khusus dari model linier campuran.

Salah satu sifat yang menarik dari model campuran adalah kemampuannya dalam menduga kombinasi linier dari pengaruh tetap dan pengaruh acak. Dalam papernya, Henderson (1953, 1975) mengembangkan teknik penyelesaian model pengaruh campuran *best linear unbiased prediction* (BLUP). BLUP menjadi metode yang *powerful* dan digunakan secara luas. Namun demikian, BLUP yang dikembangkan Henderson (1948-1975) mengasumsikan diketahuinya ragam pengaruh acak dalam model campuran (komponen ragam). Dalam praktek, komponen ragam tidak diketahui dan harus diduga berdasarkan data. Harville (1977) melakukan review terhadap beberapa metode pendugaan komponen ragam, dengan memasukkan metode kemungkinan maksimum dan metode kemungkinan maksimum terkendala serta tiga metode yang diajukan Henderson. Penduga BLUP yang diperoleh ketika komponen ragam yang tidak diketahui disubstitusi oleh penduganya, disebut *empirical best linear unbiased predictor* (EBLUP) seperti yang dikembangkan Harville (1991).

Model campuran telah digunakan untuk meningkatkan akurasi pendugaan pada kasus area kecil berdasarkan data survei dan data sensus oleh Fay dan Herriot (1979), Ghosh dan Rao (1994), Rao (1999), Pfeiffermann (1999), Kubokawa (2006) serta Jiang dan Lahiri

(2006). Pada aplikasi ini, model campuran diturunkan dari konsep bahwa vektor nilai populasi terbatas merupakan realisasi dari superpopulasi. Dalam kasus ini, pendugaan rata-rata area kecil ekuivalen dengan pendugaan dari perwujudan pengaruh acak area yang tidak diobservasi dalam model campuran untuk sebaran superpopulasi yang dicari rata-ratanya.

Selain EBLUP, pendugaan dan inferensi pada pendugaan area kecil juga menggunakan *empirical Bayes* (EB) dan *hierarchical Bayes* (HB). Pada pendekatan EB, pendugaan dan inferensi berdasarkan pada sebaran posterior yang parameternya diduga dari data. Persoalan mendasar dalam metode EB adalah lebih sulitnya dalam pendugaan galat model (*uncertainty model*). Beberapa pendekatan telah dikembangkan untuk mengatasi persoalan ini, di antaranya yang umum digunakan adalah metode *delta* dan *bootstrap*. Deely dan Lindley (1981) serta Kass dan Steffey (1989) telah mendiskusikan metode *delta*, sementara Laird dan Louis (1987), Butar (1997) serta Butar dan Lahiri (2003) mengembangkan metode *bootstrap*. Selain itu, pendekatan lain yang juga diajukan adalah metode jackknife seperti dibahas oleh Wan (1999), Chen (2001), Jiang, Lahiri dan Wan (2002), serta Chen dan Lahiri (2005).

Adapun pada pendekatan HB, parameter model yang tidak diketahui (termasuk komponen ragam) diperlakukan sebagai komponen acak yang masing-masing memiliki sebaran prior tertentu. Ghosh dan Rao (1994) mengulas penggunaan HB pada pendugaan area kecil, sedangkan secara spesifik Maiti (1998) menggunakan *non-informative prior* untuk kasus hiperparameter pada penggunaan HB. You dan Rao (2000) menggunakan HB untuk menduga rata-rata area kecil berdasarkan model pengaruh acak.

Metode HB mempunyai keuntungan karena pemodelannya dilakukan secara bertahap. Setiap tahap bisa relatif sederhana dan mudah

dipahami, meskipun proses pemodelannya secara keseluruhan sangat rumit. Di samping itu, karena sebaran prior dari parameter modelnya didasarkan pada beberapa asumsi, maka penduga HB mempunyai kuadrat tengah galat yang lebih kecil dibandingkan dengan penduga BLUP (Ghosh dan Rao, 1994).

2.2. Konsep Pendugaan Berbasis Disain

Model statistik mempunyai sejarah yang panjang dalam analisis data survei. Salah satu contoh penelitian awal yang populer adalah pada saat Watson menggunakan model regresi luas daun terhadap berat daun dalam menduga rata-rata luas daun. Namun demikian, perkembangan konsep ini tidak semarak sampai dengan tahun 70-an, dimana informasi tambahan hanya digunakan dalam disain survei.

Dimulai tahun 70-an, dimana Brewer, Cochran dan Royall melakukan kajian dengan teknik penduga rasio, penggunaan peubah penyerta dalam melakukan pendugaan parameter mulai banyak digunakan para peneliti.

Dalam konteks pendugaan berbasis disain, peubah penyerta dapat digunakan pada tahap pendugaan parameter baik melalui teknik kalibrasi maupun pemodelan. Namun demikian, pada pendekatan *model-assisted*, pemodelan statistik yang menghubungkan peubah yang menjadi perhatian dengan peubah penyerta dimanfaatkan untuk menduga nilai-nilai dari peubah yang menjadi perhatian. Model paling umum yang digunakan dalam *model-assisted* adalah *generalized regression* (GREG) yang menggunakan pola hubungan linier dan hanya mengandung komponen tetap. Pembahasan lebih lanjut model ini bisa dilihat dalam Särndal, Swensson dan Wretman (1992).

GREG merupakan suatu metode pendugaan parameter dalam survei sampling yang memungkinkan untuk menggunakan beberapa informasi tambahan dan dirancang untuk meningkatkan keakuratan

dengan menggunakan informasi tambahan x_i yang berkorelasi dengan peubah yang menjadi perhatian, y_i . Metode ini dapat digunakan untuk menduga total populasi, nilai tengah populasi ataupun proporsi populasi. Di sisi lain GREG juga relatif sederhana dan mudah untuk digunakan dalam praktek. Kelebihan lainnya adalah dalam kondisi yang baku GREG masih bersifat *approximately unbiased* (Rao, 2003).

Model-assisted GREG secara umum ditulis sebagai berikut

$$\hat{T}_i(y) = \sum_{\forall j} \hat{y}_{ij}(\cdot) + \sum_{\forall j} w_{ij} e_{ij} \text{ dengan } e_{ij} = y_{ij} - \hat{y}_{ij}.$$

Adapun model standar GREG didasarkan atas model linier $y_i = x_i^T \beta + \varepsilon_i$ dengan $\varepsilon_i \sim N(0, \sigma_i^2)$. Penduga GREG-linier untuk rata-rata area kecil dengan disain penarikan contoh acak sederhana (*simple random sampling*, SRS) adalah

$$\begin{aligned} \hat{Y}_i^{GREG} &= \frac{1}{W_i} \sum_{\forall j} w_{ij} y_{ij} + \left(\bar{X}_i - \frac{1}{W_i} \sum_{\forall j} w_{ij} x_{ij} \right)^T \hat{\beta} \\ &= \hat{Y}_i^{DIRECT} + (\bar{X}_i - \hat{\bar{X}}_i) \hat{\beta} \end{aligned}$$

dengan :

- $\bar{X}_i = (\bar{X}_{i1}, \dots, \bar{X}_{ip})^T$ adalah vektor dari nilai tengah populasi p peubah penyerta, dan
- $W_i = \sum_{\forall j} w_{ij}$, $w_{ij} = 1 / \pi_{ij}$, dengan π_{ij} adalah peluang contoh.

Penduga langsung (\hat{Y}_i^{DIRECT}) dan besarnya pembobot (w_{ij}) diperoleh berdasarkan disain penarikan contoh yang digunakan dalam pelaksanaan survei. Penduga bagi koefisien regresi (β) dapat diperoleh dengan menggunakan metode *generalized weighted least square* (GWLS) :

$$\hat{\beta} = \left(\sum_{\forall ij} w_{ij} \frac{x_{ij} x_{ij}^T}{\sigma_i^2} \right)^{-1} \sum_{\forall ij} w_{ij} \frac{x_{ij} y_{ij}}{\sigma_i^2}.$$

Metode GREG-linier ini sangat tergantung pada asumsi yang seringkali tidak terpenuhi dalam prakteknya. Ketika terdapat pencilan dalam data, GREG berdasarkan metode kuadrat terkecil seringkali memiliki performa yang rendah (Kurnia, Sartono dan Wulandari, 2007).

Regresi kekar digunakan untuk memberikan metode alternatif yang sama baiknya dengan metode kuadrat terkecil, tetapi tidak terlalu dipengaruhi oleh adanya pencilan atau pelanggaran asumsi model. Regresi kekar yang banyak digunakan dalam praktek adalah metode yang diperkenalkan oleh Huber, *M-Estimator*, yang bekerja dengan meminimumkan fungsi tertentu dalam data.

Andaikan, dalam model dasar dengan contoh yang berasal dari sebaran kontinu dengan fungsi sebaran kumulatif $F(x)$ dan fungsi kepekatan $f(x)$, prinsip untuk menduga nilai tengah μ dari $T_n = T_n(x_1, \dots, x_n)$ dilakukan dengan meminimumkan fungsi

$$\sum_{i=1}^n \rho(x_i - T_n)$$

atau dengan menyelesaikan persamaan

$$\sum_{i=1}^n \Psi(x_i - T_n) = 0$$

dimana $\rho(\cdot) = -\log f(\cdot)$, $\Psi(x|\theta) = \frac{\partial}{\partial \theta} \rho(x|\theta)$ dan $T_n = \frac{\sum w_i x_i}{\sum w_i}$.

M-estimator pada prinsipnya mendefinisikan masalah pemilihan fungsi Ψ yang memenuhi prinsip efisiensi dan kekekaran (*robustness*). Efisiensi pada fungsi F berarti mendapatkan masalah lokasi dengan mengambil Ψ proporsional dari logaritma fungsi kemungkinan. Kekekaran diperoleh dengan memilih Ψ yang sesuai dan dibatasi untuk mengurangi pengaruh dari proporsi sebagian kecil pengamatan. Jika F berasal dari sebaran normal, maka fungsi Ψ berdasarkan metode kemungkinan maksimum adalah fungsi identitas yang tidak dibatasi. Fungsi Ψ ini disesuaikan menjadi (Huber, 1996)

$$\Psi_k(x) = \begin{cases} -k, & x < -k \\ x, & -k \leq x \leq k \\ +k, & +k < x \end{cases}$$

Penghitungan pendugaan Huber *M-Estimator* menggunakan berbagai algoritma, salah satunya adalah *iterative reweighted least squares* (IRLS).

Metode *robust* lain yang cukup sederhana dalam aplikasinya adalah *L-estimator*. Penduga ini dikenal juga sebagai *Linear Order Statistic Estimator*. Pendugaan koefisien regresi berdasarkan *L-estimator* dapat dilakukan dengan beberapa teknik : (1) Metode *Rupert dan Carroll*, (2) Metode *Koenker and Bassett*, (3) Metode *Trimming Absolute Order Residuals*, (4) *Tukey's biweight* dan (5) *Siegel's repeated median*.

Lebih lanjut, jika kita dihadapkan pada kasus *domain class frequencies* atau respon survei bersifat Bernoulli, maka sudah tentu GREG-linier memiliki keterbatasan. Lebih detail pembahasan keterbatasan alat analisis yang linier bisa dilihat pada McCullagh dan Nelder (1989) maupun Agresti (2002) dan Agresti (2007).

Misalkan peubah respon kita adalah y_{ij} bersifat biner dan $p_i = P(y_{ij} = 1)$, maka koefisien $\hat{\beta}$ dapat diperoleh dengan memaksimum fungsi kemungkinan

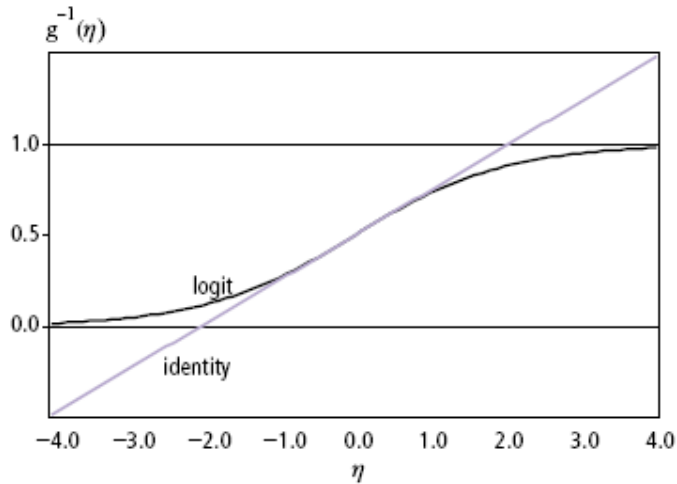
$$l(\beta | y_i) = \sum_{\forall i} w_i [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$

dengan fungsi hubung $\eta_i = \log(p_i / (1 - p_i)) = x_i^T \beta$ atau menggunakan fungsi hubung lainnya : probit atau complementary log-log.

Dengan mengadopsi bentuk umum GREG, GREG-Logistik dapat ditulis sebagai berikut

$$\hat{T}_i(y) = \sum_{\forall j} \hat{y}_{ij}(\cdot) + \sum_{\forall j} w_i e_{ij}$$

dengan $e_{ij} = y_{ij} - \hat{y}_{ij}$ dan $\hat{y}_{ij} = \frac{\exp(x_i^T \hat{\beta})}{1 + \exp(x_i^T \hat{\beta})}$.



Gambar 2.1. Plot kurva linier dan logistik

Gambar 2.1 menunjukkan bahwa garis linier mendekati kurva fungsi logistik jika p berkisar dalam selang $(0.25, 0.75)$. Tetapi jika p dekat dengan batas intervalnya $(0$ atau $1)$ maka perbedaan dari fungsi logistik dan linier menjadi sangat jelas. Pendugaan dari model linier bersifat tidak terbatas, sehingga pendugaan galatnya pun bersifat demikian. Namun model logistik, karena pendugaannya terbatas, maka galatnya pun berkisar pada selang $(-1, 1)$. Dengan demikian, jika p diduga di sekitar batas intervalnya, maka kesalahan pendugaan dengan model linier akan lebih besar dibandingkan kesalahan yang dihasilkan model logistik.

Lebih lanjut, jika dukungan peubah penyerta cukup bagus, maka GREG-logistik diharapkan akan lebih akurat dibanding GREG-linier. Akan tetapi, jika peubah penyerta tidak begitu bagus menerangkan parameter yang menjadi perhatian, menggunakan GREG-logistik maupun GREG-linier diharapkan mampu memberikan akurasi yang sama untuk pendugaan p pada selang $(0.25, 0.75)$.

Pengembangan lain dari kelompok penduga langsung adalah *model based design estimator* (MBDE). Chambers dan Chandra (2006) mengembangkan MBDE berdasarkan pada model umum yang digunakan dalam pemodelan pendugaan area kecil yaitu model Fay-Herriot

$$y_i = x_i^T \beta + v_i + e_i$$

dengan e_i dan v_i saling bebas, $E(e_i) = E(v_i) = 0$ serta $\text{Var}(e_i) = \sigma_i^2$ dan $\text{Var}(v_i) = \tau^2$ untuk $i = 1, 2, 3, \dots, m$ dengan m adalah banyaknya area kecil yang menjadi perhatian.

Pendugaan $\theta_i = x_i^T \beta + v_i$ dapat menggunakan teknik EBLUP, EB ataupun HB. Lebih lanjut Chandra, Salvati dan Chambers (2007) menunjukkan bahwa berdasarkan model linier campuran dapat diperoleh suatu pembobot EBLUP untuk total populasi (Y), yaitu :

$$w_{EBLUP} = (w_{j,EBLUP}) = 1_n + \hat{H}^T (X^T 1_N - X_s^T 1_n) + (I_n - \hat{H}^T X_s^T) \hat{V}_{ss}^{-1} \hat{V}_{sr} 1_r$$

dengan

$$\hat{H} = \left(\sum_i X_{is}^T \hat{V}_{iss}^{-1} X_{is} \right)^{-1} \left(\sum_i X_{is}^T \hat{V}_{iss}^{-1} \right) = \left(X_s^T \hat{V}_{ss}^{-1} X_s \right)^{-1} \left(X_s^T \hat{V}_{ss}^{-1} \right),$$

$$V = \begin{bmatrix} V_{ss} & V_{sr} \\ V_{rs} & V_{ss} \end{bmatrix} \text{ dimana } s \text{ adalah indeks untuk unit dalam contoh}$$

sedangkan r adalah indeks untuk unit bukan anggota contoh.

Penduga MBDE (Chambers dan Chandra, 2006) untuk nilai tengah area kecil ke- i didefinisikan sebagai berikut

$$\hat{\bar{Y}}_{i,MBD} = \sum_{j \in s_i} w_{j,EBLUP} y_j / \sum_{j \in s_i} w_{j,EBLUP}$$

sedangkan kuadrat tengah galat dari suatu penduga MBDE adalah

$$M(\hat{\bar{Y}}_{i,MBD}) = v(\hat{\bar{Y}}_{i,MBD}) + \left\{ b(\hat{\bar{Y}}_{i,MBD}) \right\}^2$$

dengan

$v(\hat{Y}_{i,MBD}) = \sum_{s_j} \lambda_j (y_j - x_j^T \hat{\beta})^2$ adalah dugaan ragam penduga MBDE dan $\lambda_j = N_i^{-2} \{a_j^2 + (N_i - n_i)(n_i - 1)^{-1}\}$ serta $a_j = (\sum_{s_i} w_k)^{-1} (N_i w_j - \sum_{s_i} w_k)$. Sedangkan $b(\hat{Y}_{i,MBD}) = (\hat{X}_{i,MBD} - \hat{X}_i) \hat{\beta}$ adalah dugaan bias penduga MBDE. Ini menandakan bahwa $\hat{X}_{i,MBD}$ sebagai rata-rata terbobot nilai contoh dari peubah penyerta pada area kecil ke-i berdasarkan pembobot EBLUP.

2.3. Konsep dan Sifat Statistik Penduga EBLUP

Perhatikan model yang dikembangkan oleh Fay dan Herriot (1979), dimana model tersebut yang kemudian digunakan sebagai dasar pengembangan model pendugaan area kecil

$$y_i = x_i^T \beta + v_i + e_i \quad (2.1)$$

dengan e_i dan v_i saling bebas serta $v_i \sim N(0, \tau^2)$ dan $e_i \sim N(0, \sigma_i^2)$ untuk $i = 1, 2, \dots, m$. Untuk selanjutnya kita perhatikan kasus dimana β dan τ^2 tidak diketahui, tetapi σ_i^2 ($i = 1, 2, \dots, m$) diasumsikan diketahui.

Penduga terbaik (*best prediction*, BP) bagi $\theta_i = x_i^T \beta + v_i$ berdasarkan model (2.1) dengan meminimumkan kuadrat tengah galat jika β dan τ^2 diketahui adalah

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i | \beta, \tau^2) = \gamma_i y_i + (1 - \gamma_i) x_i^T \beta \quad (2.2)$$

dengan $\gamma_i = \tau^2 / (\tau^2 + \sigma_i^2)$.

Misalkan $X = (x_1, x_2, \dots, x_m)^T$, $Y = (y_1, y_2, \dots, y_m)^T$ dan $V = \text{Diag}(\tau^2 + \sigma_1^2, \tau^2 + \sigma_2^2, \dots, \tau^2 + \sigma_m^2)$. Jika τ^2 diketahui, β dapat diduga dengan metode kemungkinan maksimum (*maximum likelihood*), yaitu

$$\log L(\beta, V) = -\frac{1}{2} \log |V| - \frac{1}{2} (Y - X\beta)^T V^{-1} (Y - X\beta)$$

$$\frac{d}{d\beta} \log L(\beta, V) = X^T V^{-1} (Y - X\beta) = X^T V^{-1} Y - (X^T V^{-1} X) \beta$$

$$\Leftrightarrow (X^T V^{-1} X) \beta = X^T V^{-1} Y$$

$$\Leftrightarrow \beta = (X^T V^{-1} X)^{-1} X^T V^{-1} Y.$$

Misal kita notasikan penduga tersebut sebagai berikut

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

dan dengan mensubstitusi β oleh $\hat{\beta}$ pada persamaan (2.2), maka diperoleh

$$\hat{\theta}_i^{BLUP} = \hat{\theta}_i(y_i | \hat{\beta}, \tau^2) = \gamma_i y_i + (1 - \gamma_i) x_i^T \hat{\beta}. \quad (2.3)$$

Dalam praktek, baik β maupun τ^2 biasanya tidak diketahui sehingga untuk kasus pendugaan $\theta_i = x_i^T \beta + v_i$ dengan BLUP maka τ^2 terlebih dulu harus diduga. Untuk menduga τ^2 dapat menggunakan metode kemungkinan maksimum (*maximum likelihood*, ML), metode kemungkinan maksimum terkendala (*restricted/residual maximum likelihood*, REML) atau metode momen. Jiang (1996) memperlihatkan bahwa pendugaan komponen ragam, dalam hal ini τ^2 , pada model linier campuran dengan menggunakan REML akan menghasilkan penduga yang konsisten walaupun ada penyimpangan asumsi kenormalan. Dengan demikian $\hat{\theta}_i$ dengan menggunakan $\hat{\tau}^2$ yang diduga melalui REML bisa digunakan walaupun asumsi kenormalan tidak sepenuhnya dipenuhi. Solusi melalui REML adalah dengan memaksimumkan fungsi kemungkinan berikut terhadap τ^2 , (SAS Guide 9.1), yang dilakukan dengan metode iterasi Newton Raphson.

$$\log L(\tau^2, \sigma_i^2) = -\frac{1}{2} \log |V| - \frac{1}{2} \log |X^T V^{-1} X| - \frac{1}{2} r^T V^{-1} r - \frac{n-p}{2} \log(2\pi)$$

dengan $r = Y - X(X^T V^{-1} X)^{-1} X^T V^{-1} Y$.

Dengan mensubstitusi τ^2 oleh $\hat{\tau}^2$ terhadap penduga BLUP ($\hat{\theta}_i^{BLUP}$) pada persamaan (2.3), maka diperoleh suatu bentuk penduga baru

$$\hat{\theta}_i^{EBLUP} = \hat{\theta}_i(y_i | \hat{\beta}, \hat{\tau}^2) = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) x_i^T \hat{\beta} \quad (2.4)$$

yang kemudian dikenal dengan penduga *empirical best linear unbiased predictor* (EBLUP).

Menurut Ghosh dan Rao (1994) $MSE(\hat{\theta}_i^{EBLUP}) = g_{1i}(\tau^2) + g_{2i}(\tau^2)$, dengan

$$g_{1i}(\tau^2) = \frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2}, \text{ dan}$$

$$g_{2i}(\tau^2) = \sigma_i^4 / (\tau^2 + \sigma_i^2) [x_i'(X'V^{-1}X)^{-1} x_i]$$

untuk $i = 1, 2, \dots, m$.

Kuadrat tengah galat dari $\hat{\theta}_i^{EBLUP}$, $MSE(\hat{\theta}_i^{EBLUP}) = E(\hat{\theta}_i^{EBLUP} - \theta_i)^2 = M_i$,

dapat diperoleh dengan menguraikan $E(\hat{\theta}_i^{EBLUP} - \theta_i)^2$ menjadi

$E[(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BP}) + (\hat{\theta}_i^{BP} - \theta_i)]^2$. Dengan menerapkan kaidah umum

$E[(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BP})(\hat{\theta}_i^{BP} - \theta_i)] = 0$, maka diperoleh $M_i = M_{1i} + M_{2i}$ dengan

$M_{1i} = E(\hat{\theta}_i^{BP} - \theta_i)^2$ dan $M_{2i} = E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BP})^2$. Lebih lanjut,

$M_{1i} = g_{1i}(\tau^2) = \gamma_i \sigma_i^2$ dan M_{2i} adalah kontribusi terhadap MSE karena

pendugaan parameter model. Melalui dekomposisi deret Taylor diperoleh $M_{2i} = g_{2i}(\tau^2) + g_{3i}(\tau^2)$, dengan $g_{2i}(\tau^2)$ adalah kontribusi

terhadap MSE akibat pendugaan β dan $g_{3i}(\tau^2)$ adalah kontribusi

terhadap MSE akibat pendugaan τ^2 . Pembahasan lebih detail untuk masalah ini dapat dilihat pada Rao (2003, Bab 7).

Penelitian-penelitian terakhir fokus pada pendugaan MSE dari $\hat{\theta}_i^{EBLUP}$.

Ketika asumsi-asumsi dipenuhi, Prasad dan Rao (1990) memberikan suatu pendekatan penduga yang tak bias bagi M_{1i} dan M_{2i} sebagai

berikut $\hat{M}_{1i} = g_{1i}(\hat{\tau}^2) + g_{3i}(\hat{\tau}^2)$ dan $\hat{M}_{2i} = g_{2i}(\hat{\tau}^2) + g_{3i}(\hat{\tau}^2)$. Dengan demikian penduga bagi M_i adalah

$$\hat{M}_i = \hat{M}_{1i} + \hat{M}_{2i} = g_{1i}(\hat{\tau}^2) + g_{2i}(\hat{\tau}^2) + 2g_{3i}(\hat{\tau}^2) \quad (2.5)$$

dengan $g_{3i}(\hat{\tau}^2) = \frac{2\sigma_i^4}{m^2(\hat{\tau}^2 + \sigma_i^2)^2} \sum_{j=1}^m (\hat{\tau}^2 + \sigma_j^2)^2$.

Menurut Datta dan Lahiri (1997) dalam Rao (1999) pendekatan MSE($\hat{\theta}_i^{EBLUP}$) pada persamaan (2.5) sesuai digunakan untuk pendugaan $\hat{\tau}^2$ dengan REML atau metode momen, tetapi tidak sesuai untuk pendugaan dengan ML. Lahiri dan Rao (1995) menjelaskan bahwa hampiran MSE($\hat{\theta}_i^{EBLUP}$) tersebut bersifat *robust* terhadap ketidaknormalan pengaruh acak area kecil v_i , dalam pengertian bahwa penduga tersebut masih tetap memiliki sifat tidak bias.

Datta dan Lahiri (2000) melakukan kajian penduga MSE menggunakan ML dan REML, sedangkan Datta, Rao dan Smith (2005) menggunakan metode Fay-Herriot. Hasil simulasi dari Datta, Rao dan Smith (2005) mengindikasikan bahwa berdasarkan metode Fay-Herriot, penduga bagi MSE memiliki performa lebih baik untuk mengatasi bias dan koefisien keragaman (*coeficient variance*, CV), khususnya untuk σ_i^2 yang kecil dan keragaman antar area relatif lebih besar. Pfeiffermann dan Glickman (2004) melakukan kajian metode bootstrap parametrik dan non-parametrik untuk menduga M_i dan menunjukkan bahwa penduga MSE bersifat *robust* walaupun v_i dan e_i tidak normal. Hal ini merupakan hasil lanjutan dari Lahiri dan Rao (1995) yang memperlihatkan bahwa penduga MSE menggunakan metode PR untuk menduga τ^2 bersifat *robust* untuk v_i yang tidak normal, sedangkan e_i normal. Pfeiffermann dan Glickman (2004) memperlihatkan pula bahwa seluruh penduga MSE memiliki nilai bias ketika galat contoh, e_i ,

tidak normal, kecuali untuk metode bootstrap parametrik yang dibangkitkan melalui contoh bootstrap dari sebaran yang tepat. Butar dan Lahiri (2003) serta Pfeiffermann dan Glickman (2004) mengajukan metode bootstrap untuk kasus normal. Penduga tersebut menggunakan penduga M_{1i} yang sudah dikoreksi terhadap biasnya dan penduga bootstrap untuk M_{2i} melalui pembangkitan bilangan besar, B , dari peubah acak \tilde{v}_i^b yang mengikuti $N(0, 1)$ dan \tilde{e}_i^b yang mengikuti $N(0, 1)$, $b = 1, 2, \dots, B$. Peubah bootstrap $\{\tilde{v}_i^b\}$ dan $\{\tilde{e}_i^b\}$ digunakan untuk membangkitkan penduga langsung bootstrap $\hat{\theta}_i^b = x_i^T \hat{\beta} + \sqrt{\hat{\tau}^2} \tilde{v}_i^b + \sqrt{\sigma_i^2} \tilde{e}_i^b$. Dengan teknik yang serupa, \tilde{v}_i^b dan \tilde{e}_i^b dapat dibangkitkan juga untuk kasus tidak normal.

Pada kasus pendugaan MSE dari penduga total Y_i , Slud dan Maiti (2006) menyampaikan suatu hampiran penduga tak bias bagi MSE, $MSE(\hat{Y}_i^{SM})$, untuk kasus normal dengan menggunakan dekomposisi deret Taylor yang mirip dengan yang dikerjakan Prasad dan Rao (1990). Metode jackknife dari Wan (1999) serta Jiang, Lahiri dan Wan (2002) mungkin digunakan untuk memperoleh pendekatan penduga $MSE(\hat{Y}_i^{EBLUP}) = M_i$ yang tak bias, walaupun hal ini memerlukan teknik Monte Carlo atau penyelesaian numerik (Rao, 2003). Penguraian M_i seperti sebelumnya menjadi dua komponen M_{1i} dan M_{2i} juga diterapkan dalam pendugaan M_i dengan metode jackknife. Untuk kasus khusus $g(Y_i) = Y_i$, Chen (2001) serta Chen dan Lahiri (2005) mengajukan penduga MSE dengan metode jackknife terboboti dan memiliki performa yang baik untuk kajian simulasi pada berbagai nilai rasio komponen ragam terhadap total keragaman. Metode jackknife terboboti menggunakan pembobot $k_i = 1 - h_{ii}$ yang tergantung pada pembobot nilai leverage $h_{ii} = \left(\sqrt{\sigma_i^2} \right)^{-1} x_i^T \sum_{i=1}^m \left(x_i x_i^T / \sqrt{\sigma_i^2} \right)^{-1} x_i$. Dengan demikian area yang memiliki nilai h_{ii} besar akan menerima

pembobotan k_i yang kecil, berbeda dengan metode jackknife biasa dimana pembobotnya merupakan konstanta $k_i = 1$ untuk semua area. Kedua metode tersebut, jackknife terboboti maupun jackknife biasa memiliki sifat *approximately unbiased* untuk m yang besar (Chen dan Lahiri, 2005).

Berbagai pengembangan model taraf area, $y_i = x_i^T \beta + v_i + e_i$, telah banyak dibahas dalam berbagai literatur meliputi model dengan *correlated sampling errors* e_i , *spatially correlated model errors* v_i , dan *time series* serta *cross-sectional models* (Rao, 2003). Beberapa penelitian yang membahas pengembangan model spasial dalam SAE seperti Petrucci, Pratesi dan Salvati (2005) serta Petrucci dan Salvati (2006). Rao (2003) menyajikan berbagai perkembangan model pendugaan area kecil dengan memanfaatkan data-data dari survei yang dilakukan secara berulang (data deret waktu), walaupun kajian ini masih terbatas pada penggunaan model AR atau random walk. Namun demikian model yang banyak digunakan untuk SAE dengan data deret waktu adalah model yang dikembangkan oleh Rao dan Yu (1994).

2.4. Konsep dan Sifat Statistik Penduga Bayes

Pada sub-bab berikut disajikan ulasan tentang inferensi pada model pendugaan area kecil dengan metode Bayes.

2.4.1. Review Metode Bayes

Perhatikan fungsi kepekatan peluang (fkp) atau fungsi massa peluang (fmp) peubah acak Y dengan syarat θ diketahui, $Y \sim f_Y(y|\theta)$. Jika θ adalah parameter yang diasumsikan sebagai peubah acak tidak diobservasi (*unobservational random variable*) dengan (*prior*) fkp/fmp $\pi(\theta)$, maka fungsi fkp/fmp bersama dari Y dan θ adalah $f(y, \theta) = f_Y(y|\theta) \pi(\theta)$. Untuk kasus Y dan θ kontinu, fkp marginal bagi Y adalah $f(y) = \int_{\theta} f(y, \theta) d\theta$, dan fkp posterior bagi θ dengan syarat $Y = y$ adalah

$$\pi(\theta|y) = \frac{f(y,\theta)}{f(y)}.$$

Misalkan $d(y)$ adalah suatu fungsi keputusan yaitu suatu fungsi yang memetakan semua kemungkinan keputusan berdasarkan contoh acak. $L(\theta, d)$ adalah fungsi kerugian (*lost function*) yang berupa bilangan nyata sedemikian sehingga $L(\theta, d) \geq 0 \forall d$ dan $L(\theta, d) = 0$ hanya untuk d dimana $\hat{\theta} = \theta$. Fungsi resiko (*risk function*) adalah nilai tengah dari kemungkinan kerugian, $R(\theta, d) = E_Y [L(\theta, d)]$ dan Resiko Bayes didefinisikan $R(\theta) = E_0 [R(\theta, d)]$.

Penduga Bayes diperoleh dengan meminimumkan resiko Bayes, $R(\theta)$, sehingga :

$$\begin{aligned} R(\theta) &= E_0 [R(\theta, d)] \\ &= \int R(\theta, d) \pi(\theta) d\theta \\ &= \iint L(\theta, d) f(y|\theta) \pi(\theta) dy d\theta \\ &= \iint L(\theta, d) f(y, \theta) dy d\theta \\ &= \iint L(\theta, d) \pi(\theta|y) f(y) dy d\theta \\ &= \int f(y) \left[\int L(\theta, d) \pi(\theta|y) d\theta \right] dy. \end{aligned}$$

Untuk meminimumkan fungsi di atas sebenarnya cukup dengan meminimumkan fungsi $\int L(\theta, d) \pi(\theta|y) d\theta$.

Jika kita ambil $L(\theta, d) = (\theta - d)^2$ adalah galat kuadrat, maka

$$\begin{aligned} \int L(\theta, d) \pi(\theta|y) d\theta &= \int (\theta - d)^2 \pi(\theta|y) d\theta \\ &= \int \theta^2 \pi(\theta|y) d\theta - 2d \int \theta \pi(\theta|y) d\theta + d^2 \int \pi(\theta|y) d\theta \\ &= \int \theta^2 \pi(\theta|y) d\theta - 2d \int \theta \pi(\theta|y) d\theta + d^2 (1) \\ &\Leftrightarrow \frac{\delta}{\delta d} \int \theta^2 \pi(\theta|y) d\theta - 2d \int \theta \pi(\theta|y) d\theta + d^2 (1) = 0 \\ &\Leftrightarrow 2d - 2 \int \theta \pi(\theta|y) d\theta = 0 \\ &\Leftrightarrow d = \int \theta \pi(\theta|y) d\theta \\ &\Leftrightarrow d = E(\theta|y) \text{ adalah penduga bayes bagi } \theta. \end{aligned}$$

Dengan demikian jika kita menggunakan fungsi kerugian kuadratik

$$L(\theta, d) = (\theta - d)^2$$

dimana d adalah fungsi keputusan, maka penduga bagi θ adalah

$$\hat{\theta}^B = E(\theta|y) = \int_{\theta} \theta \pi(\theta|y) d\theta, \quad (2.6)$$

sedangkan ragam dari $\hat{\theta}^B$, $\text{Var}(\hat{\theta}^B)$, adalah $\text{Var}(\theta|y)$.

2.4.2. Empirical Bayes pada Model SAE

Perhatikan $\hat{\theta}_i^{\text{BP}} = \hat{\theta}_i(y_i|\beta, \tau^2)$ juga adalah penduga Bayes bagi θ_i berdasarkan model Bayes berikut:

- (i) $y_i|\theta_i \sim N(\theta_i, \sigma_i^2)$
- (ii) $\theta_i \sim N(x_i^T \beta, \tau^2)$ adalah sebaran prior bagi θ_i untuk $i = 1, 2, \dots, m$.

Penjabaran model Bayes di atas adalah:

$$f(y_i|\theta_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(y_i - \theta_i)^2\right) \text{ dan}$$

$$\pi(\theta_i|\beta, \tau^2) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}(\theta_i - x_i^T \beta)^2\right), \text{ sehingga}$$

$$f(y, \theta|\beta, \tau^2) = \prod_{i=1}^m \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(y_i - \theta_i)^2\right) \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}(\theta_i - x_i^T \beta)^2\right) \right)$$

untuk $y = (y_1, y_2, \dots, y_m)^T$, $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$.

Kita perhatikan dua bentuk fungsi eksponensial tanpa faktor $(-1/2)$ dari $f(y, \theta|\beta, \tau^2)$,

$$\begin{aligned} & \frac{1}{\sigma_i^2}(y_i - \theta_i)^2 + \frac{1}{\tau^2}(\theta_i - x_i^T \beta)^2 \\ &= \frac{1}{\sigma_i^2}(y_i^2 - 2y_i\theta_i + \theta_i^2) + \frac{1}{\tau^2}(\theta_i^2 - 2\theta_i x_i^T \beta + (x_i^T \beta)^2) \\ &= \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right) \theta_i^2 - 2 \left(\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2} \right) \theta_i + \alpha_i \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right) \left\{ \theta_i^2 - \frac{2\theta_i \left(\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2} \right)}{\left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)} + \left(\frac{\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2}}{\frac{1}{\sigma_i^2} + \frac{1}{\tau^2}} \right)^2 \right\} + \alpha_i^* \\
&= \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right) \left\{ \theta_i - \frac{\left(\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2} \right)}{\left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)} \right\}^2 + \alpha_i^*
\end{aligned}$$

dengan α_i^* adalah suatu konstanta yang bebas dari θ_i . Dengan demikian aproksimasi sebaran posterior bagi θ_i jika y_i , β dan τ^2 masing-masing diketahui adalah

$$\begin{aligned}
(\theta_i | y_i, \beta, \tau^2) &\sim N \left(\left(\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2} \right) \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)^{-1}, \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)^{-1} \right) \\
&\sim N \left(\left(\frac{\tau^2 y_i + \sigma_i^2 x_i^T \beta}{\tau^2 \sigma_i^2} \right) \left(\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} \right), \left(\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} \right) \right) \\
&\sim N \left(\left(\frac{\tau^2 y_i + \sigma_i^2 x_i^T \beta}{\tau^2 + \sigma_i^2} \right), \left(\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} \right) \right) \\
&\sim N \left(\left(\frac{\tau^2}{\tau^2 + \sigma_i^2} y_i + \frac{\sigma_i^2}{\tau^2 + \sigma_i^2} x_i^T \beta \right), \left(\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} \right) \right) \\
&\sim N \left(\left(\gamma_i y_i + (1 - \gamma_i) x_i^T \beta \right), \left(\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} \right) \right)
\end{aligned}$$

dengan $\gamma_i = \tau^2 / (\tau^2 + \sigma_i^2)$.

Jadi dapat diperlihatkan bahwa

$$\begin{aligned}
\hat{\theta}_i^B &= E(\theta_i | y_i, \beta, \tau^2) \\
&= \gamma_i y_i + (1 - \gamma_i) x_i^T \beta ; \text{ dengan } \gamma_i = \tau^2 / (\tau^2 + \sigma_i^2).
\end{aligned}$$

Berdasarkan uraian di atas, dapat diperlihatkan bahwa penduga $\hat{\theta}_i^B$ dan $\hat{\theta}_i^{BP}$ adalah identik untuk kasus normal. Namun kelebihan metode

Bayes dapat diterapkan secara lebih umum untuk berbagai sebaran gabungan dari y_i dan θ_i .

Lebih lanjut, jika τ^2 dan β terlebih dulu diduga dari data, penduga Bayes kemudian dikenal sebagai penduga *empirical* Bayes, $\hat{\theta}_i^{EB}$. Menurut Rao (1999), pendekatan *empirical* Bayes tidak berbeda dengan pendekatan frequentis karena hanya digunakan untuk menghubungkan model percontohan yang kemudian divalidasi oleh data pengamatan. Pada pendekatan ini tidak ada sebaran prior untuk parameter model seperti pada pendekatan hierarki Bayes.

Perhatikan $MSE(\hat{\theta}_i^{EB})$ yang diperoleh dari ragam sebaran posterior untuk kasus normal, yaitu

$$(\theta_i | y_i, \beta, \tau^2) \sim N \left(\left(\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2} \right) \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)^{-1}, \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)^{-1} \right),$$

$$\text{sehingga } MSE(\hat{\theta}_i^{EB}) = \text{Var}(\theta_i | y_i, \beta, A) = \frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} = g_{1i}(\tau^2).$$

Penduga $MSE(\hat{\theta}_i^{EB})$ ini jelas bersifat *underestimate* karena pengaruh pendugaan τ^2 dan β .

Jika τ_0^2 adalah suatu bilangan tidak negatif yang diketahui dan I_E adalah suatu fungsi indikator yaitu bernilai 1 jika kejadian E terjadi dan bernilai 0 untuk kondisi lainnya, maka metode momen terboboti menurut Chen (2001) untuk menduga τ^2 adalah

$$\hat{\tau}^2 = \tau_0^2 I(\tilde{\tau}^2 \leq 0) + \tilde{\tau}^2 I(\tilde{\tau}^2 > 0)$$

dengan : $\tilde{\tau}^2 = \delta - c$

$$\delta = (m - p)^{-1} \sum_{i=1}^m (y_i - x_i^T \hat{\beta}_{OLS})^2$$

$$c = (m - p)^{-1} \sum_{i=1}^m w_i \sigma_i^2$$

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$$

$$w_i = 1 - h_{ii}$$

$$h_{ij} = x_i^T (X^T X)^{-1} x_j, \text{ untuk } i, j = 1, 2, \dots, m.$$

Adapun β biasanya disubstitusi oleh penduganya yang diperoleh berdasarkan metode kuadrat terkecil terboboti

$$\hat{\beta} = \hat{\beta}(\hat{\tau}^2) = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} Y$$

dengan $\hat{V} = \text{Diag}(\hat{\tau}^2 + \sigma_1^2, \hat{\tau}^2 + \sigma_2^2, \dots, \hat{\tau}^2 + \sigma_m^2)$.

Chen (2001) serta Chen dan Lahiri (2005) mengusulkan suatu metode yang diberi nama *weighted jackknife* yang merupakan pengembangan dari teknik jackknife Wan (1999) untuk menduga MSE dari $\hat{\theta}_i(y_i | \hat{\tau}^2)$, yaitu

$$\text{MSE}[\hat{\theta}_i(y_i | \hat{\tau}^2)]^{WJ} = h_{1i}^{WJ} + h_{2i}^{WJ}$$

dengan

$$h_{1i}^{WJ} = H_{1i}(\hat{\tau}^2) - \sum_{u=1}^m w_u (H_{1i}(\hat{\tau}_{-u}^2) - H_{1i}(\hat{\tau}^2))$$

$$h_{2i}^{WJ} = \sum_{u=1}^m w_u (\hat{\theta}_i(y_i | \hat{\tau}_{-u}^2) - \hat{\theta}_i(y_i | \hat{\tau}^2))^2, \text{ dan}$$

$$\hat{\tau}_{-u}^2 = \tau_0^2 I(\tilde{\tau}_{-u}^2 \leq 0) + \tilde{\tau}_{-u}^2 I(\tilde{\tau}_{-u}^2 > 0)$$

$$\tilde{\tau}_{-u}^2 = \delta_{-u} - c$$

$$\delta_{-u} = (m - p - 1)^{-1} \sum_{i \neq u} (y_i - x_i^T \hat{\beta}_{OLS,-u})^2$$

$$\hat{\beta}_{OLS,-u} = (X_{(u)}^T X_{(u)})^{-1} X_{(u)}^T Y_{(u)}$$

$$\hat{\theta}_i(y_i | \hat{\tau}_{-u}^2) = x_i^T \hat{\beta}_{OLS,-u} + (1 - \hat{B}_{i,-u})(y_i - x_i^T \hat{\beta}_{OLS,-u})$$

$$\hat{\beta}_{-u} = \hat{\beta}_{-u}(\hat{\tau}_{-u}^2) = (X_{(u)}^T \hat{V}_{(u)}^{-1} X_{(u)})^{-1} X_{(u)}^T \hat{V}_{(u)}^{-1} Y_{(u)}$$

$$\hat{B}_{i,-u} = \sigma_i^2 / (\hat{\tau}_{-u}^2 + \sigma_i^2)$$

Dalam persamaan di atas $X_{(u)}$ diperoleh dari matriks data X dengan menghapus baris ke- u ($u = 1, 2, \dots, m$), begitu pula dengan $Y_{(u)}$. $\hat{V}_{(u)}$ diperoleh dengan menghapus baris ke- u dan kolom ke- u dari \hat{V} dan substitusikan $\hat{\tau}_{-u}^2$ terhadap $\hat{\tau}^2$ untuk $u = 1, 2, \dots, m$.

2.4.3. Bayes Berhierarkhi pada Model SAE

Dalam pendekatan hierarki Bayes (HB), kita harus terlebih dahulu dapat menetapkan sebaran prior untuk seluruh parameter model, sedangkan sebaran posteriornya kemudian diturunkan berdasarkan sebaran prior sebelumnya. Inferensi terhadap parameter model didasarkan pada sebaran posterior, yaitu parameter akan diperoleh berdasarkan nilai harapan sebaran posteriornya, sedangkan presisinya diukur berdasarkan ragam sebaran posterior yang bersesuaian.

Perhatikan model Fay dan Herriot (1979) yang juga mempertimbangkan $Y_i = \log(Z_i)$ dimana Z_i adalah penduga langsung berdasarkan metode survei. Mereka mengeksplorasi hubungan antara penduga koefisien keragaman (CV) dengan ukuran populasi area kecil. CV untuk Z didefinisikan sebagai $[\text{Var}(Z)]^{1/2}/E(Z)$ dimana $\text{Var}(Z)$ adalah ragam dari Z dan $E(Z)$ adalah nilai harapan Z . CV kemudian diduga dengan s_z/Z , dimana s_z adalah standar deviasi penduga langsung Z . Mereka menemukan bahwa $s_z/Z = 3/\sqrt{N}$ dengan N adalah ukuran populasi area. Dengan demikian, Fay dan Herriot menyatakan bahwa $\text{Var}(Z)$ dapat diduga dengan baik oleh $(9/N)[E(Z)]^2$.

Pengembangan terhadap model Fay-Herriot dimana Y_i adalah fungsi dari suatu penduga langsung Z_i , sehingga $Y_i = g(Z_i)$. Model diasumsikan :

- (i) $Y_i|\theta_i \sim N(\theta_i = g(\mu_i), \sigma_i^2)$ dimana $g(\cdot)$ merupakan suatu fungsi hubung dan σ_i^2 diketahui.
- (ii) Sebaran prior bagi θ_i adalah $N(x_i^T\beta, \tau^2)$,

Untuk $g(\cdot) = \log(\cdot)$, model tersebut dapat ditulis menjadi $Y_i = \log(Z_i) = x_i^T\beta + v_i + e_i$ dengan $v_i \sim N(0, \tau^2)$ dan bersifat saling bebas dengan $e_i \sim N(0, \sigma_i^2)$ untuk $i = 1, 2, \dots, m$. Model di atas tidak lain adalah bentuk khusus *generalized linear mixed model* dengan fungsi hubung logaritma. Masalah yang dihadapi dalam kasus seperti ini

adalah pendugaan terhadap $\mu_i = g^{-1}(\theta_i)$. Jika $g(\cdot)$ adalah fungsi identitas, sehingga $g(\mu_i) = \mu_i = \theta_i$, maka model tersebut adalah sama dengan model pada pembahasan sebelumnya.

Perhatikan untuk kondisi-kondisi berikut:

$$\begin{aligned} y &= (y_1, y_2, \dots, y_m)^T \\ \theta &= (\theta_1, \theta_2, \dots, \theta_m)^T \\ x^T &= (x_1, x_2, \dots, x_m)^T \\ D &= \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2). \end{aligned}$$

Fungsi sebaran bersama untuk (y, θ) jika β dan τ^2 diketahui adalah

$$f(y, \theta | \beta, \tau^2) = \prod_{i=1}^m \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(y_i - \theta_i)^2\right) \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}(\theta_i - x_i^T \beta)^2\right) \right)$$

Untuk kasus ini pada Sub-bab 2.4.2 telah diperlihatkan bahwa

$$\begin{aligned} (\theta_i | y_i, \beta, \tau^2) &\sim N \left(\left(\frac{y_i}{\sigma_i^2} + \frac{x_i^T \beta}{\tau^2} \right) \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)^{-1}, \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2} \right)^{-1} \right) \\ &\sim N \left((\gamma_i y_i + (1 - \gamma_i) x_i^T \beta), \left(\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} \right) \right). \end{aligned}$$

Penduga bagi θ_i adalah $\hat{\theta}_i^B = E(\theta_i | y_i, \beta, \tau^2)$ sehingga penduga bagi μ_i adalah

$$\hat{\mu}_i^B = E[g^{-1}(\theta_i) | y_i, \beta, \tau^2]$$

dengan $\text{Var}(\hat{\mu}_i^B) = \text{Var}\{E[(g^{-1}(\theta_i) | y_i, \beta, \tau^2)]\}$.

Dalam prakteknya baik β maupun τ^2 tidak diketahui. Jika τ^2 diketahui tetapi β tidak diketahui sehingga β diasumsikan mengikuti suatu sebaran tertentu, maka akan diperoleh $f(y_i, \theta_i, \beta | \tau^2)$ dan fungsi kepadatan peluang marginal bersama (y, θ) adalah $f(y_i, \theta_i | \tau^2)$ merupakan solusi dari integral $f(y_i, \theta_i, \beta | \tau^2)$ terhadap β dengan asumsi

β kontinu. Sebaran posterior bagi θ adalah $f(\theta_i|y_i, \tau^2)$ dan penduga bagi θ_i adalah $\hat{\theta}_i^B = E(\theta_i|y_i, \tau^2)$.

Lebih lanjut, dalam prakteknya sangat jarang τ^2 diketahui. Jika diasumsikan $\tau^2 \sim \pi(\tau^2)$, sehingga $f(y_i, \theta_i, \tau^2) = f(y_i, \theta_i | \tau^2) \pi(\tau^2)$. Fungsi marginal bersama (y_i, θ) , setelah diintegrasikan terhadap τ^2 , adalah $f(y_i, \theta_i)$. Sebaran posterior bagi θ adalah $f(\theta_i|y_i)$ dan penduga bagi θ_i adalah $\hat{\theta}_i^B = E(\theta_i|y_i)$.

BAB III

METODE PENDUGAAN LANGSUNG

2.5. Latar Belakang

Survei Sosial Ekonomi Nasional (Susenas) yang dilakukan oleh BPS didisain untuk skala nasional. Apabila dari survei tersebut ingin dilakukan pendugaan untuk sub-populasi yang biasanya memiliki ukuran contoh yang kecil, pendugaan dengan menggunakan penduga langsung (π -estimator) akan memiliki presisi yang rendah. Usaha untuk meningkatkan presisi dalam kasus contoh berukuran kecil dilakukan melalui peningkatan efektifitas ukuran contoh yang dikenal dengan istilah pendugaan area kecil.

Penduga parameter yang efisien untuk suatu area kecil merupakan tujuan penting dalam analisis data survei. Pendekatan klasik untuk menduga parameter area kecil didasarkan pada model disain penarikan contoh. Namun demikian adakalanya kita memiliki informasi tambahan yang dapat digunakan untuk pendugaan parameter. Dalam beberapa kasus kita bisa memperoleh nilai parameter yang menjadi perhatian dari area kecil lain yang memiliki karakteristik serupa, atau nilai pada waktu yang lalu, atau nilai dari peubah yang memiliki hubungan dengan peubah yang sedang diamati.

Dengan merujuk pada data Susenas, metode penarikan contoh yang dilakukan BPS pada level desa/kelurahan adalah dengan terlebih dahulu menentukan blok sensus, baru kemudian menentukan rumah tangga objek survei di dalam blok sensus terpilih. Namun demikian, pada penelitian ini dikaji dua pendekatan, yaitu metode penarikan contoh acak sederhana dan metode penarikan contoh acak gerombol dua tahap, dengan alasan : (1) banyak pengguna data Susenas mengasumsikan disain penarikan contoh acak sederhana dalam melakukan analisis dengan berbagai alasan masing-masing, (2) jika

blok sensus sudah ditentukan atau tidak diacak serta biasanya satu desa/kelurahan hanya terdiri atas satu blok sensus yang disurvei, maka metode penarikan contoh acak sederhana dapat digunakan, dan (3) adapun jika blok sensus diacak, maka disain surveinya adalah penarikan contoh acak gerombol dua tahap.

Secara umum berdasarkan disain survei yang dilakukan BPS, penduga langsung untuk data Susenas memiliki bentuk sebagai berikut

$$\hat{\bar{Y}}_i^{DIRECT} = \frac{1}{\sum_{j,k \in s_i} w_{ijk}} \sum_{j,k \in s_i} w_{ijk} y_{ijk} \quad (3.1)$$

dengan pembobot w_{ijk} merupakan kebalikan dari peluang pengambilan

contoh yaitu $w_{ijk} = \frac{1}{\sum_{\{j,k \in s_i\}} p(s)}$ dengan notasi i merupakan indeks untuk

setiap area kecil, notasi j merupakan indeks untuk setiap blok sensus, notasi k merupakan indeks untuk setiap rumah tangga, dan notasi s adalah indeks untuk unit di dalam contoh.

2.6. GREG pada Penarikan Contoh Acak Sederhana dan Gerombol Dua Tahap

Generalized regression (GREG) merupakan suatu metode pendugaan parameter yang memungkinkan untuk menggunakan beberapa informasi tambahan dan dirancang untuk meningkatkan keakuratan dengan menggunakan informasi tambahan x_i yang memiliki hubungan fungsional dengan peubah yang menjadi perhatian, y_i . Metode ini dapat digunakan untuk menduga total populasi, nilai tengah populasi ataupun proporsi populasi. Metode GREG pada penelitian ini didasarkan atas model linier, yaitu

$$y_i = x_i^T \beta + \varepsilon_i, \text{ dengan } \varepsilon_i \sim N(0, \sigma_i^2). \quad (3.2)$$

Penduga GREG dari model ini adalah

$$\begin{aligned}\hat{Y}_i^{GREG} &= \frac{1}{W_i} \sum_{j,k \in s_i} w_{ijk} y_{ijk} + \left(\bar{X}_i - \frac{1}{W_i} \sum_{j,k \in s_i} w_{ijk} x_{ijk} \right)^T \hat{\beta} \\ &= \hat{Y}_i^{DIRECT} + \left(\bar{X}_i - \hat{\bar{X}}_i \right)^T \hat{\beta}\end{aligned}$$

dengan :

- $\bar{X}_i = (\bar{X}_{i,1}, \dots, \bar{X}_{i,p})^T$ adalah vektor nilai tengah p peubah pendukung
- $W_i = \sum_{j,k \in s_i} w_{ijk}$
- $w_{ijk} = 1/\pi_{ijk}$, dengan $\pi_{ijk} = \sum_{j,k \in s_i} p(s)$
- $\hat{\bar{X}}_i = \frac{1}{W_i} \sum_{j,k \in s_i} w_{ijk} x_{ijk} = \hat{Y}_i(x)$
- $\hat{Y}_i^{DIRECT} = \frac{1}{W_i} \sum_{j,k \in s_i} w_{ijk} y_{ijk} = \hat{Y}_i(y)$.

Penduga langsung (\hat{Y}_i^{DIRECT}) dan besarnya pembobot (w_{ijk}) diperoleh berdasarkan teknik penarikan contoh yang digunakan dalam pelaksanaan survei. Penduga koefisien regresi β dapat diperoleh dengan menggunakan beberapa metode seperti metode kuadrat terkecil dan secara umum bentuknya adalah sebagai berikut :

$$\hat{\beta} = \left(\sum_{j,k \in s_i} w_{ijk} x_{ijk} x_{ijk}^T \right)^{-1} \sum_{j,k \in s_i} w_{ijk} x_{ijk} y_{ijk} . \quad (3.3)$$

Formulasi pendugaan langsung dan GREG yang diturunkan dari penarikan contoh acak sederhana (*simple random sampling, SRS*) dan penarikan contoh gerombol dua tahap (*two stage cluster sampling, TSCS*) disajikan sebagai berikut :

Tabel 3.1. Formula penduga langsung berdasarkan metode SRS dan TSCS

Penduga Langsung		
	SRS	TSCS
Pembobot	$: w_{ik} = \frac{1}{\pi_{ik}} = \frac{M_{i.}}{m_{i.}}$	$w_{ijk} = \frac{1}{\pi_{ijk}} = \frac{N_i}{n_i} \frac{M_{ij}}{m_{ij}}$
Penduga nilai tengah	$: \hat{\bar{Y}}_i^{DIRECT} = \bar{y}_i = \frac{1}{\sum_{k=1}^{m_{i.}} w_{ik}} \sum_{k=1}^{m_{i.}} w_{ik} y_{i.k}$	$\hat{\bar{Y}}_i^{DIRECT} = \bar{\bar{y}}_i = \frac{1}{\sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ijk}} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ijk} y_{ijk}$
Penduga ragam	$: V(\hat{\bar{Y}}_i^{DIRECT}) = \frac{S_i^2}{m_{i.}} \left(1 - \frac{m_{i.}}{M_{i.}}\right)$	$V(\hat{\bar{Y}}_i^{DIRECT}) = \left(1 - \frac{n_i}{N_i}\right) \frac{N_i^2}{n_i M_{i.}^2} S_c^2 + \frac{N_i}{n_i M_{i.}^2} \sum_{j=1}^{n_i} M_{ij}^2 \left(\frac{M_{ij} - m_{ij}}{M_{ij}}\right)$
Keterangan : $s_i^2 = \sum_{k=1}^{m_{i.}} \frac{(y_{i.k} - \bar{y}_i)^2}{m_{i.} - 1}$, $s_c^2 = \sum_{j=1}^{n_i} \frac{(\bar{y}_{ij} - \bar{\bar{y}}_i)^2}{n_i - 1}$, $s_{ij}^2 = \frac{\sum_{k=1}^{m_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{(m_{ij} - 1)}$, $\hat{e}_{ijk} = y_{ijk} - \hat{\beta} x_{ijk}$		

Tabel 3.2. Formula penduga GREG berdasarkan metode SRS

Penduga GREG untuk SRS	
Pembobot	$: w_{ik} = \frac{1}{\pi_{ik}} = \frac{M_{i.}}{m_{i.}}$
Penduga nilai tengah	$: \hat{\bar{Y}}_i^{GREG} = \hat{\bar{Y}}_i^{DIRECT} + \left(\bar{X}_i - \hat{\bar{X}}_i\right)^T \hat{\beta}$
Penduga ragam	$: V(\hat{\bar{Y}}_i^{GREG}) = \frac{1}{m_{i.}} \left(1 - \frac{m_{i.}}{M_{i.}}\right) \sum_{k=1}^{m_{i.}} \frac{(\hat{e}_{i.k} - \bar{\bar{e}}_i)^2}{m_{i.} - 1}$

Tabel 3.3. Formula penduga GREG berdasarkan metode TSCS

Penduga GREG untuk TSCS	
Pembobot	$: w_{ijk} = \frac{1}{\pi_{ijk}} = \frac{N_i}{n_i} \frac{M_{ij}}{m_{ij}}$
Penduga nilai tengah	$: \hat{\bar{Y}}_i^{GREG} = \hat{\bar{Y}}_i^{DIRECT} + \left(\bar{X}_i - \hat{\bar{X}}_i\right)^T \hat{\beta}$
Penduga ragam	$: V(\hat{\bar{Y}}_i^{GREG}) = \frac{N_i^2}{n_i M_{i.}^2} \left(1 - \frac{n_i}{N_i}\right) \sum_{j=1}^{n_i} \frac{(\bar{\bar{e}}_{ij} - \bar{\bar{e}}_i)^2}{n_i - 1} + \frac{N_i}{n_i M_{i.}^2} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \frac{M_{ij}^2}{m_{ij}} \left(\frac{M_{ij} - m_{ij}}{M_{ij}}\right) \frac{(\hat{e}_{ijk} - \bar{\bar{e}}_{ij})^2}{m_{ij} - 1}$

2.7. Penerapan pada Data Susenas

Sumber data yang digunakan pada kajian ini adalah Susenas 2005 dengan materi informasi berbasis rumah tangga, serta Podes 2005 sebagai sumber data populasi. Peubah yang diamati dan menjadi perhatian adalah pengeluaran per kapita (y_i) pada beberapa kelurahan di Kota Bogor. Peubah penyerta yang digunakan adalah luas lantai (m^2). Sedangkan untuk data populasi, peubah yang digunakan adalah luas lahan untuk pemukiman (m^2).

Kajian yang dilakukan didekati dengan dua cara yaitu SRS dan TSCS. Dua pendekatan ini digunakan karena dalam praktek banyak pengguna statistik yang sering kali mengabaikan teknik TSCS yang sebenarnya digunakan dalam survei seperti pada Susenas, sehingga dianggap bahwa contoh diambil dengan SRS dari suatu area tertentu.

Evaluasi hasil kajian menggunakan *Relative Root Mean Square Error* (RRMSE) yang diperoleh dengan perhitungan sebagai berikut

$$RRMSE(\hat{\theta}_i) = \frac{\sqrt{MSE(\hat{\theta}_i)}}{\hat{\theta}_i} \times 100\%.$$

Peubah penyerta dipilih berdasarkan eksplorasi dan disesuaikan dengan ketersediaan data populasi pada data Podes 2005. Pada penelitian ini hanya digunakan satu peubah penyerta, yaitu x_i (luas lantai) karena tidak tersedianya data populasi yang sesuai untuk peubah-peubah lain yang sebelumnya dikaji dari data Susenas 2005.

Data populasi untuk peubah x_i adalah luas lahan untuk pemukiman (X_i) yang merupakan selisih dari lahan untuk non pertanian dengan koreksi dari beberapa data yang berkaitan dengan bangunan, kemudian dibagi dengan jumlah keluarga atau bisa ditulis sebagai berikut : $X_i = \frac{(a_i - b_i)}{c_i}$

dengan :

a_i = lahan untuk non pertanian pada desa ke-i

b_i = faktor koreksi yang terdiri dari bangunan untuk pendidikan, sarana kesehatan, tempat ibadah, bangunan untuk sarana angkutan, komunikasi dan informasi, bangunan untuk kepentingan ekonomi seperti perusahaan, restoran, hotel, bank, bengkel dan lain-lain.

c_i = jumlah keluarga pada desa ke-i

X_i = luas lahan pemukiman terkoreksi pada desa ke-i

Dengan demikian, peubah yang digunakan pada penelitian ini adalah pengeluaran per kapita (y_i), luas lantai (x_i) dan luas lahan pemukiman terkoreksi (X_i).

Pendugaan langsung rata-rata pengeluaran per kapita dilakukan dengan dasar dua teknik penarikan contoh SRS dan TSCS. Hasil yang diperoleh dari pendugaan langsung dengan kedua metode penarikan contoh dapat dilihat pada Tabel 3.4. Pada penelitian ini diamati 36 desa/kelurahan dengan banyaknya contoh yang diambil pada masing-masing kelurahan sebanyak 16 rumah tangga, kecuali untuk kelurahan Kedung Halang (15 rumah tangga) dan kelurahan Kedung Badak (32 rumah tangga).

Hasil pada Tabel 3.4 menunjukkan bahwa pendugaan dengan menggunakan teknik SRS menghasilkan nilai pendugaan yang sama dengan pendugaan menggunakan teknik TSCS. Hal ini dikarenakan :

1. Untuk setiap desa banyaknya gerombol (blok sensus) yang diambil sebanyak satu gerombol saja, kecuali pada kelurahan Kedung Badak (2 gerombol).
2. Total rumah tangga dalam setiap blok sensus (M_{ij}) tidak diketahui pada setiap desa sehingga besarnya M_{ij} diduga dengan besaran yang merupakan rasio antara total rumah tangga dalam suatu desa

ke-i dengan jumlah blok sensus dalam desa ke-i atau bisa ditulis dengan $\bar{M} = \frac{M_{i.}}{N_i}$.

Dengan demikian, besarnya bobot (w) untuk teknik TSCS akan sama dengan bobot pada teknik SRS. Hal ini dapat dilihat pada pembuktian di bawah ini :

$$w_{TSCS} = \frac{1}{\pi_{jk}} = \frac{N_i}{n_i} \frac{M_{ij}}{m_{ij}} \text{ karena } M_{ij} \text{ tidak diketahui maka diduga dengan } \bar{M}$$

menjadi

$$\begin{aligned} w_{TSCS} &= \frac{N_i}{n_i} \frac{\bar{M}}{m_{ij}} \\ &= \frac{N_i}{n_i} \frac{M_{i.}}{m_{ij} N_i} \\ &= \frac{1}{n_i} \frac{M_{i.}}{m_{ij}}, \quad n_i = 1 \\ &= \frac{M_{i.}}{m_{ij}} \\ &= w_{SRS} \end{aligned}$$

Koefisien regresi β yang digunakan dalam pendugaan GREG diperoleh dengan menggunakan metode kuadrat terkecil. Pendugaan β dengan teknik SRS, penggunaan bobot dapat diabaikan karena masing-masing individu dalam suatu desa memiliki bobot yang sama dan diperoleh berdasarkan model regresi sederhana tanpa intersep. Nilai dugaan β yang diperoleh dengan teknik SRS memiliki nilai yang sama dengan menggunakan teknik TSCS. Hal ini dikarenakan pada teknik TSCS bobot yang diberikan kepada setiap rumah tangga sama seperti yang telah disebutkan pada bagian sebelumnya. Berikut adalah pembuktian bahwa dugaan β pada teknik TSCS sama dengan dugaan β pada teknik SRS :

$$\hat{\beta}_{TSCS} = \left(\sum_{j,k \in s_i} w_{ijk} x_{ijk} x_{ijk}^T \right)^{-1} \sum_{j,k \in s_i} w_{ijk} x_{ijk} y_{ijk} \cdot$$

Karena bobot setiap rumah tangga sama maka $w_{ijk} = a$ (konstanta) sehingga diperoleh :

$$\begin{aligned}
\hat{\beta}_{TSCS} &= \left(\sum_{j,k \in s_i} a x_{ijk} x_{ijk}^T \right)^{-1} \sum_{j,k \in s_i} a x_{ijk} y_{ijk} \\
&= \left(a \sum_{j,k \in s_i} x_{ijk} x_{ijk}^T \right)^{-1} \left(a \sum_{j,k \in s_i} x_{ijk} y_{ijk} \right) \\
&= a^{-1} a \left(\sum_{j,k \in s_i} x_{ijk} x_{ijk}^T \right)^{-1} \left(\sum_{j,k \in s_i} x_{ijk} y_{ijk} \right) \\
&= \left(\sum_{j,k \in s_i} x_{ijk} x_{ijk}^T \right)^{-1} \left(\sum_{j,k \in s_i} x_{ijk} y_{ijk} \right) \\
&= \hat{\beta}_{SRS}
\end{aligned}$$

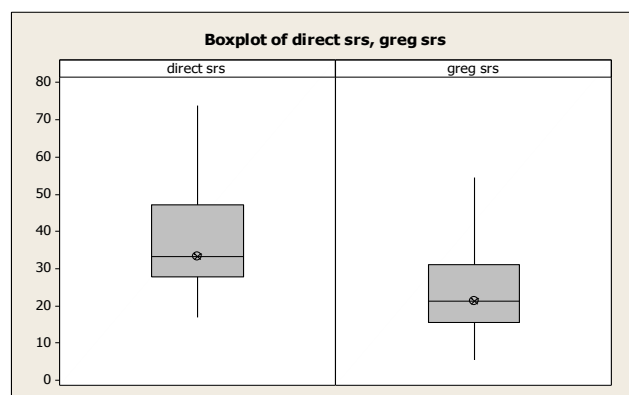
Kasus di atas berlaku juga jika diasumsikan bobot setiap rumah tangga pada setiap gerombol sama dan setiap gerombol memiliki ukuran yang sama. Dengan demikian, hasil kajian menunjukkan bahwa pendugaan dengan teknik SRS menghasilkan nilai pendugaan yang sama dengan teknik TSCS. Hal ini karena pengaruh dari pembobot pada nilai dugaan β dapat dihilangkan seperti yang telah dijelaskan sebelumnya pada bagian pendugaan β . Lebih lanjut, nilai dugaan β untuk teknik SRS dan TSCS adalah sama dan mengakibatkan hasil pendugaan rata-rata pengeluaran per kapita setiap teknik penarikan contoh juga menghasilkan nilai yang sama.

Namun demikian, berdasarkan nilai RRMSE, pendugaan GREG dengan teknik TSCS menghasilkan dugaan yang memiliki presisi lebih baik dibandingkan pendugaan dengan teknik SRS. RRMSE penduga pengeluaran per kapita dengan menggunakan metode GREG dan penduga langsung untuk masing-masing metode penarikan contoh disajikan pada Tabel 3.4.

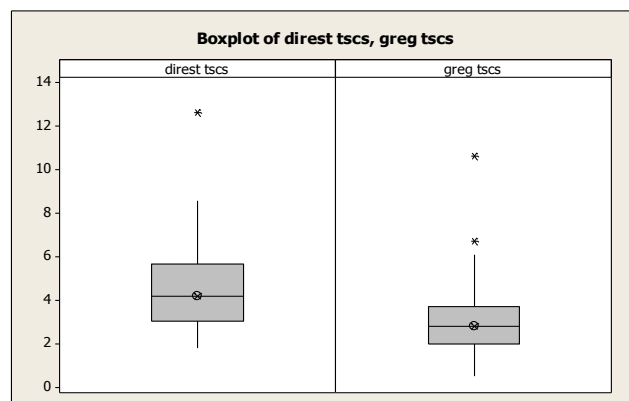
Tabel 3.4. RRMSE penduga langsung dan GREG untuk pendugaan pengeluaran per kapita berdasarkan metode SRS dan TSCS di Kota Bogor berdasarkan data Susenas tahun 2005

Nama Desa/Kelurahan	RRMSE SRS (%)		RRMSE TSCS (%)	
	Penduga Langsung	Penduga GREG	Penduga Langsung	Penduga GREG
Pamoyanan	21.5574	8.0066	2.9874	1.1095
Genteng	27.5936	10.4679	5.0593	1.9193
Harjasari	18.1412	11.6805	2.6878	1.7306
Cipaku	36.7955	27.2583	4.4511	3.2974
Batutulis	59.6609	46.7982	8.5743	6.7257
Empang	39.4423	33.6350	2.7502	2.3453
Cikaret	32.4812	25.6825	3.7192	2.9407
Sindangrasa	17.0725	11.2922	1.9336	1.2789
Katulampa	23.7905	5.4860	2.2689	0.5232
Baranangsiang	48.3173	29.4597	4.8392	2.9505
Sukasari	24.1097	19.1093	3.4116	2.7040
Bantarjati	48.3053	31.5539	4.9695	3.2462
Tegal Gundil	73.9721	50.7916	6.9470	4.7701
Tanahbaru	30.5676	18.3027	3.4144	2.0444
Cimahpar	35.3752	15.4625	4.5569	1.9918
Cibuluh	28.8154	21.1941	4.0118	2.9507
Kedunghalang	65.7174	54.4366	7.4156	6.0952
Ciparigi	53.3672	31.4717	6.1549	3.6297
Babakanpasar	28.3430	22.0152	4.0947	3.1805
Tegallega	47.6329	32.4829	5.9495	4.0572
Pabaton	64.4803	54.2820	12.6393	10.6402
Kebonkelapa	31.3071	29.1277	4.0220	3.7420
Pasirmulya	20.2177	11.5342	4.0258	2.2967
Pasirjaya	31.1810	12.4085	3.5797	1.4245
Gunungbatu	16.8672	12.7975	1.8086	1.3722
Menteng	30.1304	15.4458	2.8261	1.4488
Cilendek Barat	29.6103	24.1059	2.9435	2.3963
Sindangbarang	43.6845	18.3000	5.5024	2.3050
Situgede	34.9995	17.6370	5.1634	2.6019
Semplak	23.4630	20.6920	4.3153	3.8057
Kedungwaringin	46.3344	33.6821	5.3521	3.8906
Kedungjaya	40.4861	30.0680	6.2173	4.6174
Kebonpedes	29.7501	27.0217	3.0138	2.7374
Kedungbadak	34.3495	17.7109	3.1487	2.9412
Kayumanis	42.5307	21.5967	5.7196	2.9043
Kencana	48.6949	17.4046	7.1071	2.5402

Secara umum kajian empirik yang dilakukan juga menunjukkan bahwa hasil pendugaan GREG memiliki nilai RRMSE yang lebih kecil dibandingkan dengan hasil yang diperoleh dari pendugaan langsung dengan teknik SRS. Hal ini berarti pendugaan GREG dengan teknik SRS cukup baik digunakan jika dibandingkan dengan pendugaan langsung dengan teknik SRS. Kajian juga menunjukkan bahwa hasil pendugaan GREG memiliki nilai RRMSE yang lebih kecil dibandingkan dengan hasil yang diperoleh dari pendugaan langsung dengan teknik TSCS.



Gambar 3.1. Boxplot nilai RRMSE berdasarkan metode SRS



Gambar 3.2. Boxplot nilai RRMSE berdasarkan metode TSCS

Jika melihat nilai RRMSE pada pendugaan langsung, terdapat perubahan yang cukup nyata antara pendugaan dengan teknik SRS dibandingkan dengan teknik TSCS. Hal ini juga bisa dilihat dari

boxplot yang menggambarkan perbedaan yang cukup nyata jika menggunakan teknik TSCS.

2.8. Kajian Simulasi Regresi Kekar pada GREG

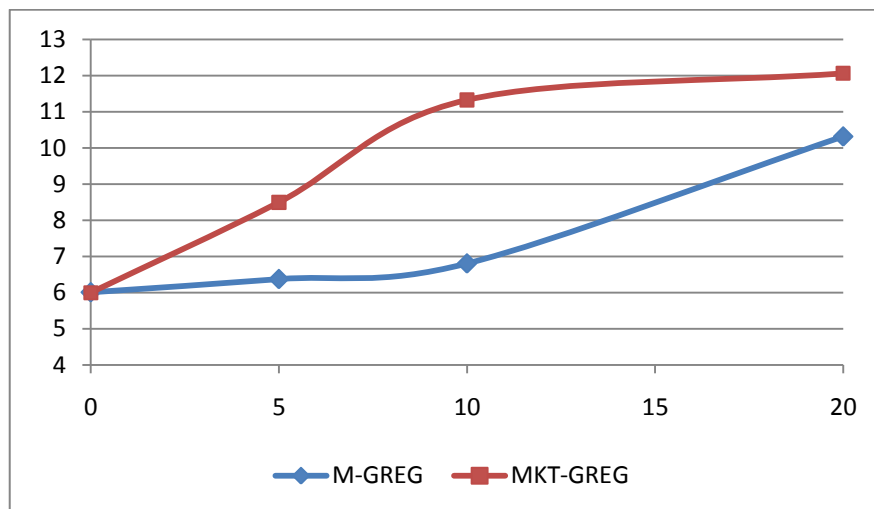
Karakteristik metode GREG adalah menggunakan metode kuadrat terkecil (MKT) dalam menduga koefisien regresi β . Seperti sudah dipahami bersama, MKT sangat tergantung pada asumsi yang seringkali tidak dipenuhi dalam prakteknya dimana data diasumsikan mengikuti sebaran normal. Ketika terdapat pencilan dalam data, MKT seringkali memiliki performa yang rendah. Regresi kekar diperlukan untuk memberikan metode alternatif yang sama baiknya dengan metode kuadrat terkecil, tetapi tidak terlalu dipengaruhi oleh pencilan atau hal lain dalam asumsi model.

Dalam kajian simulasi ini, GREG dipadukan dengan regresi kekar penduga Huber M (*M-estimator*). *M-estimator* merupakan penduga yang meminimumkan fungsi tujuan dalam data. Metode ini banyak digunakan dalam praktek dibandingkan metode penduga kekar lainnya.

Kajian simulasi dilakukan dengan proporsi pencilan yang berbeda-beda, yaitu pada proporsi 0% (tanpa pencilan), 5%, 10% dan 20%. Pemilihan proporsi ini bersifat subjektif dengan alasan bahwa pencilan 10% dari banyaknya data sudah termasuk banyak. Angka 20% ditambahkan dalam penelitian ini untuk memperbesar rentang proporsi pencilan yang diamati. Sedangkan pembentukan data pencilan dilakukan pada data-data ekstrim dalam suatu kumpulan data yang dibangkitkan dan kemudian ditambah atau dikurangi dengan 3 kali standar deviasi dari data Y_{ij} .

Simulasi tanpa ada data pencilan menunjukkan bahwa MKT-GREG (metode GREG yang didasari dengan metode kuadrat terkecil dalam menduga parameter regresinya) dan M-GREG (metode GREG yang

didasari dengan metode regresi kekar *M-estimator* dalam menduga parameter regresinya) menghasilkan RRMSE yang relatif sama. Perbedaan nilai RRMSE semakin besar dimana RRMSE MKT-GREG lebih besar dari RRMSE M-GREG sejalan dengan meningkatnya proporsi pencilan dalam data. Namun demikian, pada kondisi proporsi pencilan 20%, nilai RRMSE dari kedua metode kembali berdekatan dengan nilai RRMSE lebih dari 10%. Nilai RRMSE pada kondisi proporsi pencilan 20% merupakan nilai terbesar yang dicapai oleh kedua metode dari kombinasi proporsi pencilan yang dicobakan. Perilaku RRMSE dari MKT-GREG dan M-GREG tersebut untuk berbagai kombinasi proporsi pencilan disajikan pada Gambar 3.3.



Gambar 3.3. RRMSE MKT-GREG vs M-GREG
berdasarkan proporsi pencilan

2.9. Kesimpulan

Pendugaan parameter nilai tengah berdasarkan GREG dengan memperhitungkan metode penarikan contoh yang benar dan dilakukan di lapangan akan menghasilkan pendugaan dengan presisi yang lebih baik. Dalam hal pendugaan pengeluaran per kapita di Kota Bogor, mengabaikan teknik penarikan contoh gerombol dua tahap tidak merubah nilai pendugaan karena gerombol (blok sensus) yang dipilih dalam setiap desa hanya satu. Selain itu, tidak diketahuinya jumlah

rumah tangga dalam setiap blok sensus juga mengakibatkan penggunaan nilai harapannya sehingga penduga dari parameter regresi akan sama. Namun demikian, dengan tidak mengabaikan teknik penarikan contoh gerombol dua tahap dalam analisis, presisi dari penduga parameter jauh lebih baik.

Penggunaan regresi kekar yang dikombinasikan dengan GREG baik digunakan untuk mereduksi pengaruh pencilan pada data dalam menduga parameter regresi. Untuk pencilan yang kecil, sampai dengan 10%, M-GREG memberikan performa yang baik. Namun untuk pencilan yang lebih besar, baik M-GREG maupun MKT-GREG relatif memiliki performa yang sama dengan RRMSE yang besar. Perlu dipertimbangkan metode lain atau mungkin mengajukan anggapan ada dua populasi pada data jika ditemukan komposisi pencilan sampai 20% atau lebih dari data.

BAB IV

METODE PENDUGAAN TIDAK LANGSUNG

4.1. Latar Belakang

Analisis regresi merupakan suatu teknik statistik yang luas pemakaiannya. Teknik ini memiliki sifat pendugaan yang sangat baik (*powerful tool*) jika asumsi-asumsi yang melandasinya terpenuhi, termasuk di dalamnya adalah hubungan antara peubah respon dengan peubah penjelas dapat digambarkan dengan suatu fungsi tertentu yang terdefinisi seperti pola garis lurus, berbentuk polinomial, atau berpola eksponensial. Di dalam banyak aplikasi, bagaimanapun untuk memperoleh fungsi-fungsi tersebut secara tepat sangat sulit bahkan banyak gejala menunjukkan data-data yang diperoleh tidak menunjukkan suatu pola hubungan yang mudah untuk digambarkan.

Dalam pendugaan area kecil, khususnya pendugaan berbasis model, konsep model linier menjadi inti dari analisis yang dilakukan. Keterpenuhan asumsi linier tentu menjadi syarat penting dalam melakukan analisis lebih lanjut.

4.2. Pendekatan Generalized Additive Mixed Model

Untuk mengatasi kesulitan-kesulitan pada model linier seperti disampaikan sebelumnya, Stone (1985) mengajukan penggunaan model aditif. Model ini menduga pendekatan secara aditif dari fungsi regresi multivariate. Keuntungan penggunaan pendekatan ini paling tidak ada dua hal. Pertama, karena setiap suku aditif diduga secara individu menggunakan pemulus univariate, maka tidak terjadi masalah "*curse of dimensionality*". Kedua, pendugaan setiap suku secara individual dapat menjelaskan bagaimana perubahan peubah respon terhadap perubahan peubah penjelas.

Untuk memperluas penggunaan model aditif dalam berbagai keluarga sebaran, Hastie dan Tibshirani (1990) mengusulkan model aditif terampat (*generalized additive model*, **GAM**). Model ini menghubungkan nilai harapan peubah respon dengan prediktor aditif melalui fungsi hubung yang tidak linier. Model ini memungkinkan sebaran dari peubah respon berasal dari keluarga sebaran eksponensial.

Rao (2003) menyajikan secara intensif ulasan berbagai teknik dalam pendugaan area kecil yang sering digunakan oleh peneliti maupun pemakai statistika, termasuk didalamnya teknik atau **pendekatan *synthetic, composite estimator, empirical best unbiased linear predictors, empirical Bayes and hierarchical Bayes***. Seluruh metode-metode tersebut menggunakan **pendekatan parametrik**.

Pendekatan GAM memiliki keuntungan yang lebih dibandingkan dengan pendekatan parametrik khususnya dalam memodelkan pola hubungan peubah respon dengan peubah penjelas (*auxiliary variable*). Kelebihan tersebut yang selanjutnya digunakan untuk pemodelan yang dilakukan dalam pendugaan area kecil.

Dengan berlandaskan pada model Fay-Herriot (1979) pada model taraf area:

$$y_i = x_i^T \beta + v_i + e_i, \quad i = 1, 2, \dots, m$$

dengan β adalah koefisien regresi, v_i adalah pengaruh acak area, dan e_i adalah galat contoh. Dalam model ini juga diasumsikan bahwa $e_i \sim (0, \sigma_i^2)$, $v_i \sim (0, \tau^2)$ dan keduanya bersifat saling bebas serta biasanya σ_i^2 diasumsikan diketahui.

Kita asumsikan bahwa y_i dan x_i memiliki suatu pola hubungan yang dapat didekati oleh suatu fungsi pemulus $m(\cdot)$. Untuk x_i sebagai peubah penjelas, maka

$$y_i = m(x_i) + v_i + e_i, i = 1, 2, \dots, m \quad (4.1)$$

dengan $v_i | x_i \sim (0, \psi)$, $e_i \sim (0, \sigma_i^2)$, serta e_i dengan v_i saling bebas. Fungsi nilai tengah area kecil dapat dituliskan sebagai berikut:

$$\theta_i(x_i) = m(x_i) + v_i$$

yang merupakan kombinasi linier dari nilai tengah $m(x_i)$ dan pengaruh acak v_i . Kita dapat menggunakan suatu teknik pendugaan untuk mendapatkan fungsi pemulus seperti menggunakan fungsi pemulus linier meliputi pemulus spline, regresi spline, dan *local polynomial regression*. Lebih lanjut pembahasan mendetail metode-metode tersebut dapat dilihat pada Hastie dan Tibshirani (1990).

Jika digunakan fungsi pemulus kernel untuk menduga $m(x_i)$, penduga terbaik (*best predictor*) bagi nilai tengah area kecil θ_i dapat dituliskan sebagai berikut

$$E(\theta_i | y_i) = \gamma_i y_i + (1 - \gamma_i) \hat{m}_h(x_i) \quad (4.2)$$

dengan $\gamma_i = \psi / (\psi + \sigma_i^2)$. Hampiran MSE bagi penduga parameter tersebut dapat dilakukan dengan mengadopsi pendekatan yang diberikan Prasad dan Rao (1990) dengan mensubstitusi $x_i^T \beta$ dalam model linier campuran dengan $\hat{m}_h(x_i)$, sehingga diperoleh formulasi sebagai berikut :

$$\text{mse}(\hat{\theta}_i) = \frac{D_i \hat{\sigma}_u^2}{D_i + \hat{\sigma}_u^2} + (1 - \hat{\gamma})^2 \text{mse}(\hat{m}_h(x_i)) + 2D_i^2 (\hat{\sigma}_u^2 + D_i)^{-3} \text{mse}(\hat{\sigma}_u^2). \quad (4.3)$$

4.3. Pendekatan Nonparametrik P-Spline

Pendekatan nonparametrik P-spline dalam pendugaan area kecil terutama diarahkan dalam menangani pengaruh ketidakcocokan model yang mungkin terjadi. Beberapa kajian yang terkait dengan pengembangan model nonparametrik P-spline dalam pendugaan area kecil telah dilakukan oleh Zheng dan Little (2004), Opsomer, et.al (2008) serta Kurnia, Notodiputro dan Ibrahim (2007). Hasil kajian

awal mereka menunjukkan bahwa pendekatan yang dilakukan mampu memberikan perbaikan hasil jika dibandingkan dengan pendekatan standar EBLUP maupun EB. Pendekatan P-spline juga cukup memberikan perbaikan yang memuaskan jika data mengandung pengaruh spasial seperti yang dilaporkan oleh Opsomer, et.al (2008) maupun pengaruh acak gerombol dalam penarikan contoh acak gerombol (Zheng dan Little, 2004).

Pendekatan P-spline yang digunakan dalam penelitian ini, dengan mengikuti metodologi yang dilakukan oleh Opsomer et.al (2008) serta Ruppert, Wand dan Carroll (2003) dapat dideskripsikan sebagai berikut.

Perhatikan model sederhana

$$y_i = m_o(x_i) + \varepsilon_i$$

dengan $\varepsilon_i \sim (0, \sigma_\varepsilon^2)$. Jika $m_o(\cdot)$ diduga dengan metode p-splines, diperoleh

$$m_o(x; \beta, \gamma) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_k \gamma_k (x - \kappa_k)_+^p$$

dengan p adalah derajat spline, $(x)_+^p$ adalah suatu fungsi dimana $x^p \mathbf{I}_{(x>0)}$, $\kappa_1 < \dots < \kappa_K$ adalah himpunan dari *fixed knot* dan $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_K)^T$ adalah koefisien parametrik dan komponen spline dari model. Jika k cukup besar, maka $m(x; \beta, \gamma)$ akan memiliki ukuran yang sangat besar sehingga model $m_o(x; \beta, \gamma)$ berpotensi *over-parameterized*. Dengan demikian biasanya digunakan suatu penalti terhadap parameter spline γ .

Untuk suatu gugus data $\{(x_i, y_i): i = 1, 2, \dots, n\}$, solusi bagi β dan γ bisa diperoleh dengan meminimumkan bentuk

$$\sum_i (y_i - m(x_i; \beta, \gamma))^2 + \lambda_\gamma \gamma^T \gamma$$

terhadap β dan γ dengan λ_γ adalah *fixed penalty parameter*. Ruppert, Wand dan Carroll (2003) memperlakukan λ sebagai pengaruh acak dalam konteks model linier campuran. Hal ini dilakukan karena penetapan nilai λ tertentu akan sangat berpengaruh dalam pendugaan model dan λ dengan sendirinya akan memiliki kemungkinan nilai yang banyak.

Karena P-spline maupun model SAE dapat dipandang sebagai model dengan pengaruh acak, sangat memungkinkan untuk menggabungkan kedua konsep tersebut dalam *nonparametric small area estimation framework* yang berdasarkan pada model linier campuran.

Kita perhatikan m small area, U_1, \dots, U_m , yang menjadi perhatian. Kemudian kita definisikan $d_{it} = I_{(i \in U_m)}$, dan untuk setiap observasi ke- i , $\mathbf{d}_i = (d_{i1}, \dots, d_{im})^T$. Misalkan pula $\mathbf{Y} = (y_1, \dots, y_n)^T$, \mathbf{X} dan \mathbf{Z} adalah matriks disain sebagai berikut

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & \dots & x_1^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \dots & x_n^p \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} (x_1 - \kappa_1)^{p_+} & \dots & (x_1 - \kappa_k)^{p_+} \\ \vdots & \vdots & \vdots \\ (x_n - \kappa_1)^{p_+} & \dots & (x_n - \kappa_k)^{p_+} \end{pmatrix}$$

dan $\mathbf{D} = (\mathbf{d}_1^T, \dots, \mathbf{d}_m^T)^T$. Dengan demikian, hubungan antar komponen tersebut dalam notasi matriks dapat dituliskan sebagai berikut

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{D}\mathbf{v} + \varepsilon$$

dengan

$$\gamma \sim (0, \Sigma_\gamma) ; \Sigma_\gamma \cong \sigma_\gamma^2 \mathbf{I}_K$$

$$\mathbf{v} \sim (0, \Sigma_v) ; \Sigma_v \cong \sigma_v^2 \mathbf{I}_m$$

$$\varepsilon \sim (0, \Sigma_\varepsilon) ; \Sigma_\varepsilon \cong \sigma_\varepsilon^2 \mathbf{I}_n$$

dan masing-masing komponen bersifat saling bebas.

Model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{D}\mathbf{v} + \varepsilon$ mencakup komponen fungsi spline yang sebagai representasi nilai tengah fungsi nonparametrik $\mathbf{X}\beta + \mathbf{Z}\gamma$, dan komponen pengaruh acak area kecil $\mathbf{D}\mathbf{v}$. Berdasarkan model ini

dengan tetap mengasumsikan Z_γ sebagai faktor pengaruh acak, maka diperoleh

$$\text{Var}(Y) = V = Z \Sigma_\gamma Z^T + D \Sigma_u D^T + \Sigma_\varepsilon.$$

Jika ragam dari komponen acak diketahui, penyelesaian BLUP mudah dikerjakan untuk memperoleh dugaan parameter β , dan *general least square* (GLS) untuk γ dan u sebagai berikut

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$\hat{\gamma} = \Sigma_\gamma Z^T V^{-1} (Y - X \hat{\beta})$$

$$\hat{u} = \Sigma_u D^T V^{-1} (Y - X \hat{\beta}).$$

Lebih lanjut, untuk suatu area kecil tertentu U_m , misalkan kita ingin mengetahui nilai tengahnya, maka kita gunakan pendekatan

$$\bar{y}_m = \bar{x}_m \hat{\beta} + \bar{z}_m \hat{\gamma} + e_m \hat{u},$$

yang merupakan suatu kombinasi linier dari penduga GLS dan BLUP sebagai komponen sintetik pada pendugaan area kecil. Sedangkan pendugaan MSE dari parameter yang menjadi perhatian, selanjutnya bisa digunakan hampiran yang dilakukan oleh Prasad dan Rao (1990) dengan mensubstitusi komponen sintetik dengan hasil dugaan P-spline.

4.4. Penerapan Model Aditif dan Nonparametrik pada Data Susenas

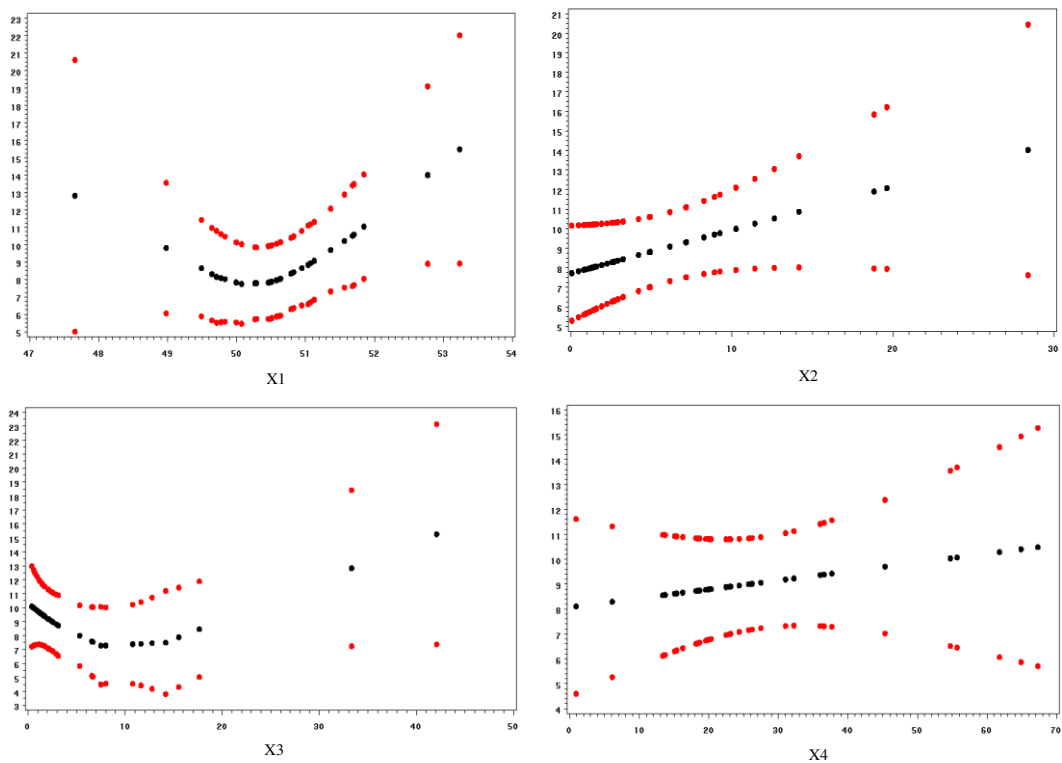
Kajian empirik menggunakan data yang dikumpulkan oleh BPS, data Podes 2005 sebagai sumber peubah penyerta dan data Susenas 2005 sebagai data survei, khususnya untuk Kota Bogor. Peubah yang menjadi perhatian adalah tingkat pengangguran yang direpresentasikan dengan persentasi tenaga kerja yang tidak sedang bekerja atau tidak memiliki pekerjaan tetap untuk setiap kelurahan di Kota Bogor. Kondisi tersebut diukur dalam satu minggu terakhir dari waktu survei. Persentasi banyaknya penduduk laki-laki (X_1), persentasi rumah tidak permanen (X_2), persentasi surat miskin yang

dikeluarkan kelurahan (X3), dan persentasi keluarga pra sejahtera dan sejahtera 1 (X4) digunakan sebagai peubah penyerta dalam kajian ini.

Hasil pendugaan untuk setiap metode yang digunakan disajikan pada Tabel 4.1. Seluruh metode pendugaan mengarah ke hasil yang diperoleh oleh teknik pendugaan langsung. **Kemungkinan faktor yang menyebabkan hal tersebut yang utama adalah pengaruh dari kondisi dimana keragaman antar area kecil yang diamati jauh lebih besar dibandingkan dengan keragaman akibat galat contoh di dalam setiap area kecil.** Walaupun demikian, **pendekatan GAMM mampu untuk mereduksi pengaruh peubah penyerta yang tidak memiliki pola hubungan linier.** Gambar 4.1 menyajikan *scater plot* dari peubah penyerta, dan peubah X1 serta X3 jelas tidak memiliki hubungan yang linier. Kedua peubah tersebut dengan menggunakan pendekatan GAMM diaproksimasi sesuai dengan gambaran yang disajikan pada Gambar 4.1 tersebut.

Table 4.1 Pendugaan tingkat pengangguran dalam persen di Kota Bogor berdasarkan data Susenas tahun 2005

Desa	Direct	GAMM	EBLUP	Desa	Direct	GAMM	EBLUP
1002 Pamoyanan	13.04	12.64	13.03	4006 Sempur	10.94	10.38	10.93
1005 Kertamaya	8.42	8.86	8.43	4010 Kebonkelapa	12.07	12.06	12.07
1006 Rancamaya	25.00	23.36	24.94	5002 Pasirkuda	20.00	17.60	19.95
1009 Muarasari	1.85	1.97	1.85	5003 Pasirjaya	13.51	12.91	13.49
1013 Batutulis	6.38	6.46	6.39	5004 Gunungbatu	10.64	10.31	10.63
1015 Empang	3.33	3.42	3.34	5006 Menteng	10.91	10.91	10.90
1016 Cikaret	9.80	9.74	9.80	5008 Cilendek Barat	16.67	15.81	16.64
2002 Sindangrasa	1.67	1.75	1.67	5009 Sindangbarang	6.38	6.72	6.39
2006 Sukasari	8.33	8.21	8.33	5012 Situgede	4.00	4.24	4.00
3001 Bantarjati	5.45	5.56	5.46	5015 Curugmekar	10.42	10.25	10.41
3002 Tegalgundil	6.90	6.98	6.90	6001 Kedungwaringin	6.38	6.33	6.39
3004 Cimahpar	3.28	3.59	3.29	6003 Kebonpedes	9.43	9.55	9.44
3006 Cibuluh	10.53	10.91	10.53	6004 Tanahsareal	11.54	10.92	11.53
3007 Kedunghalang	9.09	8.94	9.09	6005 Kedungbadak	6.38	6.35	6.38
3008 Ciparigi	4.88	5.16	4.88	6007 Sukadamai	12.50	11.99	12.49
4002 Gudang	14.81	14.48	14.79	6009 Kayumanis	5.45	5.56	5.47
4004 Tegallega	2.27	2.53	2.28	6011 Kencana	6.25	6.57	6.26



Gambar 4.1 Diagram pencar peubah penyerta

Berdasarkan kajian yang dilakukan, mampu ditunjukkan keunggulan *generalized additive mixed model* (GAMM) dibandingkan dengan *generalized linear mixed model* (GLMM) di dalam pendekatan EBLUP, setidaknya dapat ditemukan dalam dua aspek. Pertama, **GAMM bersifat bebas dari asumsi kelinieran hubungan di antara peubah penyerta dan peubah respon sehingga mampu untuk mereduksi masalah jika terjadi ketidaktepatan (*misspecification*) pemodelan di dalam EBLUP.** Aspek yang kedua, dengan kemampuannya untuk mengelaborasi pengaruh nonlinier dalam model, **GAMM mampu untuk mengatasi pola-pola yang tersembunyi dari peubah penyerta dan pada akhirnya akan meningkatkan akurasi dari pendugaan yang dilakukan.**

Pendekatan nonparametrik yang dilakukan untuk kasus yang sama juga menghasilkan pendugaan yang mirip dengan hasil dari pendekatan GAMM. Suatu kelemahan yang diperlihatkan dari pendekatan nonparametrik dalam kasus ini adalah sempitnya selang

pengukuran peubah penyerta. Dengan demikian, kemampuan pendekatan nonparametrik menjadi kurang memberikan makna untuk mengatasi pola hubungan antara peubah respon yang menjadi perhatian dengan peubah penyerta.

4.5. Pendekatan EBLUP Baku untuk Pendugaan Pengeluaran per Kapita di Kabupaten dan Kota Bogor

Peubah yang diamati dan menjadi perhatian dalam kajian ini adalah rata-rata pengeluaran perkapita rumah tangga. Sumber data yang digunakan adalah Susenas 2005 dengan materi informasi berbasis rumah tangga, serta Podes 2005 sebagai sumber data peubah penyerta. Peubah penyertanya adalah peubah-peubah yang diasumsikan mempengaruhi dan atau menggambarkan pengeluaran rumah tangga pada suatu wilayah, meliputi: (1) status administrasi desa/kelurahan, (2) persentasi rumah tangga petani, (3) persentasi buruh tani, (4) persentasi rumah tangga pelanggan listrik PLN, (5) persentasi luas daerah yang digunakan untuk persawahan, (6) persentasi rumah tangga penerima surat jaminan kesehatan, (7) jarak ke pusat kota, dan (8) waktu tempuh ke pusat kota.

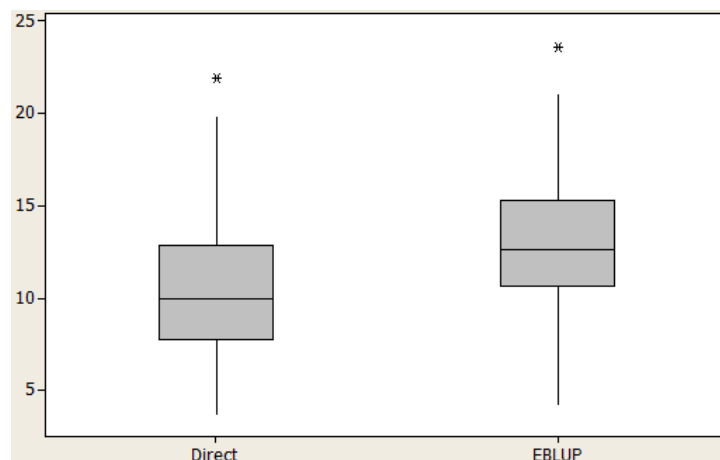
Kajian dilakukan pada 105 desa/kelurahan contoh yang disurvei pada Susenas 2005 dengan total rumah tangga sebanyak 1705. Sebaran contoh untuk masing-masing Kabupaten dan Kota Bogor disajikan pada Tabel 4.2.

Secara lengkap hasil analisis data disajikan pada Lampiran 4. Berdasarkan sebaran RRMSE seperti disajikan pada Gambar 4.2, hasil kajian dengan metode EBLUP memberikan dugaan RRMSE yang lebih bervariasi, diperlihatkan oleh bentang boxplot yang lebih panjang. Selain itu, median RRMSE berdasarkan metode EBLUP juga lebih besar dibandingkan dengan RRMSE berdasarkan metode pendugaan langsung.

Tabel 4.2 Sebaran contoh desa/kelurahan pada Susenas 2005 di Kabupaten dan Kota Bogor

Kabupaten/Kota	Ukuran contoh	Populasi desa/kelurahan
3201 Kab. Bogor	68	426
3271 Kota Bogor	37	68
Total	105	494

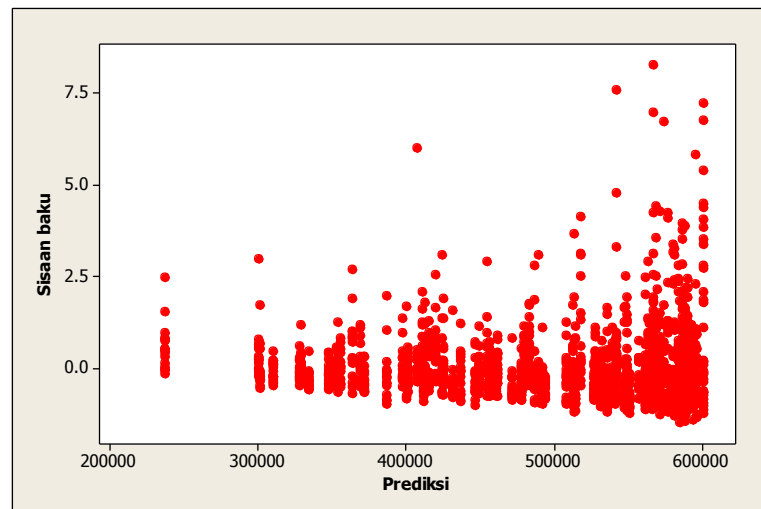
Kondisi tersebut menunjukkan bahwa metode pendugaan langsung relatif lebih bisa diterima dibandingkan dengan metode EBLUP. Namun demikian, secara teori semestinya metode EBLUP akan memberikan perbaikan pendugaan jika asumsi-asumsi yang mendasarinya terpenuhi.



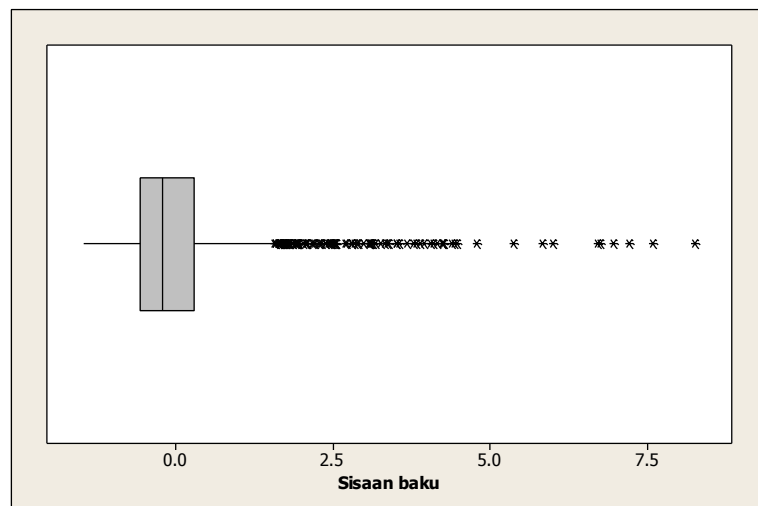
Gambar 4.2 Boxplot RRMSE penduga langsung dan EBLUP

Analisis sisaan berdasarkan model yang digunakan pada EBLUP menunjukkan bahwa model kurang baik seperti disajikan pada Gambar 4.3 sampai dengan Gambar 4.6. Plot sisaan dengan prediksi model menunjukkan pola yang mengindikasikan sisaan masih membentuk pola tertentu. Sedangkan histogram dan boxplot menunjukkan pola sisaan yang menceng ke sisi kanan, begitu pula plot quantil-quantil yang tidak membentuk pola linier. Kondisi tersebut mengindikasikan

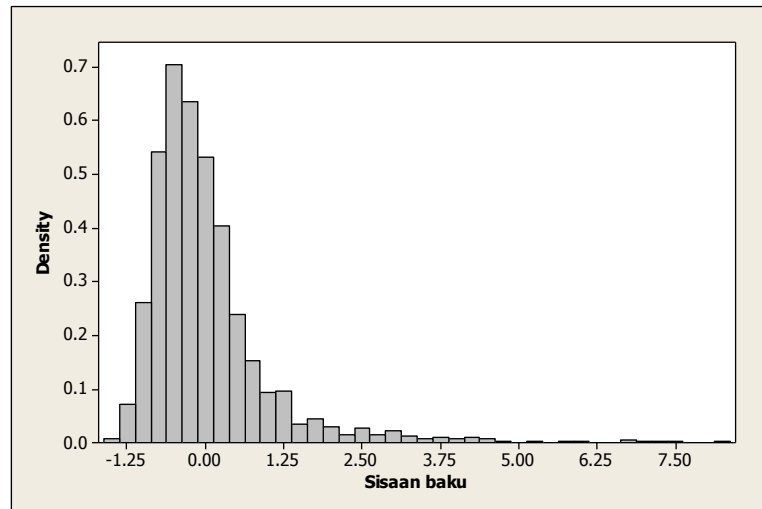
perlunya penanganan (transformasi) data terlebih dahulu, setidaknya pada peubah respon sebelum dilakukan pendugaan berdasarkan metode SAE. Gambar 4.7 memperjelas kondisi ini, dimana histogram peubah yang menjadi perhatian menunjukkan pola kemencengan yang cukup besar.



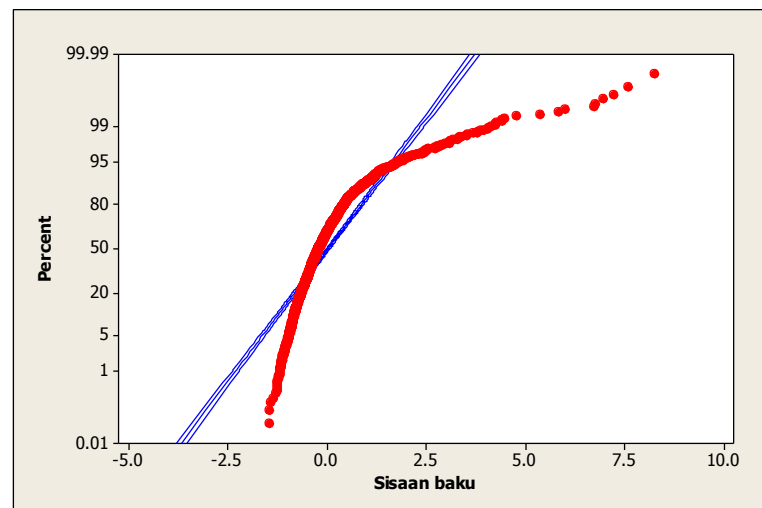
Gambar 4.3. Plot sisaan baku vs prediksi pada model dasar



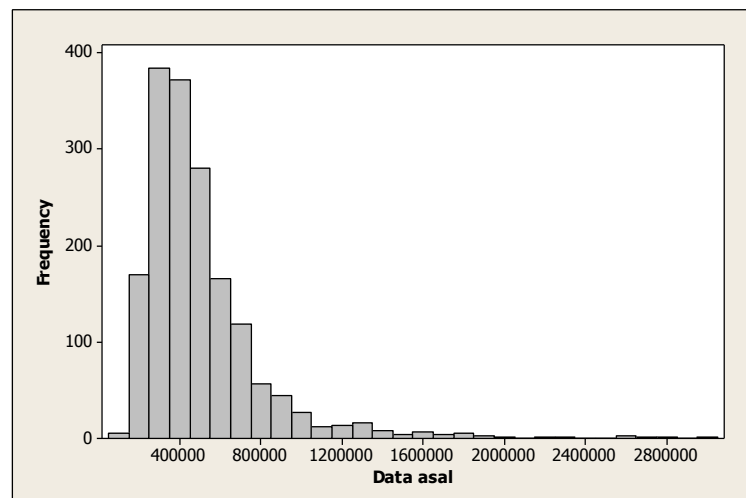
Gambar 4.4. Boxplot sisaan baku pada model dasar



Gambar 4.5. Histogram sisaan baku pada model dasar



Gambar 4.6. Plot QQ – sisaan baku pada model dasar



Gambar 4.7. Histogram data asal (peubah respon)

4.6. Kesimpulan

Pendugaan area kecil berbasis model yang diterapkan langsung pada data BPS (Susenas dan Podes) tidak memberikan perbaikan hasil pendugaan yang memuaskan. Model aditif dan nonparametrik yang diharapkan mampu mereduksi pengaruh ketidaktepatan penggunaan model linier, walaupun memberikan perbaikan masih belum memuaskan. Salah satu kelemahan yang diperlihatkan dari pendekatan aditif dan nonparametrik dalam kasus ini adalah sempitnya selang pengukuran peubah penyerta. Namun demikian, baik model aditif maupun nonparametrik karena memiliki sifat bebas dari asumsi kelinieran sehingga mampu mereduksi masalah jika terjadi ketidaktepatan pemodelan linier. Begitu pula keduanya mampu mengatasi pola-pola yang tersembunyi dari peubah penyerta.

Penggunaan EBLUP dalam konteks pendugaan pengeluaran per kapita atau data sosial secara umum menunjukkan hasil yang juga tidak memuaskan. Evaluasi sisaan dari model pendugaan pengeluaran per kapita memberikan petunjuk perlunya dilakukan upaya transformasi baik pada peubah yang menjadi perhatian, peubah penyerta atau kedua-duanya.

BAB V

PREDIKSI TERBAIK EMPIRIK

UNTUK MODEL TRANSFORMASI LOGARITMA

5.1. Latar Belakang

Dalam statistika, secara umum transformasi biasanya digunakan untuk mengatasi atau memenuhi asumsi kenormalan, kelinieran dan kehomogenan ragam (Carroll dan Ruppert, 1988). Namun demikian, literatur yang membahas model-model transformasi dalam SAE tidak mudah ditemukan. Sementara itu kebutuhan terhadap pengembangan model-model transformasi dalam analisis data survei semakin berkembang. Beberapa penelitian yang terkait dengan masalah ini seperti yang dilakukan oleh Chen dan Chen (1996) serta Karlberg (2000) yang membahas suatu penduga berdasarkan model superpopulasi lognormal-logistik untuk menduga total populasi dari suatu data survei dengan kemencengan (*skewness*) yang tinggi. Chambers dan Dorfman (2003) mendiskusikan penduga nilai tengah populasi secara umum dengan transformasi yang lazim digunakan untuk data kontinu.

Di sisi lain, seperti yang diuraikan pada Bab III dan Bab IV, untuk kasus data BPS, terutama dalam pendugaan tingkat kemiskinan yang direpresentasikan oleh pengeluaran per kapita, penerapan metode SAE baku tidak memberikan perbaikan yang memuaskan. Dalam Bab V ini, disajikan upaya perbaikan metode SAE baku dengan melakukan transformasi terlebih dahulu terhadap data respon sebelum menggunakan teknik SAE untuk menduga parameter nilai tengah dari peubah yang menjadi perhatian. Akibat dari transformasi ini adalah dihasilkannya penduga parameter dari transformasi-balik (*back-transformation*) yang bersifat berbias serta munculnya kesulitan untuk menurunkan kuadrat tengah galatnya. Pembahasan dua hal tersebut disajikan dalam Sub-bab 5.2 yang disertai pula dengan kajian simulasi

pada Sub-bab 5.3 serta kajian empirik untuk data BPS di Kabupaten dan Kota Bogor pada Sub-bab 5.4.

5.2. Model Transformasi Logaritma

Ada dua pendekatan yang digunakan dalam pendugaan area kecil yang tergantung pada ketersediaan data peubah penyerta. Pertama model taraf area, yaitu dicirikan dengan ketersediaan data peubah penyerta hanya ada pada taraf area kecil yang menjadi perhatian. Model kedua adalah taraf unit, yaitu jika peubah penyerta yang tersedia sampai pada unit contoh dalam survei yang dilakukan dan biasanya peubah penyerta yang digunakan berasal dari catatan administratif atau sensus.

Model taraf area memiliki bentuk $\theta_i = g(Y_i) = x_i^T \beta + v_i$ dengan v_i bersifat bebas stokastik identik (iid), dan x_i adalah vektor berukuran p yang memuat peubah penyerta untuk masing-masing area kecil, sedangkan $g(Y_i)$ adalah fungsi dari total pengamatan suatu respon pada area kecil yang menjadi perhatian. Model tersebut biasanya digabung dengan model penarikan contoh menjadi $y_i = \theta_i + e_i$, dimana $y_i = g(\hat{Y}_i)$ adalah penduga langsung dari θ_i . Adapun $e_i | \theta_i$ diasumsikan menyebar $N(0, \sigma_i^2)$ dengan galat contoh σ_i^2 yang diketahui serta e_i saling bebas antar area. Jika seluruh area diamati, maka area akan berfungsi seperti strata dalam percontohan berstrata, akan tetapi dalam prakteknya tidak seluruh area dapat diamati. Untuk kasus kedua ini, diasumsikan kombinasi dari model taraf area $y_i = x_i^T \beta + v_i + e_i$ juga merepresentasikan area yang tidak diamati, dalam hal ini diasumsikan bahwa tidak ada bias pemilihan contoh (Pfeffermann dan Sverchkov, 2003).

Dalam penelitian ini parameter yang menjadi perhatian adalah nilai tengah populasi terhitung $\mu_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$ dengan N_i adalah banyaknya anggota populasi untuk area ke- i dan y_{ij} adalah nilai peubah yang menjadi perhatian pada pengamatan ke- j dari area ke- i . Didefinisikan suatu transformasi logaritma dalam model linier campuran (*log-scale linear mixed model*) untuk peubah ini sebagai berikut

$$\log(y_{ij}) = x_i^T \beta + \nu_i + \varepsilon_{ij} \quad (5.1)$$

dengan pengaruh individu ε_{ij} mengikuti sebaran *iid* $N(0, \sigma_i^2)$, pengaruh area ν_i menyebar *iid* $N(0, \tau^2)$, dan β_k , σ_i^2 serta τ^2 adalah parameter yang tidak diketahui. Adapun peubah penyerta yang diamati tersedia pada level area atau dikenal dengan istilah *contextual covariates*.

Dengan mengikuti teori EBLUP baku untuk model (5.1), yaitu EBLUP untuk nilai tengah θ_i dari $\log(y_{ij})$, maka penduga bagi θ_i dapat ditulis sebagai berikut

$$\hat{\theta}_i^{EBLUP} = \gamma_i \hat{\theta}_i^D + (1 - \gamma_i) x_i^T \hat{\beta} \quad (5.2)$$

dengan $\hat{\beta}$ biasanya diperoleh berdasarkan metode kuadrat terkecil terboboti untuk parameter regresi β dari *log-scale linear mixed model*, $\gamma_i = \hat{\tau}^2 / (\hat{\tau}^2 + n_i^{-1} \hat{\sigma}_i^2)$, $\hat{\theta}_i^D = n_i^{-1} \sum_{j \in s(i)} \log(y_{ij})$ adalah penduga langsung bagi θ_i berdasarkan data contoh $s(i)$ untuk area ke- i dan 'tanda topi' melambangkan penduga parameter yang diperoleh dari contoh.

Karena yang diinginkan adalah suatu penduga aktual untuk nilai tengah pada setiap area ke- i , maka digunakan sifat sebaran lognormal untuk melakukan transformasi-balik dari model (5.2). Lebih lanjut, diasumsikan bahwa sebaran penarikan contoh bagi $\hat{\theta}_i^{EBLUP}$ adalah

$N\{\theta_i, \text{Var}(\hat{\theta}_i^{EBLUP})\}$. Dengan demikian, penduga nilai tengah aktual (*raw-scale*) untuk area ke- i adalah

$$\hat{\mu}_i = \exp\left(\hat{\theta}_i^{EBLUP} + \frac{1}{2}\hat{v}_i^{EBLUP}\right) \quad (5.3)$$

dengan \hat{v}_i^{EBLUP} adalah penduga bagi *mean squared prediction error* (MSPE) dari $\hat{\theta}_i^{EBLUP}$. Dalam hal ini penduga MSPE dihitung berdasarkan formula yang dikembangkan oleh Prasad dan Rao (1990). Kemudian, penduga MSE bagi penduga nilai tengah pada (5.3) dapat diperoleh sebagai berikut

$$\hat{V}(\hat{\mu}_i) = e^{\hat{v}_i^{EBLUP}} \left(e^{\hat{v}_i^{EBLUP}} - 1 \right) e^{2\hat{\theta}_i^{EBLUP}}. \quad (5.4)$$

Alternatif lain dikembangkan suatu pendekatan prediktor terbaik empirik (*empirical best predictor*, EBP) untuk nilai tengah area kecil yang didasarkan pada *log-scale linear mixed model* (persamaan 5.1) yaitu dengan mengganti nilai y_{ij} yang bukan anggota contoh dengan nilai harapan dugaannya.

Jika kita definisikan total pengamatan $T_i = \sum_{j \in s(i)} y_{ij} + \sum_{j \in r(i)} y_{ij}$ adalah penjumlahan dari nilai pengamatan yang merupakan anggota contoh ($j \in s(i)$) dan nilai unit yang bukan anggota contoh ($j \in r(i)$), maka dugaan total pengamatan bisa dituliskan sebagai berikut

$$\hat{T}_i = \sum_{j \in s(i)} y_{ij} + \sum_{j \in r(i)} \hat{y}_{ij}. \quad (5.5)$$

Dengan demikian, pendekatan baku untuk membentuk suatu EBP dapat dibentuk dari transformasi-balik nilai harapan *log-scale* berdasarkan *log-scale linear mixed model* untuk menduga nilai unit yang bukan anggota contoh. Hal ini akan menghasilkan penduga bagi μ_i sebagai berikut

$$\hat{\mu}_i^{naive} = f_i \bar{y}_i^s + (1 - f_i) \exp(x_i^T \hat{\beta}) \quad (5.6)$$

dengan $f_i = n_i N_i^{-1}$ adalah fraksi contoh pada area ke- i , \bar{y}_i^s adalah rata-rata contoh dari y_{ij} pada area ke- i , $\exp(x_i^T \hat{\beta})$ adalah penduga nilai tengah aktual y_{ij} yang bukan anggota contoh dan diperoleh dari transformasi-balik, serta s merupakan indeks untuk unit yang terpilih sebagai anggota contoh. Akan tetapi, dapat diperlihatkan bahwa $\hat{\mu}_i^{naive}$ berbias untuk nilai tengah dari *log-scale linear mixed model*. Bukti $\hat{\mu}_i^{naive}$ berbias disajikan pada Lampiran 1.

Untuk mengatasi hal tersebut didefinisikan $\eta_i^2 = \tau^2 + \sigma_i^2$, dan berdasarkan asumsi lognormal untuk y_{ij} maka diperoleh alternatif pendugaan sebagai berikut

$$\tilde{\mu}_i^{BP} = f_i \bar{y}_i^s + (1 - f_i) \exp\left(x_i^T \hat{\beta} + \frac{\eta_i^2}{2}\right). \quad (5.7)$$

Lebih lanjut diperoleh :

$$\begin{aligned} \text{Var}(\tilde{\mu}_i^{BP} - \mu_i) &= (1 - f_i)^2 \text{Var}\left\{\exp\left(x_i^T \hat{\beta} + \frac{\eta_i^2}{2}\right) - \bar{y}_{r(i)}\right\} \\ &= \left(\frac{N_i - n_i}{N_i}\right)^2 \text{Var}\left\{\exp\left(x_i^T \hat{\beta} + \frac{\eta_i^2}{2}\right) - \bar{y}_{r(i)}\right\} \\ &= N_i^{-2} \left\{ (N_i - n_i)^2 \text{Var}\left\{\exp\left(x_i^T \hat{\beta} + \frac{\eta_i^2}{2}\right) - \bar{y}_{r(i)}\right\} \right\} \\ &\simeq N_i^{-2} \left\{ \text{Var}\left(\sum_{j \in r(i)} \tilde{y}_{ij}\right) + \text{Var}\left(\sum_{j \in r(i)} y_{ij}\right) \right\} \end{aligned} \quad (5.8)$$

dengan

$$\tilde{y}_{ij} = \exp\left(X_i^T \hat{\beta} + \frac{\eta_i^2}{2}\right)$$

$$\text{Var}\left(\sum_{j \in r(i)} y_{ij}\right) = (N_i - n_i) \sigma_i^2 + (N_i - n_i)(N_i - n_i - 1) \tau^2$$

dan

$$\begin{aligned}
\text{Var}\left(\sum_{j \in r(i)} \tilde{y}_{ij}\right) &= \sum_{j \in r(i)} \sum_{k \in r(i)} \text{Cov}(\tilde{y}_{ij}, \tilde{y}_{ik}) \\
&= (N_i - n_i)^2 \text{Var}\left\{\exp\left(x_i^T \hat{\beta} + \frac{\eta_i^2}{2}\right)\right\} \\
&= (N_i - n_i)^2 e^{\eta_i^2} \text{Var}\left\{\exp\left(x_i^T \hat{\beta}\right)\right\}.
\end{aligned}$$

Berdasarkan pendekatan deret Taylor orde pertama untuk

$\text{Var}\left\{\exp\left(x_i^T \hat{\beta}\right)\right\}$ sebagai berikut :

Misalkan $Z_i(\beta) = e^{x_i^T \beta}$, maka $Z_i(\hat{\beta}) \simeq Z_i(\beta) + (\hat{\beta} - \beta) Z_i^{(1)}(\beta)$ dan

$Z_i(\hat{\beta}) - Z_i(\beta) \simeq (\hat{\beta} - \beta) Z_i^{(1)}(\beta)$. Dengan mensubstitusi $Z_i^{(1)}(\beta) = x_i^T e^{x_i^T \beta}$ pada

persamaan $Z_i(\hat{\beta}) - Z_i(\beta)$, maka diperoleh

$$\begin{aligned}
E\left[Z_i(\hat{\beta}) - Z_i(\beta)\right]^2 &\simeq E\left[\left(x_i^T e^{x_i^T \beta}\right)^2 (\hat{\beta} - \beta)^2\right] \\
&= e^{2x_i^T \beta} x_i^T \text{Var}(\hat{\beta}) x_i,
\end{aligned}$$

sehingga diperoleh $\text{Var}\left(\sum_{j \in r(i)} \tilde{y}_{ij}\right) \approx (N_i - n_i)^2 e^{2x_i^T \hat{\beta} + \eta_i^2} x_i^T \text{Var}(\hat{\beta}) x_i$.

Dengan demikian diperoleh pendekatan bagi penduga $\text{Var}(\tilde{\mu}_i^{BP} - \mu_i)$ adalah

$$\begin{aligned}
\hat{V}(\tilde{\mu}_i^{BP} - \mu_i) &\simeq N_i^{-2} \left((N_i - n_i)^2 e^{2x_i^T \hat{\beta} + \eta_i^2} x_i^T \text{Var}(\hat{\beta}) x_i \right) + (N_i - n_i) \sigma_i^2 \\
&\quad + (N_i - n_i)(N_i - n_i - 1) \tau^2 \\
&\simeq (1 - f_i)^2 \left\{ e^{2x_i^T \hat{\beta} + \sigma_i^2 + \tau^2} x_i^T \text{Var}(\hat{\beta}) x_i + \tau^2 \right\} + N_i^{-2} (N_i - n_i) \sigma_i^2. \quad (5.9)
\end{aligned}$$

Di dalam prakteknya, komponen ragam σ_i^2 dan τ^2 dari *log-scale linear mixed model* biasanya tidak diketahui dan diduga dari data contoh.

Dengan demikian, kita mengganti $\tilde{\mu}_i^{BP}$ dengan EBP

$$\hat{\mu}_i^{EBP} = f_i \bar{y}_i^s + (1 - f_i) \exp\left(x_i^T \hat{\beta} + \frac{\hat{\sigma}_i^2 + \hat{\tau}^2}{2}\right). \quad (5.10)$$

Dalam aplikasinya pendugaan MSPE dari $\hat{\mu}_i^{EBP}$ didasarkan pada keseluruhan ukuran contoh yang relatif besar sehingga pengaruh

keragaman karena mensubstitusi σ_i^2 oleh $\hat{\sigma}_i^2$ dapat diabaikan. Asumsi ini tidak bertentangan dengan konsep kekonvergenan pendugaan σ^2 untuk n yang besar. Oleh karena itu, kajian lebih lanjut difokuskan pada pendugaan *extra variability* yang disebabkan oleh pendugaan τ^2 . Dengan demikian, MSPE dimaksud dapat ditulis sebagai berikut

$$MSPE(\hat{\mu}_i^{EBP}) = Var(\tilde{\mu}_i^{BP} - \mu_i) + 2E(\hat{\mu}_i^{EBP} - \tilde{\mu}_i^{BP})(\tilde{\mu}_i^{BP} - \mu_i) + E(\hat{\mu}_i^{EBP} - \tilde{\mu}_i^{BP})^2.$$

Sebelumnya telah diuraikan penduga bagi $Var(\tilde{\mu}_i^{BP} - \mu_i)$ seperti pada persamaan (5.9). Mengikuti kaidah praktis, diasumsikan bahwa $E(\hat{\mu}_i^{EBP} - \tilde{\mu}_i^{BP})(\tilde{\mu}_i^{BP} - \mu_i) = 0$, sehingga perhatian sekarang tinggal pada $E(\hat{\mu}_i^{EBP} - \tilde{\mu}_i^{BP})^2$. Dengan menggunakan deret Taylor, kita bisa memperoleh

$$\hat{\mu}_i^{EBP} - \tilde{\mu}_i^{BP} \simeq (1 - f_i)(\hat{\tau}^2 - \tau^2) \frac{1}{2} \exp\left(x_i^T \hat{\beta} + \frac{\hat{\sigma}_i^2 + \hat{\tau}^2}{2}\right)$$

dan dengan demikian, dapat ditunjukkan bahwa

$$E(\hat{\mu}_i^{EBP} - \tilde{\mu}_i^{BP})^2 \simeq \frac{1}{4}(1 - f_i)^2 e^{2x_i^T \hat{\beta} + \hat{\sigma}_i^2 + \hat{\tau}^2} \nabla(\hat{\tau}^2) \quad (5.11)$$

dengan $\nabla(\hat{\tau}^2)$ adalah ragam asimtotik (*asymptotic variance*) dari $\hat{\tau}^2$. Akhirnya dengan menggabungkan persamaan (5.9) dengan (5.11) diperoleh penduga MSPE bagi $\hat{\mu}_i^{EBP}$ sebagai berikut :

$$\begin{aligned} \hat{V}(\hat{\mu}_i^{EBP}) = (1 - f_i)^2 \left\{ e^{2x_i^T \hat{\beta} + \hat{\sigma}_i^2 + \hat{\tau}^2} \left(x_i^T \hat{V}(\hat{\beta}) x_i + \frac{1}{4} \hat{V}(\hat{\tau}^2) \right) + \hat{\tau}^2 \right\} \\ + N_i^{-2} (N_i - n_i) \hat{\sigma}_i^2. \end{aligned} \quad (5.12)$$

5.3. Evaluasi Sifat Statistik Model EBP Berdasarkan Simulasi

Kajian simulasi dibangun berdasarkan *model-based simulation* berdasarkan karakteristik dari 105 contoh desa/kelurahan di Kabupaten dan Kota Bogor seperti telah diuraikan pada Sub-bab 4.6. Populasi terhitung untuk kajian ini dibangun atas dasar *log-scale linear*

mixed model dengan peubah penyerta terdiri dari delapan karakteristik yang bersesuaian berdasarkan sensus desa pada Podes 2005. Sedangkan peubah yang menjadi perhatian didefinisikan diperoleh dari level rumah tangga yang menjadi anggota contoh di Kabupaten dan Kota Bogor pada Susenas 2005.

Kajian simulasi ini dilakukan untuk empat penduga yaitu : (1) penduga langsung untuk data asli (Direct), (2) EBLUP untuk data asli (EBLUP), (3) transformasi-balik EBLUP, persamaan 5.3 (Log-scale EBLUP), dan (4) lognormal-EBP, persamaan 5.10 (Lognormal EBP).

Table 5.1 Ringkasan hasil kajian simulasi

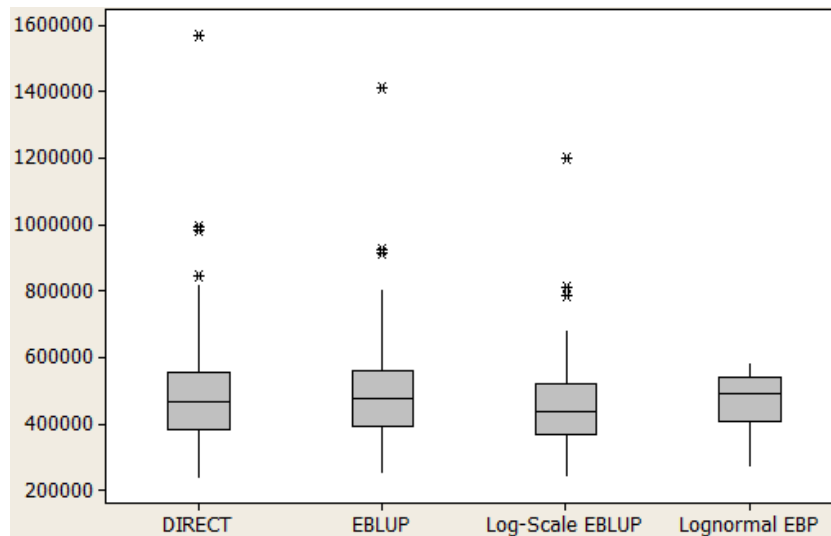
	Direct	EBLUP	Log-scale EBLUP	Lognormal EBP
ARB (%)	0.0998	0.1076	-9.4858	-8.2409
ARRMSE (%)	11.3181	6.3050	4.8122	3.5035
ARBMSE (%)	1.8812	-2.4099	-2.6634	-2.8505
ACR (%)	91.1111	93.7698	91.5635	94.3651

Hasil kajian simulasi ditunjukkan pada Tabel 5.1 yang merupakan evaluasi rata-rata untuk : (1) bias relatif (ARB), (2) *root mean squared error* relatif (ARRMSE), (3) bias MSE relatif (ARBMSE), dan (4) *coverage rates* untuk selang kepercayaan 95 (ACR) yang didasarkan pada masing-masing penduga dan MSE yang bersesuaian.

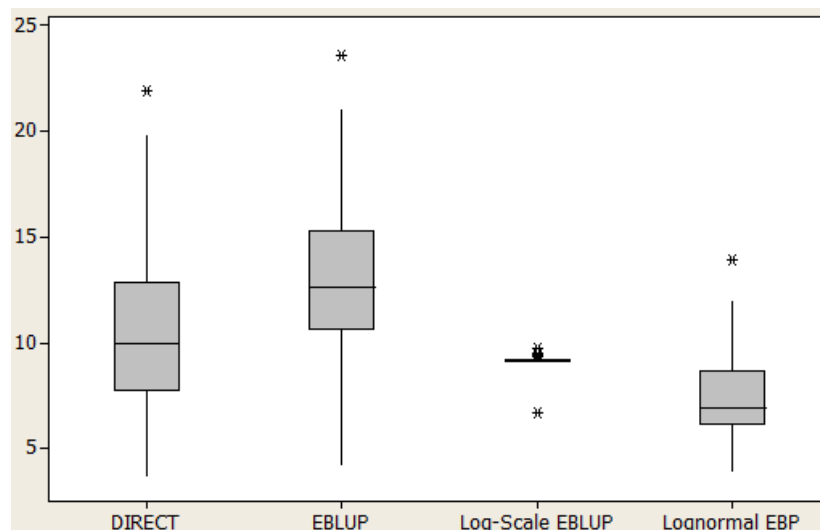
Dari kajian simulasi ini diperoleh bahwa model yang diajukan baik log-scale EBLUP maupun lognormal EBP memberikan hasil yang lebih baik terutama untuk RMSE relatif. Sifat *underestimate* bias MSE dari lognormal EBP sampai saat ini belum bisa ditentukan penyebabnya. Namun demikian diindikasikan setidaknya dipengaruhi oleh dua hal yaitu karena mengabaikan pengaruh pendugaan σ_i^2 dan bias pendugaan.

5.4. Penerapan pada Data Susenas

Sub-bab ini menyajikan aplikasi dari metode SAE yang telah dibahas pada sub-bab sebelumnya. Data yang digunakan sama dengan contoh kasus pada Sub-bab 4.6 yaitu pendugaan pengeluaran per kapita di 105 desa/kelurahan di Kabupaten dan Kota Bogor yang merupakan desa contoh dalam Susenas 2005.



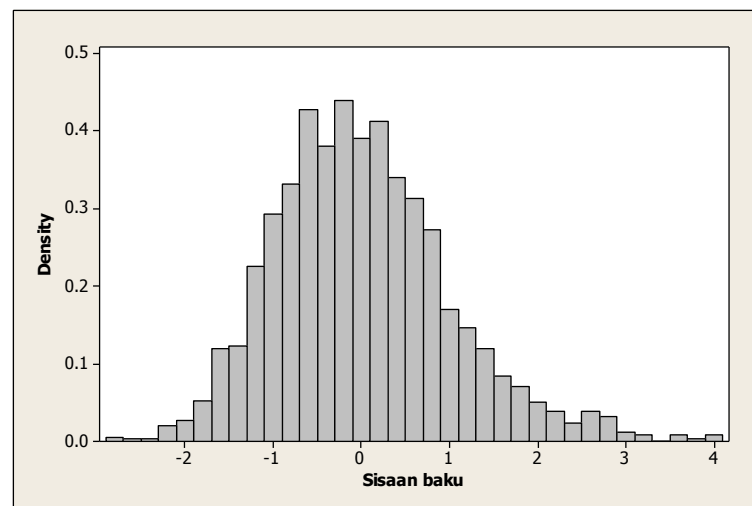
Gambar 5.1. Boxplot penduga pengeluaran per kapita desa/kelurahan di Kabupaten dan Kota Bogor berdasarkan empat metode pendekatan



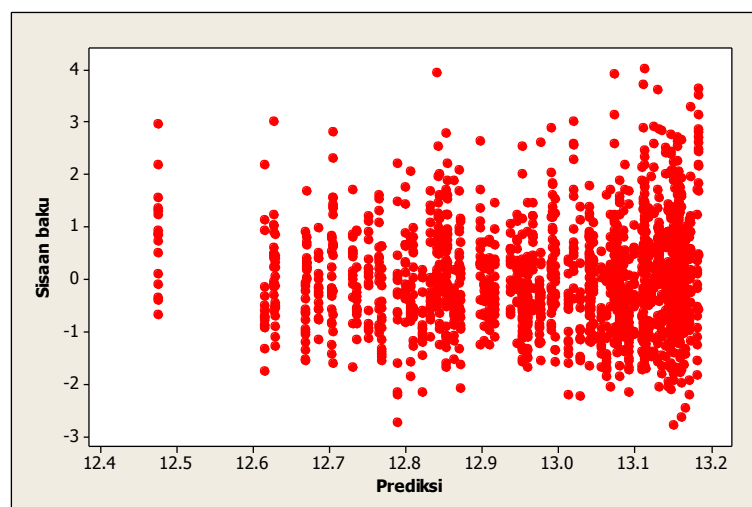
Gambar 5.2. Boxplot penduga MSE relatif (%) pada pendugaan pengeluaran per kapita desa/kelurahan di Kabupaten dan Kota Bogor berdasarkan empat metode pendekatan

Hasil kajian menunjukkan bahwa metode yang dikembangkan (lognormal EBP) memberikan dugaan MSE relatif yang lebih baik seperti ditunjukkan pada Gambar 5.2 maupun Lampiran 9.

Evaluasi sisaan untuk model transformasi logaritma pun sudah cukup baik memenuhi asumsi kenormalan seperti ditunjukkan pada Gambar 5.3. Begitu pula plot sisaan baku dengan prediksi (Gambar 5.4) menunjukkan pola acak yang cukup baik.



Gambar 5.3. Histogram sisaan baku pada model transformasi



Gambar 5.4. Plot sisaan baku vs prediksi pada model transformasi

5.5. Kesimpulan

Model yang diajukan, lognormal EBP, memberikan hasil yang lebih baik seperti yang diperlihatkan oleh RRMSE yang paling kecil, walaupun bias relatifnya masih cukup besar. Sifat *underestimate* MSPE dari lognormal EBP diindikasikan setidaknya dipengaruhi oleh dua hal yaitu karena (1) mengabaikan pengaruh pendugaan σ_i^2 dan (2) pengaruh bias pendugaan. Namun demikian, secara umum ada indikasi bahwa model yang diajukan mampu meningkatkan efisiensi pendugaan.

BAB VI

PEMBAHASAN

Metode pendugaan area kecil berupaya mencari solusi efisien pada kondisi ukuran contoh sedikit (kecil) dengan cara meningkatkan efisiensi ukuran contoh. Dalam metode statistika klasik, ukuran contoh kecil sangat tidak diharapkan karena akan menghasilkan statistik dengan ragam yang besar. Namun demikian, pembentukan *post-stratification* dalam analisis data survei biasanya akan menghasilkan kelompok-kelompok yang anggota contohnya sedikit atau bahkan tidak terwakili di dalam survei. Hal yang sama juga bisa terjadi jika kita melakukan *break-down* area dengan batas-batas wilayah administratif.

Upaya meningkatkan efisiensi ukuran contoh dalam SAE dilakukan dengan menambah informasi baik dari luar maupun dari dalam area itu sendiri yang berupa informasi data sensus, catatan administrasi atau pun hasil survei terdahulu. Dengan pembentukan model yang tepat, informasi-informasi tambahan tersebut dirangkai sehingga mampu memperbaiki akurasi dan presisi pendugaan parameter pada setiap area kecil yang menjadi perhatian.

Penerapan SAE di Indonesia, khususnya dalam memanfaatkan data Susenas, memiliki kekhasan tersendiri. Pada level desa/kelurahan data dikumpulkan dari contoh acak blok-sensus sebanyak 16 rumah tangga untuk setiap blok-sensus terpilih. Setiap desa/kelurahan mayoritas diwakili oleh satu blok-sensus. Kendala yang dihadapi dalam menyediakan peubah penyerta adalah tidak tersedianya informasi populasi sampai ke unit rumah tangga. Informasi populasi diperoleh dari Podes sehingga sifatnya kontekstual karena tersedia hanya untuk tingkat area, dalam hal ini desa/kelurahan.

Pada Bab III dilakukan kajian eksplorasi untuk menelaah metode GREG pada data Susenas. Metode GREG digunakan karena relatif mudah dalam aplikasinya. Metode GREG juga bebas dari pengaruh secara langsung jika terjadi ketidaktepatan dalam memodelkan hubungan peubah yang menjadi perhatian dengan peubah penyerta. Selain itu, metode GREG juga secara formula tidak dipengaruhi secara langsung oleh faktor rasio antara keragaman antar area kecil dengan keragaman total.

Penyusunan pembobot dalam metode GREG yang sesuai dengan metode penarikan contoh yang dilakukan akan sangat membantu dalam meningkatkan presisi pendugaan parameter. Kendala yang dihadapi dalam eksplorasi metode ini adalah tidak tersedianya informasi jumlah rumah tangga untuk setiap blok-sensus terpilih sehingga digunakan nilai harapannya. Ketidaktersediaan jumlah rumah tangga ini karena blok-sensus BPS tidak bersifat administratif melainkan berdasarkan kemudahan geografis. Walaupun demikian, metode GREG dengan tidak mengabaikan metode penarikan contoh gerombol dua tahap mampu meningkatkan presisi pendugaan parameter dengan baik. Modifikasi GREG dengan memasukkan *M-Estimator* pada pendugaan parameter regresi mampu meningkatkan presisi pendugaan untuk data yang memuat pencilan sampai dengan 10%. Namun demikian, perlu mempertimbangkan metode lain jika ditemukan komposisi pencilan sampai 20% atau lebih dari data.

Dalam pendugaan area kecil, khususnya pendugaan berbasis model, konsep model linier menjadi inti dari analisis yang dilakukan. Keterpenuhan asumsi linier tentu menjadi syarat penting dalam melakukan analisis lebih lanjut. Dalam Bab IV disajikan eksplorasi metode berbasis model aditif dan nonparametrik untuk mengurangi pengaruh ketidaktepatan pemodelan linier yang digunakan. Namun demikian, dalam kasus data Susenas, walaupun model aditif dan nonparametrik memberikan perbaikan presisi pendugaan namun

masih belum memuaskan. Salah satu kelemahan yang diperlihatkan dalam kasus ini adalah sempitnya selang pengukuran peubah penyerta.

Penggunaan EBLUP baku dalam konteks pendugaan pengeluaran per kapita atau data sosial secara umum menunjukkan hasil yang juga tidak memuaskan. Evaluasi sisaan dari model pendugaan pengeluaran per kapita memberikan petunjuk perlunya dilakukan upaya transformasi baik pada peubah yang menjadi perhatian, peubah penyerta atau kedua-duanya.

Upaya perbaikan metode SAE berbasis model dengan terlebih dahulu melakukan transformasi terhadap data respon sebelum diterapkan teknik SAE menjadi fokus utama dalam disertasi ini. Akibat yang dihadapi dari melakukan transformasi ini adalah diperolehnya penduga parameter dari transformasi-balik yang bersifat berbias serta kesulitan penurunan kuadrat tengah galatnya. Pembahasan dua hal tersebut disajikan secara mendetail dalam Bab V yang juga disertai dengan evaluasi sifat-sifat statistiknya melalui simulasi serta contoh aplikasi untuk data BPS (Susenas dan Podes) di Kabupaten dan Kota Bogor.

Model yang diajukan, lognormal EBP, memberikan hasil yang lebih baik seperti yang diperlihatkan oleh RRMSE yang paling kecil, walaupun bias relatifnya masih cukup besar. Namun demikian, secara umum model yang diajukan mampu meningkatkan efisiensi pendugaan.

Dalam penelitian ini belum dilakukan upaya-upaya untuk memperbaiki bias pada pendugaan nilai tengah dari model lognormal EBP maupun bias pada MSPE-nya. Namun demikian suatu alternatif untuk memperbaiki bias pendugaan nilai tengah dapat dipertimbangkan konsep sebagai berikut :

Jika kita perhatikan $\hat{\delta}_i = (\hat{\beta}, \hat{\eta}_i^2)$ dan $Z(\hat{\delta}_i) = \exp(x_{ij}^T \hat{\beta} + \hat{\eta}_i^2 / 2)$, maka dengan penguraian deret Taylor diperoleh

$$Z(\hat{\delta}_i) \simeq Z(\delta_i) + (\hat{\delta}_i - \delta_i) Z^{(1)}(\delta_i) + \frac{1}{2} (\hat{\delta}_i - \delta_i)^T Z^{(2)}(\delta_i) (\hat{\delta}_i - \delta_i).$$

Nilai harapan bagi $Z(\hat{\delta}_i)$ adalah

$$\begin{aligned} E\{Z(\hat{\delta}_i)\} &\simeq E\{Z(\delta_i)\} + \frac{1}{2} \text{tr} \left[Z^{(2)}(\delta_i) E\{(\hat{\delta}_i - \delta_i)(\hat{\delta}_i - \delta_i)^T\} \right] \\ &\simeq \exp(x_{ij}^T \beta + \eta_i^2 / 2) + \exp(x_{ij}^T \beta + \eta_i^2 / 2) [k] \\ &\simeq \exp(x_{ij}^T \beta + \eta_i^2 / 2) [k^*] \end{aligned}$$

Dengan demikian untuk mengoreksi bias pendugaan lognormal EBP, maka komponen $\exp\left(X_i^T \hat{\beta} + \frac{\hat{\eta}_i^2}{2}\right)$ pada persamaan (5.7) dapat dikoreksi dengan mengalikan konstanta $[k^*]^{-1}$.

Adapun upaya untuk memperbaiki MSPE bisa dilakukan dengan memformulasikan pengaruh pendugaan σ_i^2 . Upaya ini bisa dilakukan melalui ekspansi deret Taylor. Teknik lain yang bisa dipertimbangkan untuk hal yang sama adalah metode berbasis *resampling* seperti jackknife dan bootstrap. Kedua teknik tersebut sudah banyak digunakan dalam pendugaan MSPE.

BAB VII

KESIMPULAN DAN SARAN

7.1. Kesimpulan

Permasalahan yang dihadapi dalam pendugaan area kecil adalah besarnya keragaman dari penduga langsung sehingga statistik yang diperoleh tidak efisien. Untuk meningkatkan efektifitas ukuran contoh, SAE menambah informasi baik dari area itu sendiri, area lain maupun survei lainnya.

Model baku SAE seperti EBLUP kadangkala tidak mampu menerangkan data dengan baik karena ketatnya asumsi kelinieran dari model. Upaya perbaikan dengan memodifikasi EBLUP untuk model yang terlebih dahulu dilakukan transformasi logaritma, disebut lognormal EBP, mampu meningkatkan efisiensi pendugaan. Perbaikan terhadap metode ini masih diperlukan mengingat hasil dugaannya masih berbias, begitu pula MSPE-nya masih bersifat *underestimate*. Namun demikian, secara umum model lognormal EBP mampu meningkatkan efisiensi pendugaan

7.2. Saran

Penelitian ini belum mampu menunjukkan bentuk eksplisit baik nilai harapan maupun ragam dari penduga yang diperoleh. Dengan demikian evaluasi sifat statistik dari penduga EBP secara analitik masih perlu pengkajian lebih lanjut terutama setelah dimasukkannya komponen $\hat{\sigma}_i^2$ dan $\hat{\tau}^2$ dalam model dugaan. Jika kedua komponen ragam tersebut konstan (diketahui) telah diperlihatkan bahwa penduga lognormal EBP memiliki sifat tak bias dan penduga keragamannya telah diturunkan menggunakan aproksimasi deret Taylor.

DAFTAR PUSTAKA

- Agresti A. 2002. *Categorical Data Analysis*, 2nd Edition. New York: Wiley Press.
- Agresti A. 2007. *An Introduction to Categorical Data Analysis*, 2nd Ed. New York: Wiley Press.
- Butar F. 1997. *Empirical Bayes Methods in Survey Sampling*. Disertasi pada The Graduate College at the University of Nebraska, Lincoln – Nebraska.
- Butar F dan Lahiri P. 2003. On measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference*, 112, 63-76.
- Carroll RJ dan Ruppert D. 1988. *Transformations and Weighting in Regression*, London: Chapman and Hall.
- Chambers R dan Chandra H. 2006. *Improved Direct Estimators for Small Areas*, Working Paper M06/07, Southampton Statistical Sciences Research Institute.
- Chambers R dan Dorfman AH. 2003. *Transformed Variables in Survey Sampling*, Working paper M03/21, Southampton Statistical Sciences Research Institute.
- Chambers R dan Tzavidis N. 2006. M-quantile Models for Small Area Estimation. *Biometrika*, 93, 2, 255–268.
- Chandra H, Salvati N dan Chambers R. 2007. *Small Area Estimation for Spatially Correlated Populations – A Comparison of Direct and Indirect Model-Based Methods*, Working Paper M07/09, Southampton Statistical Sciences Research Institute.
- Chand N dan Alexander CH. 1995 *Using Administrative Records for Small Area Estimation in the American Community Survey*. US Bureau of the census.
- Chen S. 2001. *Empirical Best Prediction and Hierarchical Bayes Methods in Small Area Estimation*. Disertasi pada The Graduate College at the University of Nebraska, Lincoln – Nebraska.
- Chen G dan Chen J. 1996. A Transformation Method for Finite Population Sampling Calibrated with Empirical Likelihood, *Survey Methodology*, 22, 139-46.
- Chen S dan Lahiri P. 2005. On mean squared prediction error estimation in small area estimation problems. In *Proceedings of the Survey Research Methods Section, Joint Statistical Meeting 2005*. American Statistical Association.

- Datta GS dan Lahiri P. 2000. A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10:613–627.
- Datta GS, Rao JNK dan Smith DD. 2005. On measuring the Variability of Small Area Estimators Under a Basic Area Level Model. *Biometrika*, 92, 183-196.
- Deely JJ dan Lindley DV. 1991. Bayes empirical Bayes. *Journal of the American Statistical Association* 76, 833-841.
- Fabrizi E dan Trivisano C. 2009. Hierarchical Bayes Prediction in log-transformed Linear Mixed Model with Applications to Small Area Estimation. Prosiding pada SAE2009 Conference on Small Area Estimation, June 29 – July 01, 2009, Elche, Spain.
- Fay RE dan Herriot RA. 1979. Estimates of income for small places1 an application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74, 269- 277.
- Ghosh M dan Rao JNK. 1994. Small area estimation: an appraisal. *Statistical Sciences* 9, 55-93.
- Harville DA. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-340.
- Harville DA. 1991. Discussion on Robinson paper. That BLUP is a good thing: the estimation of random effects. *Statistical Science*. 6, 15-51.
- Hastie T dan Tibshirani R. 1990. *Generalized Additive Models*. London: Chapman and Hall.
- Henderson CR. 1953. Estimation of variance and covariance components. *Biometrics* 9, 226-252.
- Henderson CR. 1975. Best linear unbiased estimation and prediction under selection model. *Biometrics* 31, 423-447.
- Huber PJ. 1996. *Robust Statistical Procedures*. Philadelphia : Society for Industrial and Applied Mathematics.
- Holt D, Smith TMF dan Tomberlin. 1979. A model-based approach to estimation for small subgroups of a population. *Journal of the American Statistical Association* 74, 405-410.
- Jiang J. 1996. REML estimation : asymptotic behavior and related topics. *Annals of Statistics* 24, 255 – 286.
- Jiang J dan Lahiri P. 2006. Mixed Model Prediction and Small Area Estimation. *Test* 15, 1, 111–999.
- Jiang J, Lahiri P dan Wan SM. 2002. A Unified Jackknife Theory For Empirical Best Prediction With M-Estimation. *The Annals of Statistics* 30, 6, 1782–1810.
- Karlberg F. 2000. Survey Estimation for Highly Skewed Population in the Presence of Zeroes, *Journal of Official Statistics*, **16**, 229-41.

- Kass RE dan Steffey D. 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association* 84, 717-726.
- Kubokawa T. 2006. Linear Mixed Models and Small Area Estimation. *Japanese Journal Appl. Statistics* 35, 3, 139-161.
- Kurnia A dan Notodiputro KA. 2008. Generalized Additive Mixed Models for Small Area Estimation. *Mathematics Journal, Universiti Teknologi Malaysia*.
- Kurnia A, Notodiputro KA dan Ibrahim NA. 2007. A Nonparametric Approach in Small Area Estimation. *Proceeding at the ICCS-IX 2007, 12 - 14 December 2007. ISSOS - Universiti of Malaya, Syah Alam - Kuala Lumpur*.
- Kurnia A, Sartono B dan Wulandari R. 2007. Pengaruh Misspesifikasi Desain Survey pada Pendugaan Area Kecil dengan Pendekatan Generalized Regression. *Prosiding pada Seminar Nasional Matematika dan Pendidikan Matematika, UNY, Yogyakarta*.
- Lahiri P. 2009. Adjusted residual maximum likelihood method in the small-area estimation and related problems. *Prosiding pada SAE2009 Conference on Small Area Estimation, June 29 - July 01, 2009, Elche, Spain*.
- Lahiri P dan Rao JNK. 1995. Robust estimation of mean squared error of small area estimator. *Journal of the American Statistical Association*, 82, 758 - 766
- Laird NM dan Louis TA. 1987. Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association* 82, 739-750.
- Levy PS dan French DK. 1977. Synthetic Estimates of State Health characteristics Based on the Health Interview Survey, *Vital and Health Statistics*, 2.
- Li Y. 2006. Analysis of Complex Survey Data using Robust Model-Based and Model-Assisted Methods. *Disertasi pada the Faculty of the Graduate School of the University of Maryland, College Park*.
- Maiti T. 1998. Hierarchical Bayes estimation of mortality rates for disease mapping. *Journal Statistical Planning and Inference* 69, 339-348.
- Marker D. 1999. Organization of small area estimators using generalized linear regression framework. *Journal of Official Statistics*. 15, 1-24.
- McCullagh P dan Nelder JA. 1989. *Generalized Linear Models*. 2nd Ed. New York: Chapman dan Hall.

- Opsomer JD et al. 2008. Nonparametric Small Area Estimation using Penalized Spline Regression. *Journal Royal Statistics Society Series B*, 70, 265–286.
- Petrucci A, Pratesi M dan Salvati N. 2005. Geographic information in small area estimation: small area models and spatially correlated random area effects. *Statistics in Transition*, 7(3), 609-623.
- Petrucci A dan Salvati N. 2006. Small area estimation for spatial correlation in watershed erosion assessment. *Journal of Agricultural, Biological and Environmental Statistics*, 11(2), 169–182.
- Pfeffermann D. 1999. Small area estimation – big developments. Keynote Paper, Conference on Small Area Statistics, Riga, Latvia, August 1999.
- Pfeffermann D dan Glickman H. 2004. Mean Squared Error Approximation in Small Area Estimation by Use of Parametric and Nonparametric Bootstrap. *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Pfeffermann D dan Sverchkov M. 2003. On Small Area Estimation under Informative Sampling. *Technical Report*, 64-69.
- Prasad NGN dan Rao JNK. 1990. The Estimation of Mean Squared Errors of Small Area Estimators. *Journal of American Statistical Association*, 85, 163-171.
- Rao JNK. 1999. Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-186.
- Rao JNK. 2003. *Small Area Estimation*. New York : John Wiley and Sons.
- Rao JNK dan Yu M. 1994. Small Area Estimation by Combining Time Series and Cross-Sectional Data. *Canadian Journal of Statistics*, 22, 511-528.
- Ruppert R, Wand M dan Carroll R. 2003. *Semiparametric Regression*. Cambridge University Press.
- Russo C, Sabbatini M dan Salvatore R. 2005. General Linear Models in Small Area Estimation : an assessment in agricultural surveys. Paper presented in The Mexsai Conference. <www.siap.sagarpa.gob.mx/mexsai/trabajos/t44.pdf>, [29 April 2005].
- Särndal CE. 1984. Design-consistent versus model-dependent estimation for small domains. *Journal of the American Statistical Association* 79, 624-631.
- Särndal CE, Swensson B dan Wretman J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

- Schaible WL, Brock DB dan Schank GA. 1977. An emperical comparison of simple inflation, syntetic and composite estimators for small area statistics. Proceeding of the American Statistical.
- Slud EV dan Maiti T. 2006. Mean-squared Error Estimation in Transformde Fay-Herriot Model. Journal Royal Statistics B, 68, 239-257.
- Stone CJ. 1985. Additive Regression and Other Nonparametric Models. Annals of Statistics, 13, 689-705.
- Wan SM. 1999. Jackknife Methoda in Small Area Estimation and Related Problems. Disertasi pada The Graduate College at the University of Nebraska, Lincoln – Nebraska.
- You Y dan Rao JNK. 2000. Hierarchical Bayes estimation of small area means using multi-level models. Survey Methodology, 26 173-181.
- Zheng H dan Little RJA. 2004. Penalized spline nonparametric mixed models for inference about a finite population mean from two-stage samples. Survey Methodology 30, 209-218.

Lampiran 1.

Sifat statistik transformasi-balik logaritma dari respon yang menyebar lognormal

Berdasarkan model (5.1), total area ke-i (T_i) dapat diduga oleh

$$\hat{T}_i = \sum_{j \in s(i)} y_{ij} + \sum_{j \in r(i)} \hat{y}_{ij}.$$

Fungsi pembangkit momen untuk $\log(y_{ij})$ jika $\log(y_{ij}) \sim N(x_i^T \beta, \eta_i^2)$ adalah

$$E(e^{t \log y_{ij}}) = e^{x_i^T \beta t + (1/2)t^2 \eta_i^2}, \quad \eta_i^2 = \tau^2 + \sigma_i^2.$$

Untuk $t = 1$ diperoleh $E(y_{ij}) = e^{x_i^T \beta + (1/2)\eta_i^2}$ dan

untuk $t = 2$ diperoleh $E(y_{ij}^2) = e^{2(x_i^T \beta + \eta_i^2)}$,

sehingga dapat diperoleh juga :

$$\begin{aligned} Var(y_{ij}) &= e^{2(x_i^T \beta + \eta_i^2)} - e^{2x_i^T \beta + \eta_i^2} \\ &= e^{\eta_i^2} (e^{\eta_i^2} - 1) e^{2x_i^T \beta}. \end{aligned}$$

Dengan demikian jika diberikan $\hat{y}_{ij} = e^{x_i^T \hat{\beta}}$, maka

$$\begin{aligned} E(\hat{T}_i - T_i) &= E\left(\sum_{j \in r(i)} (\hat{y}_{ij} - y_{ij})\right) \\ &= \sum_{j \in r(i)} \left(e^{x_i^T \hat{\beta} + \frac{1}{2} x_i^T Var(\hat{\beta}) x_i} - e^{x_i^T \beta + \frac{1}{2} \eta_i^2}\right), \end{aligned}$$

dan persamaan tersebut tidak sama dengan nol.

Karena $\hat{\mu}_i^{naive} = f_i \bar{y}_i^s + (1 - f_i) \exp(x_i^T \hat{\beta})$ merupakan fungsi linier dari penduga total, maka $\hat{\mu}_i^{naive}$ juga akan berbias.

Lampiran 2.

Pemrograman GAMM untuk SAE

```

data gamm1;
input id E A X X2 theta y D;
cards;
1 0.21158      -0.83740      1.66235 2.76340 4.30771 4.51928 1
2 -2.67786     -0.98365      1.88545 3.55491 5.34872 2.67086 1
3 -0.36950      0.54167      1.93171 3.73149 7.13890 6.76940 1
4 -1.47021      0.94912      0.85711 0.73463 3.05107 1.58086 1
5 -0.09701     -0.87631      1.20418 1.45005 2.29877 2.20176 1
6 -0.44101     -0.56693      1.25368 1.57173 2.79066 2.34964 1
7 -0.73799     -0.29014      0.80664 0.65067 1.68587 0.94787 1
8 -0.58166     -0.05866      0.80272 0.64436 1.90788 1.32622 1
9 0.54416       2.21148      0.06729 0.00453 3.21827 3.76244 1
10 -1.32666     0.83795      0.66029 0.43598 2.49192 1.16527 1
11 -1.78347    -0.29599      1.78356 3.18108 5.47564 3.69216 1
12 -0.29568     0.15685      0.10466 0.01095 1.17328 0.87760 1
13 -1.09901     1.90999      0.68025 0.46274 3.60409 2.50508 1
14 -0.84734    -1.06838      0.96386 0.92903 1.32516 0.47782 1
15 0.31523     -0.94589      0.17180 0.02952 0.09838 0.41361 1
16 1.05108     -0.91765      1.79949 3.23817 4.93960 5.99068 1
17 1.16795     -1.03830      0.10197 0.01040 -0.0227 1.14525 1
18 0.34950      1.36723      0.86123 0.74171 3.47979 3.82929 1
19 1.54116     -0.23898      1.11211 1.23678 2.61619 4.15734 1
20 2.23720     -0.40237      1.98858 3.95447 6.52932 8.76652 1
21 0.35898      0.54489      0.86729 0.75219 2.67318 3.03215 1
22 -0.18869    -1.23445      0.45441 0.20649 0.07528 -0.1134 1
23 -0.34858    -1.52062      1.31913 1.74011 2.08955 1.74096 1
24 1.52234     -0.11838      0.79395 0.63035 1.82715 3.34949 1
25 0.77825      0.68879      1.48613 2.20860 5.00168 5.77993 1
26 -1.24000     0.73781      1.80012 3.24044 6.59847 5.35848 1
27 -0.02065     0.18565      1.62775 2.64956 5.15999 5.13934 1
28 -0.57056     0.07006      1.89235 3.58099 6.44154 5.87098 1
29 -0.29832    -0.79523      0.95187 0.90605 1.56385 1.26553 1
30 0.26860      1.03008      1.26375 1.59707 4.42568 4.69428 1
31 0.19245     -1.12590      1.66497 2.77211 4.03226 4.22471 1
32 0.48049     -0.09290      1.82938 3.34662 5.92702 6.40751 1
33 0.72682     -0.02095      0.72153 0.52061 1.75996 2.48679 1
34 -1.27803    -0.11302      0.35675 0.12727 1.07789 -0.2001 1
35 -0.27860     1.40665      1.67770 2.81467 6.62866 6.35006 1
;run;

*call in the macro additive_mixed_model.sas;
%include "d:\sss\gamm\additive_mixed_model.sas";
%sm1(data=gamm1,y=y,x= x x2,randomid=id,smoothvr=x x2);

*plot the scatter plot;
goption reset=symbol;
symbol1 c=red v=dot;
symbol2 c=black v=dot;
symbol3 c=red v=dot;
proc gplot data=yhat;
run;

proc iml;
M = 35;
load _all_;
use gamm1;
read all;
X = (D||X);
Y = (Y);
D = (D);

```



```

*** EBLUP of Theta;
beta_OLS = INV(X`X)*X`Y;
L = X*INV(X`X)*X`;
W = 1 - VECDIAG(L);
C = sum(W#D)/(M-1);
delta = SSQ(Y-X*beta_OLS)/(M-1);
A_wave = delta - C;
A_hat = MAX(0,A_wave);

sum_S = 0;
sum_T = 0;
sum_G = 0;
do I = 1 to M;
    S = X[I,]*X[I,]/(A_hat + D[I]);
    T = X[I,]*Y[I]/(A_hat + D[I]);
    G = (A_hat + D[I])*(A_hat + D[I]);
    sum_S = sum_S + S;
    sum_T = sum_T + T;
    sum_G = sum_G + G;
end;

beta_hat = INV(sum_S)*sum_T;
theta_hat = X*beta_hat + (A_hat/(A_hat + D))*(Y - X*beta_hat);
glA_hat = (A_hat*D)/(A_hat + D);

*** MSE of theta_hat : Jackknife Estimator;
sum1 = 0;
sum2 = 0;

do R = 1 to M;
    if R = 1 then sub_R = (2:M);
    if (1 < R & R < M) then sub_R = ((1:(R-1)) || ((R+1):M));
    if R = M then sub_R = (1:(M-1));
    X_u = X[sub_R,];
    Y_u = Y[sub_R];
    D_u = D[sub_R];

    betaOLSu = INV(X_u`X_u)*X_u`Y_u;
    L_u = X_u*INV(X_u`X_u)*X_u`;
    W_u = 1 - VECDIAG(L_u);
    C_u = sum(W_u#D_u)/(M-2);
    delta_u = SSQ(Y_u - X_u*betaOLSu)/(M-2);
    A_wave_u = delta_u - C_u;
    A_hat_u = MAX(0,A_wave_u);

    sum_S_u = 0;
    sum_T_u = 0;
    do K = 1 to (M-1);
        S_u = X_u[K,]*X_u[K,]/(A_hat_u + D_u[K]);
        T_u = X_u[K,]*Y_u[K]/(A_hat_u + D_u[K]);
        sum_S_u = sum_S_u + S_u;
        sum_T_u = sum_T_u + T_u;
    end;

    beta_hat_u = INV(sum_S_u)*sum_T_u;
    theta_hat_u = X_u*beta_hat_u + (A_hat_u/(A_hat_u + D_u))*(Y_u - X_u*beta_hat_u);
    delta2 = (theta_hat_u - theta_hat)*(theta_hat_u - theta_hat);
end;

```

```

sum2 = sum2 + delta2;

g1A_hat_u = (A_hat_u*D[R])/(A_hat_u + D[R]);
delta1 = g1A_hat_u - g1A_hat[R];
sum1 = sum1 + delta1;

end;

h1 = g1A_hat - ((m-1)/m) * sum1;
h2 = (m-1)/m * sum2;
MSE_J = h1 + h2;

*** MSE of theta_hat;
g1 = (A_hat*D)/(A_hat + D);
g2 = (D##2/(A_hat + D)##2)#VECDIAG(X*INV(sum_S)*X`);
g3 = (2*D##2/(M##2 * (A_hat + D)##3))*sum_G;

*** MSE of theta_hat : Prasad & Rao Estimator;
MSE_P = g1 + g2 + 2*g3;
MSE_N = g1 + g2;

print Y theta_hat MSE_N MSE_P MSE_J;
quit;

```

Catatan: macro additive_mixed_model.sas oleh Long Ngo

Lampiran 3.

Pemrograman P-Spline untuk SAE

```

proc glimmix data=gamm1;
class id;
model y = x;
random x / type = rsmooth knotmethod=kdtree(bucket=30 treeinfo
knotinfo);
random id;
output out=a pred(blup)=mu stderr=stderr;
run;

proc print data=a; run;

proc glimmix data=gamm1;
class id;
model y = x ;
random id;
output out=a pred(blup)=mu stderr=stderr;
run;

proc print data=a; run;

proc iml;
M = 35;
load _all_;
use a;
read all;
X = (D||X);
Y = (Y);
D = (D);
A_hat = 1.4360;
* X*beta_hat = mu;
mu = mu;
gamma = A_hat / (A_hat + D);
theta_hat = mu + gamma#(Y - mu);

sum_S = 0;
sum_T = 0;
sum_G = 0;
do I = 1 to M;
    S = X[I,]`*X[I,]/(A_hat + D[I]);
    T = X[I,]`*Y[I]/(A_hat + D[I]);
    G = (A_hat + D[I])*(A_hat + D[I]);
    sum_S = sum_S + S;
    sum_T = sum_T + T;
    sum_G = sum_G + G;
end;

*** MSE of theta_hat;
g1 = (A_hat*D)/(A_hat + D);
g2 = (D##2/(A_hat + D)##2)#VECDIAG(X*INV(sum_S)*X`);
g3 = (2*D##2/(M##2 *(A_hat + D)##3))*sum_G;

*** MSE of theta_hat : Prasad & Rao Estimator;
MSE_P = g1 + g2 + 2*g3;
MSE_N = g1 + g2;
print Y theta_hat MSE_N MSE_P;

```

```

quit;

proc glimmix data=iccsix.datakotabogor2;
class desa;
model y = x1 x2 x3 x4;
random x1 x2 x3 x4/ type = rsmooth knotmethod=kdtree(bucket=30
treeinfo knotinfo);
random desa;
output out=a pred(blup)=mu stderr=stderr;
run;

proc print data=a; run;

proc glimmix data=iccsix.datakotabogor2;
class desa;
model y = x1 x2 x3 x4;
random desa;
output out=a pred(blup)=mu stderr=stderr;
run;

proc print data=a; run;

proc iml;
M = 34;
load _all_;
use a;
read all;
X = (x1||x2||x3||x4);
Y = (Y);
D = (Di);
A_hat = 25.5378;
* X*beta_hat = mu;
mu = mu;
gamma = A_hat / (A_hat + D);
theta_hat = mu + gamma#(Y - mu);

sum_S = 0;
sum_T = 0;
sum_G = 0;
do I = 1 to M;
    S = X[I,]`*X[I,]/(A_hat + D[I]);
    T = X[I,]`*Y[I]/(A_hat + D[I]);
    G = (A_hat + D[I])*(A_hat + D[I]);
    sum_S = sum_S + S;
    sum_T = sum_T + T;
    sum_G = sum_G + G;
end;

*** MSE of theta_hat;
g1 = (A_hat*D)/(A_hat + D);
g2 = (D##2/(A_hat + D)##2)#VECDIAG(X*INV(sum_S)*X`);
g3 = (2*D##2/(M##2 * (A_hat + D)##3))*sum_G;

*** MSE of theta_hat : Prasad & Rao Estimator;
MSE_P = g1 + g2 + 2*g3;
MSE_N = g1 + g2;
print Y theta_hat MSE_N MSE_P;
quit;

```

Lampiran 4.

Penduga pengeluaran per kapita dalam rupiah di Kabupaten dan Kota Bogor serta penduga RMSE relatifnya menggunakan metode penduga langsung dan EBLUP berdasarkan data Susenas tahun 2005

Desa/Kelurahan	Penduga parameter		Penduga RRMSE (%)	
	Penduga Langsung	EBLUP	Penduga Langsung	EBLUP
3201010002 Bantar Karet	527,541	500,845	12.26599	12.07658
3201010010 Parakan Muncang	364,312	377,754	9.96789	15.98887
3201020005 Karacak	308,603	311,835	6.4734	19.32912
3201020006 Barengkok	391,584	385,196	8.53178	16.1477
3201030005 Gunung Sari	355,579	362,994	8.04356	16.5776
3201030006 Gunung Bunder 2	316,928	321,964	7.72431	18.66518
3201040004 Ciaruten Udik	276,806	292,381	8.80755	20.5069
3201040005 Cibatok 1	359,929	379,137	8.07166	15.77677
3201040013 Leuweung Kolot	339,212	360,805	9.05583	16.57621
3201050013 Cihideung Udik	599,059	568,404	7.7453	10.64869
3201050018 Benteng	624,363	612,086	10.78972	9.83048
3201060003 Sukadamai	415,175	407,620	9.7023	14.74386
3201060009 Babakan	474,440	483,954	7.29715	12.38375
3201070012 Ciomas	358,673	387,599	13.46225	15.45155
3201070018 Ciomas Rahayu	995,648	925,746	15.71241	6.47451
3201071006 Tamansari	381,576	394,597	6.54225	15.19513
3201080011 Cijeruk	432,161	429,479	10.57447	13.93243
3201080018 Sukaharja	584,697	557,837	9.02741	10.82295
3201090005 Muara Jaya	660,470	631,634	7.59814	9.5547
3201100001 Cileungsi	451,993	447,529	15.58217	13.47039
3201100005 Jambu Luwuk	466,020	459,338	10.77943	13.15481
3201100013 Pandansari	535,499	535,559	9.25694	11.18714
3201110005 Batu Layang	534,945	531,501	15.97666	11.39046
3201120005 Sukamanah	345,159	352,030	15.55117	17.22728
3201120007 Sukamahi	380,354	398,559	9.32956	15.10892
3201130009 Cijujung	712,927	690,285	19.04966	8.65924
3201130012 Cilebut Timur	846,890	801,307	17.63483	7.46193
3201140001 Cijayanti	485,611	485,846	14.90714	12.41115
3201140009 Sentul	538,960	542,738	14.67999	10.99799
3201150001 Sukawangi	388,928	379,225	5.41922	15.9231
3201150008 Pabuaran	365,811	364,134	8.23919	16.64133
3201160018 Kuta Mekar	421,814	402,084	13.02386	14.9431
3201170007 Balekambang	427,522	422,667	8.46966	14.33354
3201170011 Sukamaju	385,400	408,459	5.90994	14.67534
3201180013 Cipeucang	466,724	465,139	5.70487	12.95255
3201180018 Cileungsi	537,866	546,392	6.79499	10.96928
3201180020 Pasir Angin	454,981	463,734	10.22734	12.97138
3201181002 Lulut	327,296	354,457	4.56215	16.94293

Desa/Kelurahan	Penduga parameter		Penduga RRMSE (%)	
	Penduga Langsung	EBLUP	Penduga Langsung	EBLUP
3201181008 Bojong	426,047	429,823	8.08505	14.123
3201190003 Tlajung Udik	817,634	768,851	12.8325	7.85282
3201190005 Cicadas	429,760	439,553	15.88976	13.6641
3201190009 Ciangsana	499,050	514,409	7.52461	11.64941
3201200004 Pasir Mukti	332,931	348,999	11.79967	17.30342
3201200008 Karang Asem Barat	498,906	508,409	6.90104	11.74282
3201200011 Gunung Sari	370,439	393,896	12.91614	15.16398
3201200014 Puspasari	668,831	654,457	15.54133	9.1268
3201210004 Cibinong	430,801	456,702	11.6991	13.43669
3201210006 Sukahati	544,741	551,732	9.78471	10.83009
3201210010 Pabuaran	678,385	670,551	6.72315	6.5566
3201210012 Ciriung	547,820	553,716	12.50421	10.78612
3201220005 Bojonggede	760,264	733,413	15.90504	8.14571
3201220012 Bojong Baru	681,096	663,259	13.47378	9.01501
3201220014 Pabuaran	441,219	466,122	9.94534	12.81585
3201230013 Tegal	508,134	491,734	21.88045	12.31027
3201230015 Jampang	455,648	469,392	9.74147	12.80279
3201240011 Pamager Sari	338,487	367,127	7.92437	16.35609
3201240017 Cogreg	340,900	364,457	6.133	16.86808
3201241004 Putat Nutug	302,146	310,499	7.57687	19.39854
3201250006 Curug	551,180	535,505	11.74308	11.20941
3201250007 Rawakalong	550,354	538,797	6.86335	11.18024
3201260005 Cibodas	459,744	452,146	11.48962	13.30344
3201270009 Sukamaju	380,082	392,560	5.5407	15.41256
3201270017 Banyu Asih	286,617	290,480	6.7273	20.90647
3201271003 Kiarapandak	284,602	287,356	14.71873	20.97563
3201280005 Sipak	243,369	258,233	7.85186	23.58304
3201280008 Curug	405,641	378,193	11.60047	16.29734
3201290004 Babakan	266,055	289,850	6.72481	20.93133
3201300005 Pingku	350,055	358,721	12.88776	17.70486
3271010002 Pamoyanan	352,221	381,537	7.93897	15.66903
3271010004 Genteng	432,471	451,295	13.29799	13.2458
3271010008 Harjasari	392,098	415,808	3.71692	14.39068
3271010011 Cipaku	443,505	466,935	11.47549	12.79361
3271010013 Batutulis	659,507	647,941	16.08239	9.21899
3271010015 Empang	577,624	569,521	6.60054	10.60834
3271010016 Cikaret	514,436	522,590	9.35147	11.42778
3271020002 Sindangrasa	605,858	601,842	8.07064	9.91955
3271020004 Katulampa	318,070	356,165	6.72544	16.76575
3271020005 Baranangsiang	505,012	518,530	12.02307	11.5217
3271020006 Sukasari	486,260	501,006	19.76298	11.91621
3271030001 Bantarjati	505,012	520,623	12.02307	11.47634

Desa/Kelurahan	Penduga parameter		Penduga RRMSE (%)	
	Penduga Langsung	EBLUP	Penduga Langsung	EBLUP
3271030002 Tegal Gundil	491,444	508,560	10.07117	11.7465
3271030003 Tanahbaru	650,851	636,670	11.5135	9.38218
3271030004 Cimahpar	401,998	407,705	10.07056	15.10014
3271030006 Cibuluh	553,789	558,692	10.27629	10.68491
3271030007 Kedunghalang	683,710	665,346	16.3884	9.23866
3271030008 Ciparigi	405,031	434,362	8.37388	13.74538
3271040003 Babakanpasar	614,331	608,906	19.77389	9.80696
3271040004 Tegallega	462,562	479,449	9.19648	12.46389
3271040007 Pabaton	1,567,773	1,409,994	10.53489	4.23833
3271040010 Kebonkelapa	982,102	917,591	11.38774	6.50912
3271050001 Pasirmulya	546,893	547,214	12.73436	10.97735
3271050003 Pasirjaya	408,920	433,822	8.45118	13.80857
3271050004 Gunungbatu	556,493	561,536	4.75927	10.63511
3271050006 Menteng	507,573	520,152	10.29911	11.47894
3271050008 Cilendek Barat	594,538	590,408	5.17781	10.12778
3271050009 Sindangbarang	750,667	721,032	15.78157	8.28092
3271050012 Situgede	478,881	476,170	6.78617	12.71934
3271050014 Semplak	616,821	609,329	8.54101	9.85577
3271060001 Kedungwaringin	464,557	481,683	11.7115	12.39672
3271060002 Kedungjaya	560,473	564,837	9.95505	10.5739
3271060003 Kebonpedes	449,201	472,276	7.37147	12.64417
3271060005 Kedungbadak	697,475	683,668	12.83002	6.44023
3271060008 Cibadak	566,478	569,426	14.48496	10.48858
3271060009 Kayumanis	505,628	506,811	10.09923	11.92309
3271060011 Kencana	386,722	411,912	10.3622	14.50523

Lampiran 5.

Pemrograman SAS untuk simulasi EBP

```

data populasi;
do i = 1 to 315340;
  if i > 313186 then
    do area = 105; x1 = 1; x2 = 5; x3 = 145; x4 = 70.334;
      x5 = 13.383; x6 = 21.9626; x7 = 10; x8 = 45; end;
    else if i > 310914 then
      do area = 104; x1 = 1; x2 = 2; x3 = 200; x4 = 77.817;
        x5 = 35.8715; x6 = 37.8601; x7 = 8; x8 = 80; end;
    else if i > 307101 then
      .
      .
      .
    else do area = 1; x1 = 0; x2 = 34; x3 = 1020; x4 = 33.277;
      x5 = 12.9481; x6 = 13.9167; x7 = 58; x8 = 150; end;

  xb = 12.9744 + 0.08488*x1 - 0.00292*x2 - 0.00009*x3 +
    0.001359*x4 - 0.00048*x5 - 0.00025*x6 + 0.000502*x7 -
    0.00097*x8;
  v = rannor(0)*sqrt(0.04849);
  e = rannor(0)*sqrt(0.1573);
  logy = xb + v + e;
  y = exp(logy);
  output;
end;

proc tabulate data = populasi out=a;
  class area ;
  var y;
  table area, n (y)*mean;
run;

data a (keep = area ypop Npop);
  set a;
  ypop = y_mean;
  Npop = n;
run;

/* ===== SAMPLE ===== */
proc surveyselect data=populasi
  method=srs n=(16 16 16 15 16 16 16 16 16 16 16 16 16 16 16
16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16
16 16 16 16 16 16 16 16 16 16 15 16 32 16 16 16 16 16 16 15
16 16 16 16 16 16 16 16 16 16 15 16 16 16 16 16 16 16 16 16
16 16 16 16 16 15 16 16 16 16 16 16 16 16 16 16 16 16 16 16
16 16 32 16 16 16)
  out=Sample;
  strata area;
run;

=====; * Direct Estimate; =====;
proc tabulate data = sample out=b;
  class area ;
  var logy y;
  table area, n (logy y)*(mean stderr);
run;

```



```

data b (keep = n area Direct Stderr RRMSE Directlog Stderrlog
RRMSElog sumy);
set b;
Directlog = logy_mean;
Stderrlog = logy_stderr;
RRMSElog = stderrlog/directlog*100;
Direct = y_mean;
stderr = y_stderr;
RRMSE = stderr/direct*100;
sumy = n*y_mean;
run;

proc mixed data=sample noclprint covtest;
class area;
model y = x1 x2 x3 x4 x5 x6 x7 x8 / s outpm = d ddfm =
kenwardroger;
* outpm = d => without random effect;
random intercept /sub = area;
run;

proc tabulate data = d out=c ;
class area ;
var pred x1 x2 x3 x4 x5 x6 x7 x8;
table area, n mean*(x1 x2 x3 x4 x5 x6 x7 x8 pred);
run;

data c (keep = area n x1 x2 x3 x4 x5 x6 x7 x8 xb Di);
set c;
x0 = 1;
x1 = x1_mean;
x2 = x2_mean;
x3 = x3_mean;
x4 = x4_mean;
x5 = x5_mean;
x6 = x6_mean;
x7 = x7_mean;
x8 = x8_mean;
xb = pred_mean;
Di = 6.738E10 /n; *diperoleh dari proc mixed;
run;

*====*; * Management data for Indirect Estimate; *====*;
proc mixed data=sample noclprint covtest;
class area;
model logy = x1 x2 x3 x4 x5 x6 x7 x8 / s outpm = d ddfm =
kenwardroger;
* outpm = d => without random effect;
random intercept /sub = area;
run;

proc tabulate data = d out=e ;
class area ;
var pred;
table area, n mean*(pred);
run;

data e (keep = area n Dilog xbldg);
set e;

```

```

xblog = pred_mean;
Dilog = 0.1573/n; *from proc mixed;
run;

*=====; * Merge data direct + data dasar ; *=====;
proc sort data=a; by area; run;
proc sort data=b; by area; run;
proc sort data=c; by area; run;
proc sort data=e; by area; run;

data baru;
merge a b c e;
by area;
run;

*** MODEL :  $Y = XB + U + E$ , where  $U \sim N(0, A)$  dan  $E \sim N(0, D)$ ;
*** PARAMETER BETHA, DESAIN MATRIK X DAN RAGAM D;
proc iml;
M = 105;
load _all_;
use baru;
read all;
X = (x0||x1||x2||x3||x4||x5||x6||x7||x8);

D = (Di);
A_hat = 2.16E10; *from proc mixed;
V = D + A_hat;

Dlog = (Dilog);
Alog = 0.04849; *from proc mixed;
*SDD = ;
*SDA = ;
Vlog = Dlog + Alog;

/* ===== EBLUP ===== */
sum_S = 0; sum_G = 0;
do I = 1 to M;
    S = X[I,]`*X[I,]/(A_hat + D[I]);
    G = (A_hat + D[I])*(A_hat + D[I]);
    sum_S = sum_S + S;
    sum_G = sum_G + G;
end;

theta_hat = XB + (A_hat / (A_hat + D))#(Direct - XB);

*** MSE of theta_hat;
g1 = (A_hat*D)/(A_hat + D);
g2 = (D##2/(A_hat + D)##2)#VECDIAG(X*INV(sum_S)*X`);
g3 = (2*D##2/(M##2*(A_hat + D)##3))*sum_G;

*** MSE of theta_hat : Prasad & Rao Estimator;
MSEP = g1 + g2 + 2*g3;
RRMSEP = sqrt(MSEP)/theta_hat*100;

gamma = (A_hat)/(A_hat + D);

/* ===== Log-Scale EBLUP ===== */
sum_Slog = 0; sum_Glog = 0;
do J = 1 to M;

```

```

    Slog = X[J,]`*X[J,]/(Alog + Dlog[J]);
    Glog = (Alog + Dlog[J])*(Alog + Dlog[J]);
    sum_Slog = sum_Slog + Slog;
    sum_Glog = sum_Glog + Glog;
end;

thetahat = XBlog + (Alog / (Alog+ Dlog))*(Directlog - XBlog);

*** MSE of theta_hat;
g1log = (Alog*Dlog)/(Alog + Dlog);
g2log = (Dlog##2/(Alog + Dlog)##2)#VECDIAG(X*INV(sum_Slog)*X`);
g3log = (2*Dlog##2/(M##2 *(Alog + Dlog)##3))*sum_Glog;

*** MSE of theta_hat : Prasad & Rao Estimator;
MSEPR = g1log + g2log + 2*g3log;

theta_hatlog = exp(thetahat + 0.5*MSEPR);
MSEPlog = exp(MSEPR)#(exp(MSEPR)-1)#exp(2*thetahat);

RRMSEPlog = sqrt(MSEPlog)/theta_hatlog*100;

/* ===== Log-Normal EBP ===== */

Xb2 = (xblog + Vlog/2);
sq = Dlog + Alog;

TB = sumy + (Npop - n)#exp(xb2 + sq/2);
MeanTB = TB / Npop;

Finite1 = exp(Vlog)#(exp(Vlog) - 1)#exp(Xb2);

Finite2 = (2*(Alog##2)/(m-1) + 2*(Dlog##2)/(n)) # (0.5 *
exp(Xb2))##2;

Finite3 = 0.25 * exp(Alog) # ((Npop-n) # exp(2*xblog+Dlog)) #
(Alog**2 * 0.005560**2 + Dlog * 0.008414**2
+ 2*Alog*Dlog*1.91732030577329E-06);

MSETB = finite1 + finite2 + 2*finite3;
RRMSETB = 1.5#sqrt(MSETB)/meanTB * 100;

deltdirect = direct-ypop;
delteBLUP = theta_hat-ypop;
deltlog = theta_hatlog-ypop;
deltlognor = meantb-ypop;

MSE = stderr#stderr;
print area deltdirect delteBLUP deltlog deltlognor Ypop MSE MSEP
MSEPlog MSETB;

quit;

```

Lampiran 6.

Ringkasan kajian simulasi untuk model lognormal EBP (bias relatif dan RRMSE)

Bias Relatif

AREA	DIRECT	EBLUP	LOG	EBP
1	-0.01472	-0.01617	-0.09943	-0.08689
2	0.01751	0.00351	-0.09487	-0.08495
3	-0.00659	-0.00256	-0.08790	-0.07414
4	0.00534	-0.00720	-0.09684	-0.08261
5	0.00797	0.00732	-0.09954	-0.08840
6	-0.01997	-0.01614	-0.10084	-0.08436
7	-0.01748	0.00098	-0.09401	-0.07743
8	0.01149	0.01620	-0.09578	-0.08559
9	0.00261	0.00832	-0.09732	-0.08459
10	-0.00198	-0.00413	-0.09502	-0.08249
11	-0.00635	-0.00669	-0.10157	-0.09022
12	0.00811	0.00523	-0.08837	-0.07648
13	-0.01576	0.00298	-0.09960	-0.08549
14	-0.00348	0.00193	-0.09506	-0.08476
15	-0.01937	-0.01244	-0.10084	-0.08727
16	0.00108	0.02034	-0.09077	-0.07868
17	0.00721	0.01366	-0.09425	-0.08238
18	0.02312	0.02710	-0.08511	-0.07269
19	-0.01148	-0.00191	-0.09571	-0.08199
20	0.00104	0.00675	-0.09035	-0.07710
21	0.00424	0.00856	-0.09490	-0.08291
22	-0.00056	-0.00472	-0.09643	-0.08312
23	-0.00740	0.00988	-0.10798	-0.09458
24	-0.00199	-0.00124	-0.09825	-0.08773
25	-0.00477	0.00763	-0.09517	-0.08259
26	0.00714	-0.00389	-0.09766	-0.08481
27	-0.00063	0.00333	-0.09776	-0.08582
28	0.01226	0.02513	-0.09059	-0.07827
29	-0.01706	-0.00019	-0.09747	-0.08501
30	-0.00635	-0.00786	-0.09338	-0.08191
31	0.03102	0.00896	-0.08335	-0.07341
32	0.00431	-0.02051	-0.09531	-0.07965
33	0.00863	0.00487	-0.09580	-0.08505
34	-0.00859	0.00038	-0.09975	-0.08618
35	-0.01362	0.00508	-0.09568	-0.08319
36	-0.00534	-0.01189	-0.10000	-0.08861
37	0.00504	0.00430	-0.10017	-0.08902

Relative Root Mean Square Error - RRMSE

AREA	DIRECT	EBLUP	LOG	EBP
1	11.30069	9.24029	5.53198	3.24607
2	11.60890	7.64156	5.55943	3.23774
3	10.87785	9.02519	5.35514	3.29863
4	12.08315	9.27941	5.77457	3.29396
5	11.54610	7.66077	5.12385	3.13154
6	11.48650	8.45756	4.93409	2.65232
7	11.60553	7.51295	4.67606	2.35612
8	11.52953	5.70571	4.31206	2.82786
9	11.88804	5.88092	4.35298	2.40850
10	11.25978	7.89404	5.54688	3.45923
11	11.22201	5.85695	5.08589	3.17353
12	11.05272	7.72444	4.89759	2.59140
13	11.11922	5.61151	4.63685	3.76841
14	11.19735	5.30214	4.35857	3.28270
15	11.25968	5.29296	4.57382	3.34086
16	11.52072	6.42094	4.72413	3.31180
17	11.29127	6.67991	4.45930	2.37959
18	11.38940	7.67346	5.55499	3.55612
19	11.17122	6.66850	5.37107	2.22034
20	11.06858	7.65320	5.42026	2.57519
21	11.15084	7.58744	5.49261	2.35424
22	11.57533	5.34985	4.40577	2.79973
23	11.59121	7.11834	5.90105	2.83913
24	11.24894	8.94652	6.02551	3.02540
25	11.34585	6.62887	5.14753	3.06219
26	11.84446	4.91504	4.27216	4.80931
27	11.38783	4.99653	4.34521	3.51788
28	11.21754	5.65433	4.22967	3.39724
29	11.31878	4.81227	4.06937	3.42193
30	11.33803	9.39590	5.50086	2.95486
31	11.52027	9.08203	5.99473	3.14697
32	11.45878	9.18656	4.97745	2.24090
33	11.35184	8.65715	5.94697	2.89804
34	10.82424	5.52017	4.43611	3.00637
35	11.09073	6.88852	5.23577	3.18514
36	10.97529	5.16120	4.68379	4.21001
37	11.52317	6.17480	4.88201	5.71982

Bias Relatif

AREA	DIRECT	EBLUP	LOG	EBP
38	-0.01008	0.02015	-0.08747	-0.07554
39	0.00296	0.02471	-0.09246	-0.07866
40	-0.00875	0.01140	-0.09470	-0.08264
41	0.00420	0.01182	-0.09448	-0.08254
42	-0.00098	-0.00628	-0.09851	-0.08673
43	0.01427	0.01763	-0.08881	-0.07708
44	0.00472	0.00470	-0.09461	-0.08300
45	0.00714	0.01203	-0.09151	-0.08126
46	-0.00150	-0.00431	-0.09754	-0.08407
47	0.01146	-0.00694	-0.09389	-0.08203
48	-0.00849	-0.00718	-0.09203	-0.08058
49	0.01513	0.00350	-0.09394	-0.08982
50	0.01206	0.00177	-0.09470	-0.08361
51	0.00669	-0.00288	-0.09534	-0.08215
52	0.03040	0.00665	-0.09147	-0.08030
53	-0.01765	-0.00819	-0.09982	-0.08370
54	0.01479	0.02216	-0.08660	-0.07505
55	-0.02346	0.00049	-0.09906	-0.08447
56	0.00120	0.01076	-0.09830	-0.08495
57	0.00669	0.02332	-0.09532	-0.08112
58	0.01301	0.00155	-0.08728	-0.07518
59	0.00529	0.01935	-0.09182	-0.07860
60	0.01037	0.02081	-0.09323	-0.08155
61	0.00623	0.01193	-0.09168	-0.07734
62	0.00189	-0.00028	-0.09581	-0.08269
63	-0.00138	-0.02076	-0.09388	-0.08109
64	-0.00089	-0.03733	-0.09659	-0.08532
65	-0.01397	-0.01968	-0.10691	-0.09457
66	-0.00092	-0.07128	-0.09794	-0.08779
67	0.00187	0.01325	-0.08870	-0.07646
68	0.00298	0.00905	-0.09477	-0.08272
69	0.00920	0.00829	-0.09557	-0.08251
70	0.01780	0.00969	-0.09539	-0.08151
71	-0.00024	0.00303	-0.09422	-0.08098
72	-0.00770	-0.00786	-0.09402	-0.08225
73	-0.00297	-0.00468	-0.09461	-0.08196
74	-0.01182	0.00482	-0.09705	-0.08302
75	0.00702	0.00581	-0.09156	-0.08071
76	-0.00783	-0.00371	-0.09638	-0.08237
77	-0.02969	-0.00588	-0.09744	-0.08249

Relative Root Mean Square Error - RRMSE

AREA	DIRECT	EBLUP	LOG	EBP
38	10.82874	6.18044	4.87801	3.54234
39	11.28530	7.38881	5.75141	2.03453
40	11.21985	6.00437	4.98016	6.67315
41	11.34379	6.21026	4.86521	5.39884
42	11.25818	5.07451	4.65568	4.81681
43	11.53601	6.66933	4.61039	2.62024
44	11.05978	4.85923	4.06849	4.31607
45	11.46488	5.24108	4.14034	3.26687
46	11.17966	4.76416	4.17818	3.65771
47	11.31124	4.83019	4.13152	3.99977
48	11.34456	4.75208	4.21652	4.09213
49	8.29932	4.45308	4.40748	4.72856
50	11.73739	4.67365	4.16028	4.67961
51	11.27391	4.61902	4.20665	5.34487
52	11.31409	4.80354	4.23906	2.95917
53	11.53885	4.66761	4.19308	4.84574
54	11.45461	8.12984	5.88116	3.38349
55	11.56349	5.67359	4.96122	3.04292
56	11.20840	5.80060	4.68597	3.08663
57	11.64648	5.70894	4.27828	3.08006
58	11.56894	8.26785	5.10662	2.51622
59	11.82814	6.71554	4.88150	3.25194
60	11.34897	6.20312	4.85107	2.62957
61	11.19994	7.40934	5.17105	2.72717
62	11.44369	7.50756	5.59334	3.01066
63	11.09438	10.84334	6.20108	1.89040
64	11.12192	10.07852	5.40511	3.11400
65	11.70982	10.82827	6.38506	3.11190
66	11.63124	15.81388	7.48385	2.87662
67	11.08202	8.53533	6.07858	2.97811
68	11.87536	10.71472	7.72125	2.53066
69	11.41511	5.28565	4.29872	3.29127
70	11.83103	5.02286	4.30755	2.63693
71	11.70035	5.18058	4.36729	3.45535
72	11.23310	4.75086	4.18747	3.48374
73	11.61341	4.68097	4.17753	3.50788
74	11.27675	5.73303	4.73585	4.34161
75	11.51567	4.79454	4.13129	4.12384
76	11.33973	4.68977	4.09836	3.12781
77	11.09021	5.02595	4.09073	4.55288

Bias Relatif

AREA	DIRECT	EBLUP	LOG	EBP
78	0.00310	-0.00341	-0.09471	-0.08205
79	-0.00496	0.00036	-0.09414	-0.08177
80	0.00902	-0.00619	-0.09480	-0.08211
81	0.00276	-0.00398	-0.09756	-0.08213
82	-0.00260	-0.00431	-0.09309	-0.08132
83	-0.01314	0.02942	-0.09902	-0.08780
84	-0.00187	-0.00901	-0.09564	-0.08192
85	-0.01038	-0.00982	-0.09404	-0.08087
86	-0.00817	-0.00236	-0.09608	-0.08262
87	0.00277	-0.00056	-0.09299	-0.08202
88	-0.00069	-0.00062	-0.09803	-0.08295
89	0.00022	-0.01085	-0.09655	-0.08121
90	-0.00389	-0.00916	-0.09587	-0.08221
91	0.01510	0.00426	-0.08964	-0.07866
92	-0.00456	-0.00097	-0.09445	-0.08041
93	0.01064	-0.00690	-0.09406	-0.08203
94	0.00322	-0.00357	-0.09329	-0.08182
95	-0.00362	-0.00065	-0.09354	-0.08098
96	0.00762	0.00162	-0.09519	-0.08219
97	0.00748	0.00455	-0.09356	-0.08134
98	-0.00704	-0.00915	-0.09214	-0.07870
99	0.00571	-0.00412	-0.09561	-0.08261
100	0.00346	0.00150	-0.09484	-0.08195
101	0.00603	-0.00821	-0.09268	-0.08173
102	0.01155	0.00814	-0.09304	-0.08635
103	0.01283	0.00117	-0.09459	-0.08299
104	0.01787	0.00619	-0.08783	-0.07702
105	0.01265	0.00332	-0.09483	-0.08239

Relative Root Mean Square Error - RRMSE

AREA	DIRECT	EBLUP	LOG	EBP
78	11.22331	4.70627	4.21182	5.18062
79	10.89414	4.68002	4.06025	3.52251
80	11.81864	4.61030	4.22367	4.75573
81	11.46387	4.61340	4.16970	5.13787
82	11.38446	4.95826	4.17364	4.38742
83	11.77316	9.71233	7.44637	3.68762
84	11.10360	4.72141	4.06858	4.56952
85	11.67405	4.98628	4.36948	4.59020
86	10.88699	4.65209	4.10617	4.56913
87	11.53193	4.66390	4.10946	3.36308
88	11.47628	4.82084	4.08180	4.39396
89	11.35509	4.67160	4.25272	1.99247
90	11.52911	4.68662	4.07620	3.49772
91	11.20688	5.25928	4.41939	2.06665
92	11.48336	5.22370	4.58527	4.31732
93	11.28981	4.68869	4.13495	4.38819
94	11.22914	4.63763	4.09398	3.86732
95	10.98935	4.94323	4.34938	3.82169
96	11.24787	4.76037	4.11964	3.59672
97	11.89297	7.23987	5.64220	2.85242
98	11.08555	5.45744	4.95881	3.33613
99	11.69437	4.81384	4.07347	4.41307
100	11.14467	4.61910	4.15793	3.45136
101	11.21595	4.66555	4.12356	4.65587
102	8.31443	5.08978	4.70683	3.64069
103	11.17401	4.67936	4.16190	4.11828
104	11.53006	6.57278	5.67751	3.17680
105	11.75939	5.05224	4.16695	3.09289

Lampiran 7.

Ringkasan kajian simulasi untuk model lognormal EBP (bias relatif MSE dan tingkat ketercakupan)

Bias Relatif MSE

AREA	DIRECT	EBLUP	LOG	EBP
1	-0.03351	-0.02452	-0.02596	-0.03009
2	0.22127	-0.02198	-0.02516	-0.03005
3	-0.09408	-0.02139	-0.02411	-0.02869
4	0.11937	-0.01815	-0.02497	-0.02970
5	-0.00965	-0.02338	-0.02666	-0.03026
6	-0.07772	-0.02405	-0.02730	-0.03095
7	0.06708	-0.01948	-0.02683	-0.03103
8	0.15535	-0.02371	-0.02799	-0.03051
9	0.33853	-0.02066	-0.02773	-0.03120
10	0.10666	-0.02605	-0.02533	-0.02937
11	-0.13133	-0.02539	-0.02717	-0.03036
12	-0.17829	-0.02245	-0.02556	-0.03043
13	0.11470	-0.02335	-0.02771	-0.02841
14	-0.15598	-0.02584	-0.02790	-0.02956
15	0.15907	-0.02277	-0.02792	-0.02963
16	0.22821	-0.02404	-0.02640	-0.02902
17	0.02682	-0.02181	-0.02737	-0.03125
18	-0.10955	-0.02529	-0.02329	-0.02805
19	-0.03545	-0.02493	-0.02560	-0.03160
20	0.04336	-0.02313	-0.02435	-0.03070
21	-0.05497	-0.02129	-0.02540	-0.03147
22	0.26962	-0.02000	-0.02784	-0.03046
23	-0.16411	-0.02414	-0.02624	-0.03117
24	-0.18904	-0.02260	-0.02465	-0.03066
25	-0.11082	-0.02615	-0.02616	-0.03019
26	0.06894	-0.02366	-0.02797	-0.02480
27	-0.05673	-0.02426	-0.02787	-0.02891
28	-0.14019	-0.02828	-0.02806	-0.02926
29	-0.02004	-0.02093	-0.02832	-0.02900
30	0.09811	-0.02279	-0.02509	-0.03037
31	0.09914	-0.02241	-0.02096	-0.02927
32	-0.07233	-0.02148	-0.02669	-0.03147
33	-0.21618	-0.02467	-0.02442	-0.03073
34	-0.14650	-0.02297	-0.02804	-0.03019
35	-0.11326	-0.02267	-0.02589	-0.02975
36	0.04444	-0.01978	-0.02749	-0.02754
37	0.20785	-0.02594	-0.02716	-0.02308

Tingkat ketercakupan (coverage rate)

AREA	DIRECT	EBLUP	LOG	EBP
1	91.6667	95.1667	96.6667	93.3333
2	92.5000	96.0000	95.8333	94.1667
3	90.0000	95.1667	97.5000	97.5000
4	95.0000	95.1667	99.1667	96.6667
5	92.5000	94.3333	96.6667	97.5000
6	90.8333	95.1667	94.1667	87.5000
7	89.1667	96.0000	95.8333	88.3333
8	90.0000	93.5000	86.6667	97.5000
9	94.1667	96.0000	90.8333	89.1667
10	93.3333	94.3333	97.5000	96.6667
11	87.5000	94.3333	93.3333	95.0000
12	87.5000	96.0000	97.5000	92.5000
13	90.0000	96.0000	93.3333	100.0000
14	85.8333	96.0000	88.3333	100.0000
15	91.6667	95.1667	91.6667	99.1667
16	92.5000	94.3333	94.1667	98.3333
17	90.8333	95.1667	91.6667	87.5000
18	92.5000	96.0000	99.1667	99.1667
19	88.3333	94.3333	97.5000	65.0000
20	93.3333	96.0000	100.0000	87.5000
21	88.3333	95.1667	98.3333	76.6667
22	94.1667	96.0000	90.8333	95.0000
23	85.8333	95.1667	95.8333	83.3333
24	90.8333	95.1667	97.5000	90.8333
25	88.3333	94.3333	93.3333	92.5000
26	92.5000	95.1667	87.5000	100.0000
27	91.6667	94.3333	92.5000	100.0000
28	89.1667	89.3333	82.5000	95.0000
29	89.1667	95.1667	80.0000	100.0000
30	93.3333	96.0000	95.8333	89.1667
31	94.1667	95.1667	99.1667	95.8333
32	87.5000	96.0000	94.1667	73.3333
33	88.3333	96.0000	97.5000	86.6667
34	89.1667	95.1667	88.3333	98.3333
35	90.0000	95.1667	99.1667	99.1667
36	92.5000	96.0000	94.1667	100.0000
37	96.6667	92.6667	94.1667	100.0000

Bias Relatif MSE

AREA	DIRECT	EBLUP	LOG	EBP
38	-0.11151	-0.02628	-0.02561	-0.02816
39	-0.12247	-0.02666	-0.02422	-0.03185
40	0.03834	-0.02758	-0.02658	-0.01822
41	0.03459	-0.02366	-0.02651	-0.02253
42	0.04313	-0.02520	-0.02742	-0.02546
43	0.06366	-0.02657	-0.02690	-0.03083
44	-0.00186	-0.02395	-0.02787	-0.02601
45	-0.15918	-0.02569	-0.02752	-0.02901
46	0.10932	-0.02350	-0.02814	-0.02827
47	-0.02609	-0.02554	-0.02788	-0.02700
48	0.03355	-0.02401	-0.02738	-0.02646
49	-0.07105	-0.02712	-0.02758	-0.02603
50	0.03029	-0.02575	-0.02784	-0.02497
51	-0.19472	-0.02723	-0.02785	-0.02209
52	0.22420	-0.02252	-0.02730	-0.02982
53	0.04042	-0.02477	-0.02852	-0.02445
54	0.05471	-0.02500	-0.02270	-0.02882
55	0.07206	-0.02141	-0.02704	-0.03015
56	-0.27614	-0.02666	-0.02756	-0.03004
57	0.04583	-0.02569	-0.02782	-0.02958
58	0.17264	-0.01921	-0.02517	-0.03068
59	0.07925	-0.02484	-0.02619	-0.02926
60	-0.08612	-0.02732	-0.02663	-0.03086
61	0.12942	-0.02373	-0.02532	-0.03027
62	-0.11270	-0.02309	-0.02515	-0.03027
63	0.21223	-0.02385	-0.02334	-0.03216
64	-0.13232	-0.02560	-0.02588	-0.03029
65	0.27884	-0.01927	-0.02521	-0.03094
66	0.01620	-0.02462	-0.02229	-0.03140
67	-0.08694	-0.02615	-0.02313	-0.03022
68	0.40004	-0.02410	-0.02005	-0.03162
69	0.06355	-0.02523	-0.02778	-0.02917
70	0.01886	-0.02423	-0.02761	-0.03056
71	-0.04052	-0.02330	-0.02745	-0.02853
72	-0.08271	-0.02631	-0.02786	-0.02862
73	0.20123	-0.02404	-0.02794	-0.02843
74	-0.09683	-0.02743	-0.02767	-0.02719
75	0.29940	-0.02173	-0.02745	-0.02628
76	-0.09480	-0.02662	-0.02826	-0.02944
77	-0.23197	-0.02374	-0.02842	-0.02508

Tingkat ketercakupan (coverage rate)

AREA	DIRECT	EBLUP	LOG	EBP
38	90.0000	92.6667	95.8333	98.3333
39	90.8333	94.3333	98.3333	65.8333
40	91.6667	92.6667	93.3333	100.0000
41	95.0000	96.0000	94.1667	100.0000
42	89.1667	95.1667	94.1667	100.0000
43	91.6667	92.6667	85.8333	82.5000
44	90.0000	94.3333	83.3333	100.0000
45	85.0000	95.1667	90.8333	100.0000
46	91.6667	96.0000	85.8333	100.0000
47	95.0000	94.3333	84.1667	100.0000
48	92.5000	95.1667	86.6667	100.0000
49	90.8333	95.1667	96.6667	100.0000
50	94.1667	95.1667	86.6667	100.0000
51	88.3333	93.5000	85.0000	100.0000
52	95.8333	94.3333	86.6667	98.3333
53	90.0000	96.0000	81.6667	100.0000
54	90.8333	96.0000	100.0000	96.6667
55	88.3333	95.1667	94.1667	95.8333
56	89.1667	93.5000	94.1667	97.5000
57	91.6667	92.6667	89.1667	100.0000
58	91.6667	96.0000	98.3333	88.3333
59	93.3333	94.3333	93.3333	98.3333
60	93.3333	90.1667	95.8333	89.1667
61	90.8333	96.0000	99.1667	96.6667
62	93.3333	95.1667	98.3333	90.0000
63	94.1667	96.0000	99.1667	53.3333
64	90.8333	96.0000	98.3333	91.6667
65	95.0000	96.0000	98.3333	82.5000
66	92.5000	94.3333	99.1667	70.8333
67	94.1667	95.1667	97.5000	90.8333
68	94.1667	95.1667	98.3333	75.0000
69	88.3333	93.5000	89.1667	100.0000
70	94.1667	94.3333	90.8333	98.3333
71	91.6667	95.1667	95.0000	100.0000
72	86.6667	93.5000	83.3333	100.0000
73	93.3333	94.3333	91.6667	100.0000
74	89.1667	91.8333	87.5000	99.1667
75	94.1667	96.0000	88.3333	100.0000
76	85.8333	95.1667	82.5000	100.0000
77	84.1667	95.1667	86.6667	100.0000

Bias Relatif MSE

AREA	DIRECT	EBLUP	LOG	EBP
78	0.17649	-0.02374	-0.02770	-0.02263
79	-0.04006	-0.02470	-0.02803	-0.02836
80	0.26873	-0.02458	-0.02776	-0.02444
81	0.05838	-0.02666	-0.02839	-0.02289
82	0.00501	-0.02341	-0.02767	-0.02549
83	0.20591	-0.02576	-0.02192	-0.02997
84	0.00359	-0.02530	-0.02825	-0.02496
85	0.02020	-0.02593	-0.02727	-0.02480
86	-0.11904	-0.02572	-0.02817	-0.02510
87	0.01755	-0.02560	-0.02779	-0.02888
88	-0.13935	-0.02588	-0.02864	-0.02593
89	-0.00668	-0.02762	-0.02807	-0.03177
90	-0.04973	-0.02671	-0.02834	-0.02855
91	0.02893	-0.02754	-0.02715	-0.03170
92	-0.04986	-0.02045	-0.02710	-0.02591
93	0.34057	-0.02270	-0.02786	-0.02565
94	0.24334	-0.02008	-0.02782	-0.02741
95	0.14163	-0.02213	-0.02731	-0.02739
96	0.16399	-0.01964	-0.02798	-0.02817
97	0.07295	-0.02626	-0.02472	-0.03054
98	-0.03558	-0.02569	-0.02594	-0.02888
99	0.09652	-0.02322	-0.02819	-0.02565
100	-0.13102	-0.02470	-0.02790	-0.02857
101	-0.10809	-0.02418	-0.02770	-0.02460
102	-0.01107	-0.02164	-0.02693	-0.02875
103	-0.03871	-0.02481	-0.02794	-0.02676
104	0.12100	-0.02587	-0.02357	-0.02949
105	-0.25302	-0.02421	-0.02789	-0.02956

Tingkat ketercakupan (coverage rate)

AREA	DIRECT	EBLUP	LOG	EBP
78	90.0000	94.3333	86.6667	100.0000
79	90.8333	94.3333	81.6667	100.0000
80	93.3333	95.1667	85.0000	100.0000
81	90.8333	94.3333	82.5000	100.0000
82	88.3333	95.1667	87.5000	100.0000
83	91.6667	95.1667	97.5000	86.6667
84	93.3333	95.1667	83.3333	100.0000
85	95.0000	93.5000	93.3333	100.0000
86	90.0000	94.3333	84.1667	100.0000
87	93.3333	96.0000	83.3333	100.0000
88	89.1667	93.5000	80.0000	100.0000
89	89.1667	93.5000	85.0000	73.3333
90	87.5000	92.6667	80.0000	100.0000
91	95.8333	90.1667	90.0000	68.3333
92	88.3333	96.0000	96.6667	100.0000
93	92.5000	95.1667	82.5000	100.0000
94	90.8333	96.0000	82.5000	100.0000
95	92.5000	96.0000	92.5000	100.0000
96	94.1667	96.0000	84.1667	100.0000
97	87.5000	96.0000	98.3333	85.8333
98	88.3333	94.3333	99.1667	100.0000
99	98.3333	94.3333	81.6667	100.0000
100	89.1667	93.5000	87.5000	100.0000
101	90.0000	96.0000	85.0000	100.0000
102	94.1667	95.1667	99.1667	100.0000
103	90.0000	96.0000	84.1667	100.0000
104	93.3333	94.3333	99.1667	96.6667
105	85.8333	95.1667	85.8333	100.0000

Lampiran 8.

Pemrograman SAS pendugaan pendapatan per kapita di Kabupaten dan Kota Bogor berdasarkan data Susenas 2005 dan Podes 2005

```
data ipb.bogorproject;
set ipb.jabarproject;
if kab = 3201 or kab = 3271;
run;

=====; * Direct Estimate; =====;
proc tabulate data = ipb.bogorproject out=b;
class desa ;
var logy y household;
table desa, n (logy y)*(mean stderr) household*mean;
run;

data b (keep = desa Direct Stderr RRMSE Directlog Stderrlog
RRMSElog household sumy);
set b;
Directlog = logy_mean;
Stderrlog = logy_stderr;
RRMSElog = stderrlog/directlog*100;
Direct = y_mean;
Stderr = y_stderr;
RRMSE = stderr/direct*100;
household = household_mean;
sumy = n*y_mean;
run;

=====; * Management data for Indirect Estimate; =====;
proc mixed data=ipb.bogorproject noclprint covtest scoring;
class desa;
model y = status pctfarmer farmworker ptelectrics PCTGrthealt
PCTRicefield citydistance citytime / s outpm = a ddfm =
kenwardroger residual;
* outpm = a => without random effect;
random intercept /sub = desa;
run;

/* Di = 6.738E10 / ni, (2.3821E9)^2
A = 2.16E10, (3.7248E9)^2 */

proc tabulate data = a out=c;
class kab desa namadesa ;
var status pctfarmer farmworker ptelectrics PCTGrthealt
PCTRicefield citydistance citytime pred;
table desa, n mean*(status pctfarmer farmworker ptelectrics
PCTGrthealt PCTRicefield citydistance citytime pred);
run;

data c (keep = desa n x0 x1 x2 x3 x4 x5 x6 x7 x8 Di xb);
set c;
x0 = 1;
x1 = status_mean;
x2 = pctfarmer_mean;
x3 = farmworker_mean;
```

```

x4 = ptelectrics_mean;
x5 = PCTGrthealt_mean;
x6 = PCTRicefield_mean;
x7 = citydistance_mean;
x8 = citytime_mean;
xb = pred_mean;
Di = 6.738E10/n; *from proc mixed;
run;

*====*; * Management data for Indirect Estimate; *====*;
proc mixed data=ipb.bogorproject noclprint covtest scoring;
class desa;
model logy = status pctfarmer farmworker ptelectrics PCTGrthealt
PCTRicefield citydistance citytime / s outpm = d ddfm =
kenwardroger residual;
* outpm = a => without random effect;
random intercept /sub = desa;
run;

/* Dilog = 0.1573 / ni, (0.005560)^2
   Alog = 0.04849, (0.008414)^2 */

proc tabulate data = d out=d;
class desa ;
var pred;
table desa, n mean*(pred);
run;

data d (keep = desa n Dilog xb);
set d;
xb = pred_mean;
Dilog = 0.1573/n; *from proc mixed;
run;

*====*; * Merge data direct + data dasar ; *====*;
proc sort data=b; by desa; run;
proc sort data=c; by desa; run;
proc sort data=d; by desa; run;

data baru;
merge b c d;
by desa;
run;

*** MODEL : Y = XB + U + E, where U ~ N(0, A) dan E ~ N(0, D); ***
PARAMETER BETHA, DESAIN MATRIK X DAN RAGAM D;
proc iml;
M = 105;
load _all_;
use baru;
read all;
X = (x0||x1||x2||x3||x4||x5||x6||x7||x8);

D = (Di);
A_hat = 2.16E10; *from proc mixed;
V = D + A_hat;
sigma = I(105)#V;

Dlog = (Dilog);

```

```

Alog = 0.04849; *from proc mixed;
Vlog = Dlog + Alog;
sigmalog = I(105)#Vlog;
fraksi = n/household;

/* ===== EBLUP ===== */
theta_hat = XB + (A_hat / (A_hat + D))#(Direct - XB);

*** MSE of theta_hat;
g1 = (A_hat#D)/(A_hat + D);
g2 = (D##2/(A_hat + D)##2)# vecdiag(X*inv(X`*inv(sigma)*X)*X`);
g3 = (2*D##2/(M##2 * (A_hat + D)##3))* sum(V##2);

*** MSE of theta_hat : Prasad & Rao Estimator;
MSEP = g1 + g2 + 2*g3;
MSEN = g1 + g2;

RRMSEN = sqrt(MSEN)/theta_hat*100;
RRMSEP = sqrt(MSEP)/theta_hat*100;

gamma = (A_hat)/(A_hat + D);

* print desa Direct RRMSE gamma theta_hat RRMSEN RRMSEP;
/* ===== Log-Scale EBLUP ===== */
thetalog = XBlog + (Alog / (Alog + Dlog))#(Directlog - XBlog);

*** MSE of theta_hat;
g1log = (Alog#Dlog)/(Alog + Dlog);
g2log = (Dlog##2/(Alog + Dlog)##2)# vecdiag(X*inv(X`*inv(sigmalog)*X)*X`);
g3log = (2*Dlog##2/(M##2 * (Alog + Dlog)##3))* sum(Vlog##2);

*** MSE of theta_hat : Prasad & Rao Estimator;
MSEPR = g1log + g2log + 2*g3log;

theta_hatlog = exp(thetalog + 0.5#MSEPR);
MSEPlog = exp(MSEPR)#(exp(MSEPR)-1)#exp(2#thetalog);
RRMSEPlog = sqrt(MSEPlog)/theta_hatlog*100;

/* ===== Log-Normal EBP ===== */
Xb2 = (xblog + Vlog/2);
TB = sumy + (household - n)#exp(xb2 + Vlog/2);
MeanTB = TB / household;
* mu = fraksi#(sumy/n) + (1-fraksi)#(exp(xb2 + Vlog/2));

Finite1 = exp(Vlog)#(exp(Vlog) - 1)#exp(Xb2);

Finite2 = (2*(Alog##2)/(m-1) + 2*(Dlog##2)/(n)) # (0.5 * exp(Xb2))##2;

Finite3 = 0.25 * exp(Alog) # ((household-n) # exp(2*xblog+Dlog)) *
(Alog**2 * 0.005560**2 + 0.1573**2 * 0.008414**2
+ 2*Alog*0.1573**2*1.91732030577329E-06);

MSETB = finite1 + finite2 + 2*finite3;
RRMSETB =sqrt(MSETB)/meanTB * 100;
print desa Direct theta_hat theta_hatlog MeanTB RRMSE RRMSEP
RRMSEPlog RRMSETB;
quit;

```

Lampiran 9.

Penduga pengeluaran per kapita dalam rupiah di Kabupaten dan Kota Bogor dan RMSE relatifnya berdasarkan data Susenas tahun 2005

Desa/ Kelurahan	Penduga Parameter (Rp.)				Penduga RMSE Relatif (%)			
	DIRECT	EBLUP	Log-Scale EBLUP	Lognormal EBP	DIRECT	EBLUP	Log-Scale EBLUP	Lognormal EBP
3201010002 Bantar Karet	527,541	500,845	442,736	350,779	12.26599	12.07658	9.23799	6.76057
3201010010 Parakan Muncang	364,312	377,754	347,918	412,781	9.96789	15.98887	9.22425	6.77341
3201020005 Karacak	308,603	311,835	304,737	342,374	6.47340	19.32912	9.20454	6.89755
3201020006 Barengkok	391,584	385,196	366,812	359,268	8.53178	16.14770	9.49978	6.66178
3201030005 Gunung Sari	355,579	362,994	344,841	387,716	8.04356	16.57760	9.18874	6.55037
3201030006 Gunung Bunder 2	316,928	321,964	308,340	348,217	7.72431	18.66518	9.17591	5.55154
3201040004 Ciaruten Udik	276,806	292,381	276,673	371,016	8.80755	20.50690	9.15410	4.94562
3201040005 Cibatok 1	359,929	379,137	358,455	449,930	8.07166	15.77677	9.13137	5.92451
3201040013 Leuweung Kolot	339,212	360,805	339,162	446,359	9.05583	16.57621	9.13011	5.05531
3201050013 Cihideung Udik	599,059	568,404	538,049	407,038	7.74530	10.64869	9.24479	7.20803
3201050018 Benteng	624,363	612,086	551,364	509,323	10.78972	9.83048	9.18798	6.62259
3201060003 Sukadamai	415,175	407,620	382,484	371,556	9.70230	14.74386	9.17650	5.41397
3201060009 Babakan	474,440	483,954	458,483	491,142	7.29715	12.38375	9.14986	7.87371
3201070012 Ciomas	358,673	387,599	347,413	498,507	13.46225	15.45155	9.14323	6.87330
3201070018 Ciomas Rahayu	995,648	925,746	787,353	528,207	15.71241	6.47451	9.15079	6.94150
3201071006 Tamansari	381,576	394,597	376,848	428,482	6.54225	15.19513	9.15430	6.92606
3201080011 Cijeruk	432,161	429,479	396,938	397,233	10.57447	13.93243	9.13475	4.97477
3201080018 Sukaharja	584,697	557,837	522,049	402,173	9.02741	10.82295	9.22039	7.41193
3201090005 Muara Jaya	660,470	631,634	595,900	468,265	7.59814	9.55470	9.21663	4.61980
3201100001 Cileungsi	451,993	447,529	397,820	405,921	15.58217	13.47039	9.20597	5.38000
3201100005 Jambu Luwuk	466,020	459,338	426,227	412,660	10.77943	13.15481	9.22846	4.91916
3201100013 Pandansari	535,499	535,559	497,940	502,751	9.25694	11.18714	9.14701	5.85102
3201110005 Batu Layang	534,945	531,501	467,743	465,361	15.97666	11.39046	9.24683	5.92894
3201120005 Sukamanah	345,159	352,030	306,478	379,820	15.55117	17.22728	9.26352	6.32925
3201120007 Sukamahi	380,354	398,559	372,273	450,227	9.32956	15.10892	9.19547	6.40854
3201130009 Cijujung	712,927	690,285	585,401	534,593	19.04966	8.65924	9.12464	10.03319
3201130012 Cilebut Timur	846,890	801,307	675,287	525,888	17.63483	7.46193	9.12776	7.32492
3201140001 Cijayanti	485,611	485,846	430,749	447,805	14.90714	12.41115	9.20838	7.09598
3201140009 Sentul	538,960	542,738	475,395	523,990	14.67999	10.99799	9.11139	7.14882
3201150001 Sukawangi	388,928	379,225	372,284	337,860	5.41922	15.92310	9.22198	6.16961
3201150008 Pabuaran	365,811	364,134	348,703	366,081	8.23919	16.64133	9.25581	6.57900
3201160018 Kuta Mekar	421,814	402,084	375,031	324,864	13.02386	14.94310	9.17409	4.67137
3201170007 Balekambang	427,522	422,667	401,423	384,281	8.46966	14.33354	9.25362	6.05404
3201170011 Sukamaju	385,400	408,459	390,723	488,937	5.90994	14.67534	9.15162	6.29519
3201180013 Cipeucang	466,724	465,139	449,908	440,652	5.70487	12.95255	9.20016	6.65386
3201180018 Cileungsi	537,866	546,392	521,335	545,723	6.79499	10.96928	9.15043	8.79412
3201180020 Pasir Angin	454,981	463,734	426,112	467,162	10.22734	12.97138	9.18507	11.94325
3201181002 Lulut	327,296	354,457	339,868	456,738	4.56215	16.94293	9.16956	7.41284

Desa/ Kelurahan	Penduga Parameter (Rp.)				Penduga RMSE Relatif (%)			
	DIRECT	EBLUP	Log-Scale EBLUP	Lognormal EBP	DIRECT	EBLUP	Log-Scale EBLUP	Lognormal EBP
3201181008 Bojong	426,047	429,823	409,179	431,875	8.08505	14.12300	9.27287	4.26304
3201190003 Tlajung Udik	817,634	768,851	677,700	479,324	12.83250	7.85282	9.22071	13.91488
3201190005 Cicadas	429,760	439,553	395,776	458,756	15.88976	13.66410	9.17047	11.27460
3201190009 Ciangsana	499,050	514,409	489,305	547,839	7.52461	11.64941	9.14888	10.06174
3201200004 Pasir Mukti	332,931	348,999	324,945	408,742	11.79967	17.30342	9.22267	5.49189
3201200008 Karang Asem Barat	498,906	508,409	484,512	517,768	6.90104	11.74282	9.11319	9.01614
3201200011 Gunung Sari	370,439	393,896	354,597	482,480	12.91614	15.16398	9.11777	6.83855
3201200014 Puspasari	668,831	654,457	566,268	540,379	15.54133	9.12680	9.11788	7.63243
3201210004 Cibinong	430,801	456,702	414,123	541,312	11.69910	13.43669	9.36682	8.09799
3201210006 Sukahati	544,741	551,732	507,583	551,381	9.78471	10.83009	9.12141	8.54779
3201210010 Pabuaran	678,385	670,551	624,925	547,423	6.72315	6.55660	6.71431	13.94099
3201210012 Ciriung	547,820	553,716	505,799	545,137	12.50421	10.78612	9.11685	9.77303
3201220005 Bojongsgede	760,264	733,413	644,062	558,223	15.90504	8.14571	9.11959	11.15025
3201220012 Bojong Baru	681,096	663,259	591,821	536,697	13.47378	9.01501	9.12775	6.17452
3201220014 Pabuaran	441,219	466,122	430,121	554,100	9.94534	12.81585	9.11893	10.12489
3201230013 Tegal	508,134	491,734	427,039	400,693	21.88045	12.31027	9.24576	7.05911
3201230015 Jampang	455,648	469,392	437,918	510,252	9.74147	12.80279	9.17592	6.36512
3201240011 Pamager Sari	338,487	367,127	342,193	473,991	7.92437	16.35609	9.16834	6.46615
3201240017 Cogreg	340,900	364,457	348,373	450,313	6.13300	16.86808	9.38454	6.24554
3201241004 Putat Nutug	302,146	310,499	300,529	359,571	7.57687	19.39854	9.19774	5.27408
3201250006 Curug	551,180	535,505	487,963	424,838	11.74308	11.20941	9.16504	6.78292
3201250007 Rawakalong	550,354	538,797	513,429	454,981	6.86335	11.18024	9.19878	5.48873
3201260005 Cibodas	459,744	452,146	414,731	402,385	11.48962	13.30344	9.18480	5.69558
3201270009 Sukamaju	380,082	392,560	377,854	425,834	5.54070	15.41256	9.24095	6.29928
3201270017 Banyu Asih	286,617	290,480	282,803	323,184	6.72730	20.90647	9.27688	3.97148
3201271003 Kiarapandak	284,602	287,356	263,376	319,255	14.71873	20.97563	9.20451	6.51452
3201280005 Sipak	243,369	258,233	247,588	336,399	7.85186	23.58304	9.30398	6.51811
3201280008 Curug	405,641	378,193	352,283	278,879	11.60047	16.29734	9.42120	5.99128
3201290004 Babakan	266,055	289,850	274,819	392,052	6.72481	20.93133	9.26738	6.24190
3201300005 Pingku	350,055	358,721	329,652	386,311	12.88776	17.70486	9.70847	5.13135
3271010002 Pamoyanan	352,221	381,537	358,167	495,607	7.93897	15.66903	9.12619	6.89204
3271010004 Genteng	432,471	451,295	408,647	513,598	13.29799	13.24580	9.12532	5.52549
3271010008 Harjasari	392,098	415,808	403,976	508,038	3.71692	14.39068	9.13487	7.23018
3271010011 Cipaku	443,505	466,935	427,433	548,273	11.47549	12.79361	9.11898	7.28740
3271010013 Batutulis	659,507	647,941	545,063	551,513	16.08239	9.21899	9.11830	7.32134
3271010015 Empang	577,624	569,521	543,185	489,436	6.60054	10.60834	9.22712	9.06129
3271010016 Cikaret	514,436	522,590	488,116	528,091	9.35147	11.42778	9.11620	8.61503
3271020002 Sindangrasa	605,858	601,842	568,795	543,901	8.07064	9.91955	9.11295	6.53198
3271020004 Katulampa	318,070	356,165	333,752	513,992	6.72544	16.76575	9.11514	9.51991
3271020005 Baranangsiang	505,012	518,530	472,032	550,026	12.02307	11.52170	9.11986	10.82003
3271020006 Sukasari	486,260	501,006	431,242	536,902	19.76298	11.91621	9.11309	7.36409
3271030001 Bantarjati	505,012	520,623	473,832	562,523	12.02307	11.47634	9.12067	9.93499

Desa/ Kelurahan	Penduga Parameter (Rp.)				Penduga RMSE Relatif (%)			
	DIRECT	EBLUP	Log-Scale EBLUP	Lognormal EBP	DIRECT	EBLUP	Log-Scale EBLUP	Lognormal EBP
3271030002 Tegal Gundil	491,444	508,560	465,841	556,832	10.07117	11.74650	9.11900	10.73189
3271030003 Tanahbaru	650,851	636,670	575,912	525,354	11.51350	9.38218	9.11830	9.15517
3271030004 Cimahpar	401,998	407,705	374,716	405,030	10.07056	15.10014	9.40981	7.70448
3271030006 Cibuluh	553,789	558,692	524,453	544,266	10.27629	10.68491	9.11226	9.54250
3271030007 Kedunghalang	683,710	665,346	568,161	544,518	16.38840	9.23866	9.38330	9.27395
3271030008 Ciparigi	405,031	434,362	401,210	545,236	8.37388	13.74538	9.11371	9.55045
3271040003 Babakanpasar	614,331	608,906	488,316	544,168	19.77389	9.80696	9.11536	7.02248
3271040004 Tegallega	462,562	479,449	444,977	525,908	9.19648	12.46389	9.12219	9.18115
3271040007 Pabaton	1,567,773	1,409,994	1,198,956	581,842	10.53489	4.23833	9.12254	4.04620
3271040010 Kebonkelapa	982,102	917,591	811,955	550,450	11.38774	6.50912	9.11726	7.27496
3271050001 Pasirmulya	546,893	547,214	502,404	512,171	12.73436	10.97735	9.17184	4.32481
3271050003 Pasirjaya	408,920	433,822	407,082	524,341	8.45118	13.80857	9.14555	9.02464
3271050004 Gunungbatu	556,493	561,536	543,575	549,976	4.75927	10.63511	9.11614	9.16501
3271050006 Menteng	507,573	520,152	483,393	546,590	10.29911	11.47894	9.11422	8.08197
3271050008 Cilendek Barat	594,538	590,408	570,256	535,326	5.17781	10.12778	9.12813	7.97942
3271050009 Sindangbarang	750,667	721,032	613,828	533,242	15.78157	8.28092	9.11424	7.49842
3271050012 Situgede	478,881	476,170	456,374	444,939	6.78617	12.71934	9.25086	5.96056
3271050014 Semplak	616,821	609,329	573,591	542,508	8.54101	9.85577	9.16936	6.96523
3271060001 Kedungwaringin	464,557	481,683	445,813	529,779	11.71150	12.39672	9.11501	9.22132
3271060002 Kedungjaya	560,473	564,837	524,308	549,321	9.95505	10.57390	9.11698	7.21135
3271060003 Kebonpedes	449,201	472,276	447,225	552,539	7.37147	12.64417	9.11540	9.72940
3271060005 Kedungbadak	697,475	683,668	588,333	503,423	12.83002	6.44023	6.72459	10.72529
3271060008 Cibadak	566,478	569,426	504,505	545,364	14.48496	10.48858	9.11688	8.60141
3271060009 Kayumanis	505,628	506,811	469,434	488,971	10.09923	11.92309	9.22883	6.63896
3271060011 Kencana	386,722	411,912	380,877	505,752	10.36220	14.50523	9.12073	6.47718