

Model *Hierarchical Bayes* pada *Small Area Estimation* untuk Pendugaan Proporsi Pengangguran pada Desain Survei Kompleks

¹Arip Juliyanto, ²Heri Kuswanto, ³Ismaini Zain

^{1,2,3}Statistika FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

email: arip_jy@mhs.statistika.its.ac.id, heri_k@statistika.its.ac.id, ismaini_z@statistika.its.ac.id

Abstrak— Tingkat pengangguran menjadi salah indikator utama dalam pembangunan. Sesuai dengan rancangan surveinya, indikator pengangguran ini hanya tersedia sampai tingkat kabupaten. Jika digunakan secara langsung untuk menghitung pengangguran tingkat kecamatan akan menghasilkan *standard error* yang besar karena jumlah sampel yang kecil, akibatnya penduga parameternya menjadi kurang bisa dipercaya. Untuk mengatasi keadaan tersebut maka dikembangkan metode *Small Area Estimation* (SAE). Salah satu metode SAE yang sering digunakan karena keunggulannya adalah model *Hierarchical Bayes* (HB). Dalam pemodelan SAE dengan HB selama ini sering digunakan model *Fay-Herriot* (FH) dengan asumsi varian diketahui. Asumsi ini sulit untuk dipenuhi karena dalam kenyataannya seringkali parameter varian tidak diketahui sehingga harus tetap diestimasi. Penelitian ini bermaksud untuk membandingkan performa pemodelan FH dengan asumsi varian diketahui dengan model HB yang lain dengan asumsi varian tidak diketahui dalam rangka mengestimasi proporsi pengangguran tingkat kecamatan. Model yang dibandingkan adalah model dengan asumsi varian diketahui yaitu model FH dan *Normal-Logistic* (NL) dengan model dengan asumsi varian sampling tidak diketahui yaitu model *Normal Logistic random sampling* (NLrs) dan *Beta-Logistic random sampling* (BLrs). Dari Hasil penghitungan SAE dengan validasi proporsi dari populasi didapat kesimpulan bahwa model *Normal-Logistic random sampling* (NLrs) merupakan model terbaik untuk melakukan prediksi pengangguran di tingkat kecamatan.

Kata Kunci— *small area estimation*, *hierarchical bayes*, survei kompleks, *two stage sampling*, pengangguran

I. PENDAHULUAN

Selama kurun waktu sepuluh tahun terakhir (2004-2013), berdasarkan data BPS, Provinsi Riau menjadi provinsi kedua dengan tingkat pengangguran terbesar di Pulau Sumatera yang mencapai 8,98 persen di bawah Provinsi NAD 9,56 persen. Pada tahun 2014, Tingkat Pengangguran Terbuka (TPT) Provinsi Riau mencapai 6,65 persen atau meningkat dari tahun sebelumnya yang mencapai 5,48 persen.

Informasi mengenai Tingkat Pengangguran Terbuka (TPK) yang tersedia sampai saat ini hanya bisa diketahui sampai pada tingkat kabupaten. Padahal untuk berbagai tujuan dan kepentingan maka informasi yang memadai yang bisa menjangkau area yang lebih kecil (*small area statistic*) menjadi sangat penting. Disebabkan jumlah sampel yang

tersedia relatif kecil dan metodologi samplingnya (Sakernas) didesain hanya untuk keperluan estimasi sampai tingkat kabupaten, maka informasi ketenagakerjaan sampai tingkat kecamatan selama ini tidak tersedia. Jika data hasil Survei Sakernas ini tetap digunakan untuk mengestimasi secara langsung tingkat pengangguran pada tingkat wilayah kecil (seperti kecamatan) maka akan menghasilkan *standard error* yang besar sehingga pendugaan indikator tingkat pengangguran hasilnya kurang bisa dipercaya.

Estimasi angka pengangguran tetap bisa dilakukan sampai tingkat kecamatan jika dilakukan penambahan jumlah sampel dari suatu survei, tetapi langkah ini biasanya terkendala dengan besarnya biaya yang diperlukan. Untuk mengatasi permasalahan ini maka dewasa ini dikembangkan metode *Small Area Estimation* (SAE). Metode SAE merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi dengan ukuran sampel kecil melalui pendekatan estimasi tidak langsung (*indirect estimation*).

Secara umum terdapat beberapa metode SAE yang sering digunakan antara lain *Empirical Best Linear Unbiased Prediction* (EBLUP), *Empirical Bayes* (EB) dan *Hierarchical Bayes* (HB) *Estimation* [12]. Model HB menurut banyak penelitian menjadi model yang relatif lebih unggul jika dibandingkan dengan model EBLUP dan EB. Hasil penelitian Arora dan Lahiri [2] menunjukkan bahwa estimator yang dihasilkan dengan metode Bayes (HB) mempunyai *Average Squared Relative Deviation* (ASRD) yang lebih kecil jika dibandingkan dengan metode EBLUP. You dan Chapman (2006) juga menyimpulkan bahwa pendekatan HB lebih baik dibandingkan EBLUP pada kasus varian yang diestimasi dan jumlah sampel yang kecil. Pereira, Mendes dan Coelho [28] juga menyimpulkan dari study simulasi bahwa HB memiliki performa lebih baik dibanding EBLUP dengan melihat dari nilai rasio absolut bias (*Absolute bias ratio*). Sementara berdasarkan penelitian Farrell [7] juga menyimpulkan bahwa metode HB relatif lebih baik dibandingkan dengan EB dengan statistik rata-rata absolut bias.

Selama ini model dasar SAE berbasis area yang secara luas sering digunakan termasuk untuk pendugaan proporsi adalah model *Fay-Herriot* (FH) [36]. Pada model ini diasumsikan bahwa proporsi dari data sampel mempunyai distribusi normal dengan varian sampling yang diketahui. Sementara pada kenyataannya varian sampling sendiri tidak diketahui sehingga harus diduga dengan hasil survei itu sendiri. Dengan tetap mengasumsikan varian diketahui maka model FH menggunakan berbagai metode untuk menduga nilai varian seperti dengan menggunakan teknik fungsi *Generalized*

Variance Function (GVF) [6] ataupun dengan menggunakan penduga varian pada wilayah yang lebih luas (sensus domain) [13].

Hawala dan Lahiri [16] menggunakan model HB-FH untuk mengestimasi tingkat kemiskinan pada Biro Sensus Amerika Serikat. Dalam mengestimasi varian yang tidak diketahui mereka menggunakan penggunaan metode GVF. Sementara Ha *et al.* [13] melakukan *treatment* untuk menduga varian tidak diketahui dengan pendekatan sensus domain dalam rangka mengestimasi proporsi perokok di Amerika Serikat pada tahun 2006-2010. Penggunaan metode GVF ataupun penduga varian pada domain sensus (wilayah yang lebih luas) dilakukan untuk menjadikan penduga varian menjadi lebih stabil mengingat jumlah sampel pada wilayah ini sangat kecil bahkan tidak terdapat sampel sehingga penggunaan penduga varian secara langsung menjadi kurang tepat.

Sebagai alternatif dari model dasar Fay-Herriot, dalam penelitiannya, Liu, Lahiri dan Kalton [22] mengajukan beberapa model *Hierarchical Bayes* (HB) lainnya sebagai pengembangan model FH dengan asumsi sampling varian tidak diketahui. Model alternatif ini kemudian dibandingkan dengan model HB FH. Hasil dari penelitian tersebut tidak memberikan rekomendasi model mana yang paling baik untuk digunakan mengestimasi proporsi bayi dengan berat badan lahir rendah di Amerika Serikat.

Dalam pemodelan SAE seringkali diasumsikan bahwa data yang digunakan khususnya data respons berasal dari desain survei sederhana (*Simple Random Sampling-SRS*), padahal dalam kenyataannya data yang digunakan dalam pembentukan model SAE selain berasal dari desain SRS juga seringkali berasal dari *survey sampling* dengan desain yang lebih kompleks seperti Stratifikasi SRS, *Two Stage Sampling* dan sebagainya. Liu *et al.* [22] dalam penelitiannya telah memper-timbangkan desain survei stratifikasi SRS pada empat model HB baik dengan asumsi varian diketahui maupun tidak. Sementara Liu dan Diallo [21] masih dengan desain sampling yang sama mengestimasi parameter proporsi dengan menggunakan transformasi *arcsin* dalam pembentukan modelnya. Yang terbaru Ha *et al.* [22] dengan menggunakan data dari NHIS menduga proporsi perokok dengan pertimbangan desain survei yang lebih kompleks (*two stage sampling survei*) dengan menggunakan tiga jenis model HB.

Dari uraian sebelumnya tergambar bahwa kebutuhan akan adanya data proporsi pengangguran pada wilayah kecil (kecamatan) menjadi hal yang penting. Kemudian dari beberapa penelitian menunjukkan bahwa SAE dengan model HB lebih tepat untuk menangani data respon yang berasal dari data kategorik. Selain itu, desain survei juga menjadi hal yang penting yang perlu dipertimbangan dalam pemodelan SAE HB. Oleh karena itu maka dalam penelitian ini bermaksud akan dilakukan pemodelan proporsi pengangguran tingkat kecamatan di Provinsi Riau dengan menggunakan beberapa model HB (dengan asumsi varian diketahui dan tidak) yaitu model *Fay-Herriot* (FH), model *Normal-Logistic* (NL), model *Normal-Logistic random sampling* (NLrs), dan model *Beta-Logistic random sampling* (BLrs) pada desain survei kompleks (dalam hal ini desain *Two stage sampling*).

Tujuan penelitian ini sendiri untuk memperoleh informasi kinerja beberapa model HB untuk pendugaan

proporsi pengangguran pada desain survei kompleks (*Two stage sampling*). Dengan demikian akan didapatkan manfaat mengenai kinerja beberapa model HB SAE dengan berbagai model estimasi sehingga menambah ragam pemodelan untuk digunakan dalam estimasi parameter (proporsi) pada wilayah kecil (kecamatan). Selain itu dari hasil perbandingan kinerja juga bisa bahan diskusi selanjutnya mengenai pemodelan yang lebih tepat dengan beberapa model HB untuk pendugaan parameter proporsi pada wilayah kecil dengan desain survei kompleks.

II. TINJAUAN PUSTAKA

A. Kebutuhan adanya *Small Area Estimation*

Suatu daerah untuk kepentingan pembangunan seringkali harus menyediakan data yang lengkap. Dengan melakukan sensus maka informasi atau data sampai level terkecil dapat didapatkan. Tetapi sebagaimana diketahui bahwa kegiatan sensus membutuhkan sumber daya yang sangat besar baik dari sisi biaya, waktu serta tenaga. Dengan alasan membutuhkan sumber daya yang besar maka sensus seringkali dilakukan dalam periodisasi waktu yang lama (di Indonesia 10 tahunan) dan ditujukan terutama untuk “hanya” mendapatkan informasi dasar dan berfungsi dalam pembentukan *frame* (kerangka sampel) untuk keperluan survei. Untuk tujuan efektifitas maka kemudian metode survei menjadi pilihan untuk lebih sering dilakukan sebagai cara untuk mendapatkan suatu penduga parameter.

Berbagai survei di Indonesia pada umumnya digunakan dan dirancang untuk menduga parameter dari suatu data pada level area yang masih relatif besar seperti tingkat provinsi maupun kabupaten. Pada area tersebut sampel yang diperoleh dari survei bisa langsung digunakan untuk estimasi karena sesuai dengan desain samplingnya (*design approach*). Di sisi lain untuk area yang lebih kecil karena ketidakcukupan sampel (jumlah sampel kecil) menyebabkan tidak bisa dilakukan estimasi secara langsung dari data sampel yang didapatkan [14]. Untuk ketersediaan informasi pada level yang lebih kecil dibanding desain samplingnya maka dikembangkan suatu teknik yang dikenal dengan istilah *Small Area Estimation* (SAE).

B. *Small Area Estimation*

Istilah *small area* bisa digunakan untuk menunjukan suatu bagian dari domain populasi. Suatu wilayah dikategorikan sebagai “*small area*” jika ukuran sampel pada wilayah bersangkutan tidak mencukupi untuk mendukung estimasi langsung dengan ketelitian yang masih baik [30]. Istilah lain yang sering digunakan untuk merujuk “*small area*” antara lain *local area*, *subdomain*, *small group*, dan *minor domain*.

Metode SAE merupakan metode estimasi tidak langsung (*indirect estimation*) yang bersifat ‘meminjam’ kekuatan (*borrowing strength*) dari pengamatan sampel area yang berdekatan dengan memanfaatkan informasi tambahan [30]. Informasi tambahan yang dikumpulkan dari luar sampel bertujuan untuk peningkatan efisiensi.

C. Model SAE Berbasis Area

Model SAE secara umum dibagi menjadi dua yaitu SAE berbasis unit dan area, dalam penelitian ini digunakan SAE berbasis area. SAE berbasis area mendasarkan pada tersedianya data pendukung pada level area. Model SAE berbasis area memuat dua model yaitu *sampling* dan *linking model*. *Sampling model* dihitung berdasarkan estimasi *sampling error* dari survei langsung. Sementara *linking model* menghubungkan suatu nilai populasi (parameter) dengan variabel tambahan (*covariate*) dari suatu wilayah atau area tertentu. Karena estimasi berbasis desain survei dimodelkan secara langsung, maka model area level biasanya menghasilkan estimator yang *design-consistent*.

D. Pemodelan *Hierarchical Bayes* pada SAE

Penelitian SAE dengan menggunakan HB sudah mulai banyak digunakan tetapi belum sebanyak SAE dengan menggunakan teknik yang lain seperti EBLUP, BLUP, maupun EB. Menurut Hajarisman [15] pengembangan SAE masih banyak yang meneliti terkait data kontinu, padahal data survei seringkali berbentuk biner, multinomial atau cacahan.

Dengan pemodelan HB, SAE yang melibatkan variabel respon yang berasal dari data kategorik menjadi lebih fleksibel. Untuk kasus data biner, pengguna metode HB untuk SAE telah mulai banyak diteliti. Malec *et al.* Meneliti penggunaan SAE pada data biner dari data Survey kesehatan AS [25]. Xie *et al.* [37] meneliti proporsi orang gemuk di Amerika Serikat. Sementara Mohadjer *et al.* [26] meneliti tentang kemampuan baca tulis penduduk AS [26] dengan HB *linking model*. Di Indonesia juga sudah mulai dikembangkan. Demikian juga You [33] yang menghitung proporsi pengangguran di Canada. Sementara ada juga beberapa penelitian di Indonesia seperti Wulansari [32] yang meneliti pendugaan area kecil proporsi penduduk miskin dengan metode HB menggunakan *Gibb Sampling*. Siswatiningsih, Saefuddin, Notodiputro, Nuryantono, dan Mangku [31] menerapkan model HB dengan fungsi *link logit* untuk *poverty mapping* di Indonesia dan dikombinasikan dengan *Scan statistic*. Penelitian yang memfokuskan metode HB pada data cacahan juga telah dilakukan oleh Li [20] yang meneliti tentang frekuensi kebakaran yang terjadi di Columbia. Untuk model HB dengan menggunakan data biner maupun cacahan, pendekatan GLMMs dalam model digunakan untuk menghubungkan variabel respon dan kovariatnya dengan adanya *link function*.

E. Model HB Area-Level untuk Proporsi

Misalkan y_{ij} adalah variabel biner dari suatu karakteristik yang diteliti (status pengangguran) untuk individu ke- j pada wilayah kecil ke- i (kecamatan), dan N_i adalah ukuran populasi pada wilayah ke- i ($i=1,2,\dots,m; j=1,2,\dots,N_i$). Jika ingin diestimasi proporsi populasi dari suatu wilayah kecil yaitu:

$$\theta_i = \sum_{j=1}^{N_i} y_{ij} / N_i \quad (1)$$

maka estimator langsung yang diperoleh dari nilai sampel y_{ij} yang berasal dari suatu desain survei tertentu adalah sebagai berikut.

$$\hat{\theta}_i = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}}, i = 1, 2, \dots, m; j = 1, 2, \dots, n_i \quad (2)$$

dengan asumsi $n_i > 0$ untuk semua wilayah kecil, dimana n_i adalah ukuran sampel untuk setiap wilayah kecil dan w_{ij} adalah penimbang survei pada wilayah tertentu. Untuk desain survei yang EPSEM (*Equal Probability of Selection Method*) maka sampel proporsi bisa dinyatakan sebagai [9][29]:

$$\hat{\theta}_{iw} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, i = 1, 2, \dots, m; j = 1, 2, \dots, n_i \quad (3)$$

Menurut Kish [18], jika misalkan $Var_{\text{dsgn}}(\theta_i)$ dan $Var_{\text{srs}}(\theta_i)$ adalah varian dari parameter θ_i dari desain survei kompleks dan SRS maka $Deff_i = Var_{\text{dsgn}}(\theta_i) / Var_{\text{srs}}(\theta_i)$ adalah perbandingan antara suatu desain survei kompleks dengan desain acak sederhana yang dikenal sebagai *design effect* parameter. Dalam prakteknya *Deff* juga sering tidak diketahui sehingga harus diestimasi.

Dikarenakan jumlah sampel pada "small area" cenderung kecil, maka dugaan proporsi langsung dari sampel sangat tidak stabil dan mempunyai variasi yang sangat tinggi sehingga tidak bisa diabaikan [14]. Dengan alasan itu seringkali pendekatan model menjadi solusi untuk meningkatkan keakuratan sekaligus mengatasi masalah variabilitas tersebut. Menurut Liu [23] untuk pendugaan proporsi suatu parameter dalam pendugaan area kecil seringkali menggunakan model Fay-Herriot.

Model 1: Fay-Herriot (FH)/ Normal-Normal model

$$\text{Sampling model} : \hat{\theta}_{iw} | \theta_i \sim N(\theta_i, \psi_i) \quad (4)$$

$$\text{Linking model} : \theta_i | \beta, \sigma_v^2 \sim N(\mathbf{x}_i' \beta, \sigma_v^2) \quad (5)$$

Dimana \mathbf{x} adalah vektor dari variabel penyerta dari wilayah yang diketahui. Pada model di atas varian dari *sampling* ψ_i diasumsikan diketahui. Model 1 (model Fay-Herriot) termasuk ke dalam "matched model" karena antara *sampling* dan *linking model* dapat digabungkan menjadi satu persamaan linier campuran [17]. Selain asumsi varian yang diketahui, model Fay-Herriot mempunyai keterbatasan khususnya jika memodelkan proporsi, yaitu pada *linking model* diasumsikan bahwa "true" proporsi (θ_i) mengikuti distribusi normal padahal domain proporsi adalah pada daerah (0,1). Dengan asumsi normalitas ini tidak menjamin untuk distribusi posterior dari θ_i untuk berada pada rentang 0 dan 1.

Oleh karena itu kemudian dikembangkan model 2 (*Normal-Logistic model*).

Model 2: Normal-Logistic model (NL)

$$\text{Sampling model} : \hat{\theta}_{iw} | \theta_i \sim N(\theta_i, \psi_i) \quad (6)$$

$$\text{Linking model} : \text{logit}(\theta_i) | \beta, \sigma_v^2 \sim N(\mathbf{x}_i' \beta, \sigma_v^2) \quad (7)$$

Menurut Ha *et al.* [13] untuk mengestimasi proporsi pada wilayah kecil, asumsi bahwa baik pada *sampling model* maupun *linking model* mengikuti distribusi normal bisa menjadi kurang beralasan khususnya ketika *true proportion* mendekati 0 atau 1. Sementara Lahiri dan Rao [30] berargumen bahwa asumsi normalitas pada *sampling model* cukup beralasan dibandingkan dengan hal yang sama dilakukan untuk *linking model*. Hal ini dikarenakan pada penduga proporsi dari sampel, sebagaimana rata-rata, cenderung mempunyai *mild central limit effect*, sementara pada *true proportion* tidak memiliki hal ini.

Dengan argumentasi pada paragraf sebelumnya maka untuk pengembangan model Fay-Herriot (model 2) mempertimbangkan penggunaan asumsi normalitas untuk level *sampling model* dan non-normalitas untuk distribusi dari *linking model* nya. Pada penelitian ini distribusi *linking model* yang diusulkan adalah dengan distribusi logistik sebagaimana disarankan oleh Rao [30], Lahiri [19] serta Ha [14]. Penggunaan fungsi hubung *logit*(θ_i) juga sesuai dengan teori *Generalized Linier Models* (GLMs) yang menyatakan bahwa fungsi hubung *logit* menjadi salah satu *link functions* yang bisa digunakan untuk transformasi variabel respon sehingga hubungan antara variabel respon (proporsi) dengan variabel kovariatnya menjadi lebih tepat. Fungsi *logit*(θ_i) akan menjamin bahwa estimasi dari proporsi θ_i selalu jatuh pada daerah (0,1). Sementara itu penggunaan distribusi normal sebagai prior *logit* telah menjadi perhatian sejak tahun 1970-an. Seringnya penggunaan normal untuk prior *logit* pada model hirarki karena memungkinkan fleksibilitas dan cakupan yang lebih besar untuk generalisasi [1].

Terdapat dua perhatian utama pada Model 1 dan Model 2 di atas. Pertama adalah bahwa kedua model diasumsikan mempunyai varians *sampling* yang telah diketahui. Pendekatan yang relatif mudah untuk memperkirakan varian adalah dengan estimasi secara langsung dari data sampel, tetapi ketika ukuran sampel daerah penelitian jumlahnya kecil maka menggunakan estimasi langsung dari sampel menjadi kurang tepat karena cenderung bias dan tidak stabil [13]. Sebagai alternatifnya digunakan estimasi untuk θ_i (*true* proporsi kecamatan ke-*i*) antara lain dengan menggunakan penduga sintetis θ_i dengan $\hat{\theta}_{j(i)}^{rgn}$ yaitu estimasi langsung dari suatu nilai proporsi pada daerah yang lebih luas (sensus domain-kabupaten) untuk kemudian digunakan sebagai penduga langsung dari proporsi pada wilayah yang lebih kecil (kecamatan).

Sehingga varian dari area kecil ke-*i* dapat diformulasikan sebagai berikut.

$$\psi_i = \frac{\theta_i(1-\theta_i)}{n_i} Deff_i \quad (8)$$

Dengan penduga variannya dapat didekati dengan formula:

$$\hat{\psi}_i = \frac{\hat{\theta}_{j(i)}^{rgn}(1-\hat{\theta}_{j(i)}^{rgn})}{n_i} deff_{iw} \quad (9)$$

$$deff_{iw} = \frac{\text{var}_{dsgn}(\hat{\theta}_{j(i)}^{rgn})}{\text{var}_{srs}(\hat{\theta}_{j(i)}^{rgn})} \quad (10)$$

dimana $deff_{iw}$ adalah penduga dari parameter $Deff_i$. Penggunaan $deff_{iw}$ ini maka berimplikasi pada asumsi bahwa *true design effect* untuk semua kecamatan dalam satu kabupaten dianggap sama.

Sebagaimana disebutkan di atas bahwa nilai varian ψ_i dari Model 1 dan 2 pada *sampling model* diasumsikan diketahui. Penelitian model HB dengan menggunakan asumsi varian tidak diketahui telah beberapa dilakukan seperti dalam penelitian You [33], Liu, *et al.* [22], dan Ha, *et al.* [13] baik untuk parameter *mean* maupun proporsi.

Dua model berikut ini mengasumsikan bahwa varian *sampling* tidak diketahui.

Model 3: Normal-logistic random sampling variance (NLRs)

$$\text{Sampling model} : \hat{\theta}_{iw} | \theta_i \sim N(\theta_i, \psi_i) \quad (11)$$

$$\text{Linking model} : \text{logit}(\theta_i) | \beta, \sigma_v^2 \sim N(\mathbf{x}_i' \beta, \sigma_v^2) \quad (12)$$

Model 4: Beta-Logistic random sampling variance (BLRs)

$$\text{Sampling model} : \hat{\theta}_{iw} | \theta_i \sim \text{beta}(a_i, b_i) \quad (13)$$

$$\text{Linking model} : \text{logit}(\theta_i) | \beta, \sigma_v^2 \sim N(\mathbf{x}_i' \beta, \sigma_v^2) \quad (14)$$

Untuk mengakomodasi distribusi *sampling* yang tidak normal pada parameter proporsi, maka sebagai alternatif bisa digunakan asumsi distribusi *Beta* pada level *sampling model* seperti pada Model 4 Level 1. Distribusi ini dipilih karena memiliki sifat berada pada range (0,1) yang sesuai dengan sifat dari proporsi. Selain itu, distribusi *Beta* juga bisa mencakup banyak kelas distribusi serta memungkinkan bentuk yang asimetris [23]. Liu *et al.* [22] pada kedua model di atas (Model 3 dan 4) mengusulkan pendekatan untuk menduga nilai varians dengan menggunakan persamaan:

$$\psi_i = \left(\frac{\theta_i(1-\theta_i)}{n_i} \right) deff_{iw} \quad (15)$$

Sedangkan untuk parameter a_i dan b_i pada distribusi *Beta* diperkirakan dengan menggunakan formula:

$$a_i = \theta_i \left(\frac{n_i}{\text{deff}_{iw}} - 1 \right), \text{ dan } b_i = (1 - \theta_i) \left(\frac{n_i}{\text{deff}_{iw}} - 1 \right) \quad (16)$$

Pada proses penghitungan distribusi posterior seringkali dijumpai penyelesaian yang tidak *close-form* atau sangat rumit sehingga perlu bantuan numerik untuk memudahkan penyelesaian. Dalam analisis bayesian penghitungan numerik ini dapat dilakukan dengan menggunakan *Marcov Chain Monte Carlo* (MCMC). Dengan metode MCMC ini dibangkitkan sampel dari sembarang fungsi kepekatan posterior $\pi(\theta|y)$ dimana y adalah *direct* proporsi ($\hat{\theta}$) dan kemudian digunakan untuk menentukan nilai harapan dari besaran posterior yang akan dikaji. Algoritma yang sering digunakan pada proses MCMC adalah *Metropolis-Hasting* dalam *Gibb sampling*. Dalam proses MCMC, secara umum keempat model HB di atas dalam penelitian ini menggunakan distribusi *prior* yang noninformatif sebagai cerminan kurangnya informasi untuk parameter. Untuk parameter model β menggunakan *flat prior* $\pi(\beta) \propto 1$ dan σ_v^2 adalah *uniform* (0,100) seperti yang direkomendasikan oleh Gelman [10].

F. Konsep Pengangguran

Menurut BPS tingkat pengangguran didefinisikan sebagai banyaknya jumlah angkatan kerja yang tidak bekerja dan aktif mencari pekerjaan. Sejak tahun 2001, pengertian pengangguran yang digunakan dalam sakernas merupakan definisi luas. Pengangguran mencakup empat kriteria yaitu: mencari pekerjaan, mempersiapkan usaha, merasa tidak mungkin mendapatkan pekerjaan (putus asa) dan sudah diterima bekerja tapi belum mulai bekerja.

Pada tahap simulasi dalam tulisan ini digunakan data hasil Sensus Penduduk 2010 (SP2010). Dalam SP2010 konsep pengangguran sedikit berbeda dengan penyederhanaan pertanyaan dalam kuesioner. Kriteria pengangguran yang mencari kerja dan mempersiapkan usaha tergabung menjadi satu, sementara kriteria putus asa dan belum mulai bekerja hanya diwakili oleh pertanyaan tentang kesediaan bekerja jika ada pekerjaan yang disediakan.

III. METODOLOGI PENELITIAN

A. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder hasil Sensus Penduduk 2010 (SP2010) dan Podes 2011. Data ini digunakan untuk melakukan simulasi kinerja model HB. Penggunaan data Podes 2011 sebagai variabel penyerta mengasumsikan bahwa sejak tahun 2010 sampai 2011 tidak terjadi perubahan yang berarti pada variabel yang digunakan. Sementara pada tahap implementasi model terbaik digunakan data Sakernas 2014 dan Podes 2014.

B. Variabel Penelitian

Adapun variabel prediktor yang digunakan pada penelitian ini antara lain:

Variabel	Deskripsi
X1	Persentase rumah tangga pertanian di tiap kecamatan
X2	Persentase desa/kelurahan yang terdapat pasar di tiap kecamatan
X3	Rasio industri kecil di tiap kecamatan
X4	Persentase desa yang terdapat SLTA/Sederajat di tiap kecamatan
X5	Rasio penduduk Laki-laki di tiap kecamatan
X6	Persentase desa yang terdapat program dana bergulir bantuan modal usaha non pertanian

C. Tahapan Penelitian

Selama ini banyak penelitian mengenai SAE terutama dalam negeri tidak melakukan validasi yang cukup memadai [38]. Dalam penelitian ini akan dilakukan validasi hasil pendugaan SAE dengan model HB dengan nilai "true" parameter dari hasil sensus. Validasi ini digunakan untuk memverifikasi apakah suatu dugaan parameter SAE sudah cukup baik atau tidak atau dengan kata lain apakah nilai dugaan parameter memiliki bias yang masih bisa dimaklumi atau tidak jika dibandingkan dengan nilai "true parameter"-nya (nilai hasil sensus pada saat yang sama). Untuk itu dalam penelitian ini digunakan metode simulasi sederhana untuk menghitung dugaan proporsi pengangguran pada tingkat kecamatan dengan menggunakan data SP2010. Data SP2010 dianggap sebagai *frame* populasi sebagai domain penelitian. Beberapa tahapannya:

1. Preprocessing Data

Sebelum melakukan pengolahan data yang siap akan digunakan dalam pemodelan dengan HB, terlebih dulu dilakukan beberapa pengolahan awal dari data yang digunakan.

2. Penarikan sampel terhadap data SP2010 (*Sampling*).

Dalam penelitian ini dilakukan simulasi sederhana pengambilan sampel. Proses *sampling* ini dilakukan untuk tujuan validasi dari dugaan parameter proporsi yang didapatkan dengan nilai "true" proporsi dari data SP2010. Desain *sampling* yang digunakan adalah desain kompleks dengan metode *two stage sampling*.

3. Melakukan pemodelan HB

Lakukan pemodelan HB menggunakan keempat model (FH, NL, NL_{rs}, BL_{rs}).

4. Pemilihan model terbaik

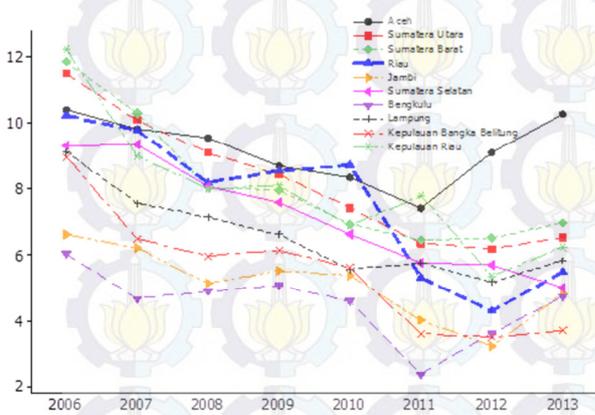
Melakukan validasi dengan cara membandingkan nilai hasil pendugaan model HB dengan nilai 'true value' populasi SP2010. Untuk tujuan ini dilakukan dengan menghitung nilai rata-rata bias (*Overall Average Bias-OAB*), rata-rata bias absolut (*Overall Average Absolute Deviation-OAAD*) dan relatif rata-rata bias absolut (*Overall Absolute Average Relative Deviation-OAARD*). Selain itu juga melihat dari arata-rata nilai *DIC* dari masing-masing model HB. Ukuran *noncoverage rate* terhadap nilai *true value* pada *credible interval* juga menjadi pembanding kebaikan model SAE.

5. Implementasi Model HB SAE terbaik untuk pendugaan proporsi pengangguran tahun 2014 dengan langkah-langkah

IV. HASIL DAN PEMBAHASAN

A. Gambaran TPT Provinsi Riau

Berdasarkan hasil pendataan Survei Angkatan Kerja Nasional (Sakernas) terlihat bahwa secara umum TPT provinsi se-Sumatera cenderung semakin menurun termasuk Provinsi Riau. Pengangguran terbuka di Provinsi Riau pada tahun 2006 mencapai 10,24 persen dan cenderung terus menurun sampai tahun 2014 sehingga menjadi tinggal separuhnya yaitu 5,50 persen (Gambar 1). Terdapat keadaan yang cukup unik dari gambaran TPT Sumatera, terlihat bahwa provinsi-provinsi kecil cenderung mempunyai TPT yang relatif kecil seperti Provinsi Jambi, Bengkulu dan Kepulauan Bangka Belitung. Keadaan ini bisa jadi dipicu oleh fenomena pengangguran yang lebih banyak terjadi di wilayah perkotaan dimana banyak terdapat pada provinsi-provinsi yang besar.



Gambar 1. Perkembangan TPT Provinsi-provinsi Se-Pulau Sumatera Tahun 2006-2013

B. Karakteristik Populasi

Pada penelitian ini dilakukan pemodelan SAE untuk menduga parameter proporsi pengangguran pada tingkat kecamatan di Provinsi Riau. Untuk kepentingan validasi model hasil penduga proporsi ini dilakukan perbandingan dengan nilai 'true' proporsi. Oleh karena itu diperlukan data yang berasal dari sensus untuk mendapatkan parameter dari proporsi yang diinginkan (dalam hal ini parameter proporsi pengangguran).

Dalam penelitian ini data sensus yang digunakan adalah data Sensus Penduduk 2010 (SP2010). Data SP2010 yang digunakan dalam penelitian ini terdiri dari seluruh rumah tangga dengan hanya diambil karakteristik anggota rumah tangga yang termasuk kedalam kelompok angkatan kerja di seluruh wilayah Provinsi Riau. Dari Tabel 1 diketahui bahwa populasi meliputi seluruh rumah tangga hasil SP2010 di seluruh kabupaten di Provinsi Riau sebesar 1,2 juta rumah tangga. Domain untuk SAE adalah semua kecamatan. Dari data SP2010 diperoleh juga parameter proporsi pengangguran untuk kecamatan sebagai domain SAE. Data proporsi ini akan

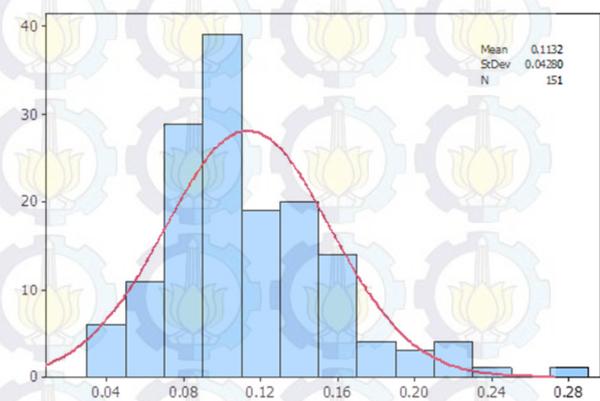
digunakan sebagai variabel untuk validasi hasil dari pemodelan SAE HB.

Tabel 1 Karakteristik Populasi yang Digunakan dalam Simulasi Sederhana (SP2010).

Id_Kab	Kabupaten/Riau	Jumlah Kec	Jml RT	Proporsi
1401	Kuantan Sengingi	12	71.002	0,076
1402	Indragiri Hulu	14	86.188	0,104
1403	Indragiri Hilir	20	158.249	0,104
1404	Pelalawan	12	74.772	0,088
1405	Siak	14	89.368	0,118
1406	Kampar	20	160.511	0,130
1407	Rokan Hulu	16	113.974	0,103
1408	Bengkalis	8	113.598	0,143
1409	Rokan Hilir	13	124.667	0,125
1410	Kepulauan Meranti	5	38.271	0,119
1471	Kota Pekanbaru	12	199.569	0,133
1473	Kota Dumai	5	57.217	0,155
1400	Riau	151	1.287.386	0,118

Sumber: BPS, data diolah

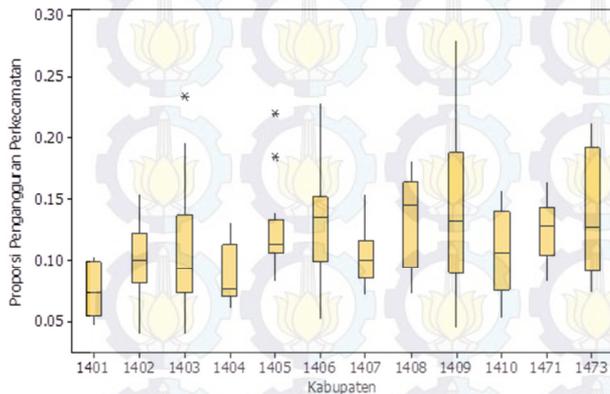
Secara keseluruhan dalam satu Provinsi Riau, terlihat bahwa rata-rata proporsi pengangguran (TPT) kecamatan di Provinsi Riau sebesar 0,11 atau 11,33 persen. Proporsi tertinggi sebesar 0,28 sementara yang terendah adalah 0,04. Dapat diketahui juga bahwa nilai varian dari TPT ini masih relatif kecil sebesar 0,002 yang bisa menjadi indikator umum bahwa tingkat pengangguran kecamatan di Provinsi Riau masih relatif homogen. Bentuk distribusi proporsi populasi ini mendekati normal meskipun sedikit menceng kanan. Hasil dari *fitting* distribusi dengan *easyfit* juga menunjukkan bahwa distribusi dari proporsi pengangguran hasil SP2010 masih terhitung berdistribusi normal dengan *mean* 0,11 dan varian sebesar 0,043.



Gambar 2. Output Deskriptif Karakteristik Proporsi Pengangguran Kecamatan di Provinsi Riau Berdasarkan SP2010

Dilihat tiap kabupaten, nilai proporsi kecamatan hasil SP2010 sesuai dengan Gambar 3 cukup bervariasi. Untuk Kabupaten Rokan Hilir (1409) terlihat proporsi pengangguran

kecamatan paling besar simpangannya, begitu juga pada Kabupaten Kampar (1406) dan Kota Dumai (1473). Kabupaten Siak (1405) proporsi pengangguran tiap kecamatannya cenderung lebih homogen, meskipun terdapat beberapa proporsi yang terindikasi outlier. Sementara itu, Kabupaten Kuantan Singingi (1401) terlihat pada saat yang sama memiliki sebaran proporsi pengangguran kecamatan yang terendah di antara kabupaten/kota yang lain dengan nilai median pada kisaran 7 persen.



Gambar 3. Box-Plot Proporsi Pengangguran Kecamatan Tiap Kabupaten Di Provinsi Riau Berdasarkan SP2010

C. Penarikan Sampel

Dalam penelitian ini dilakukan penarikan sampel terhadap populasi dari data SP2010. Sebagaimana telah dijelaskan sebelumnya bahwa tujuan dari penarikan sampel ini adalah agar parameter proporsi pengangguran dari hasil model SAE HB dapat diperbandingkan atau divalidasi dengan besarnya nilai proporsi ('true' proporsi) dari populasinya. Pada tahap ini dilakukan penarikan sampling dua tahap dengan mengabaikan desa dan kota (semuanya diikutkan) dengan tujuan agar jumlah kecamatan yang tidak terpilih sebagai sampel dapat diminimalkan.

Sampai saat ini total jumlah Blok Sensus (BS) yang sering digunakan oleh BPS Provinsi Riau untuk survei ketenagakerjaan (Sakernas) berjumlah di atas 500-an BS yang terdistribusi pada masing-masing kabupaten. Penarikan sampel tiap kabupaten dilakukan secara independen (terpisah). BS ini digunakan sebagai *Primary Sampling Unit* (PSU) untuk kemudian dipilih dari masing-masing BS terpilih sejumlah rumah tangga (dalam hal ini 10 rumah tangga tiap BS). Tabel 2 menunjukkan bahwa secara umum jumlah sampel BS yang dipilih relatif proporsional terhadap jumlah BS pada tiap-tiap kabupaten.

Hasil dari penarikan sampel sebanyak sepuluh kali untuk masing-masing kabupaten dengan pendekatan *two stage sampling* disajikan dalam Tabel 3. Meskipun dalam setiap kali penarikan sampel besarnya proporsi bervariasi tetapi secara rata-rata dapat dilihat bahwa proporsi pengangguran untuk sepuluh sampel hampir mendekati nilai proporsi dari populasinya.

Tabel 2. Deskripsi *Frame* Sampling Dua Tahap tiap Kabupaten

Id_Kab	Kabupaten/Riau	Frame SP2010		Sampel	
		BS	Rata-rata Jumlah RT	BS	RT tiap BS
1401	Kuantan Singingi	950	75	35	10
1402	Indragiri Hulu	1096	79	42	10
1403	Indragiri Hilir	1918	82	54	10
1404	Pelalawan	980	76	41	10
1405	Siak	1146	78	42	10
1406	Kampar	2224	72	54	10
1407	Rokan Hulu	1490	76	46	10
1408	Bengkalis	1247	91	46	10
1409	Rokan Hilir	1456	86	46	10
1410	Kepulauan Meranti	544	70	38	10
1471	Kota Pekanbaru	2531	79	54	10
1473	Kota Dumai	614	93	38	10

Sumber: BPS

Tabel 3. Perbandingan Rata-rata Proporsi Sampel dengan Proporsi Populasi

Id_Kab	Kabupaten/Riau	Rata-rata Proporsi Sampel	Proporsi Populasi (SP2010)
1401	Kuantan Singingi	0.072	0.076
1402	Indragiri Hulu	0.100	0.104
1403	Indragiri Hilir	0.103	0.104
1404	Pelalawan	0.091	0.088
1405	Siak	0.117	0.118
1406	Kampar	0.132	0.130
1407	Rokan Hulu	0.101	0.103
1408	Bengkalis	0.149	0.143
1409	Rokan Hilir	0.123	0.125
1410	Kepulauan Meranti	0.125	0.119
1471	Kota Pekanbaru	0.121	0.133
1473	Kota Dumai	0.161	0.155

D. Perbandingan Model HB

Terdapat beberapa kriteria pemilihan model terbaik yang digunakan yaitu nilai DIC, OAB, OAAD, OAARD dan *noncoverage rate* yaitu persentase *true value* yang tidak tercakup dalam 95% nilai *credible interval* (95% *credible interval fail to cover true value*). Untuk membandingkan model yang paling *fit* menerangkan variabel responnya maka model BLrs dengan nilai *DIC* terkecil dianggap model yang terbaik. Meskipun demikian model ini masih tidak lebih baik dibandingkan dengan model NLrs dalam rangka memprediksi nilai *true* proporsi pengangguran tiap kecamatan yang diperlihatkan oleh nilai dari OAARD dan 95% *fc*.

Tabel 4. memperlihatkan beberapa ukuran perbandingan model SAE HB. Terlihat bahwa semua model menghasilkan nilai OAB yang negatif dan mendekati nol yang menandakan bahwa hasil prediksi proporsi pengangguran berdasarkan model HB secara umum masih *under estimate* terhadap proporsi pengangguran SP2010 meskipun relatif kecil. Untuk membandingkan model yang paling *fit* menerangkan variabel responnya maka model BLrs dengan nilai *DIC* terkecil dianggap model yang terbaik. Meskipun demikian, menurut Congdon [4] model yang *fit* bukan berarti

model yang terbaik jika tidak bisa merepresentasikan nilai parameter yang diduga dengan baik.

Tabel 4. Perbandingan Keempat Model HB SAE dan Direct Estimate (DE) dari Pengambilan 10x Sampling

Model	DIC	OAB	OAAD	OAARD	95% fc
FH	-312.78	-0.004	0.0314	0.3087	0.06181
NL	-321.71	-0.007	0.0318	0.3011	0.04194
NLrs	-341.01	-0.007	0.0325	0.3019	0.03532
BLrs	-426.13	-0.002	0.0319	0.3153	0.04783
DE	-	-0.004	0.0601	0.5716	-

Nilai dari OAARD dan 95% fc menjadi cara untuk membandingkan model terbaik dalam merepresentasikan parameter yang diduga. Terlihat bahwa dari Tabel 4 Model FH menjadi model yang relatif paling berkinerja kurang baik dalam memprediksi nilai parameter proporsi pengangguran tingkat kecamatan dengan nilai OAARD yang tinggi dan tingkat *noncoverage* terhadap *true value* yang paling besar. Meskipun demikian, model FH ini masih jauh lebih baik dibandingkan dengan *Direct Estimate* (DE) yang menghasilkan nilai OAARD yang jauh lebih besar.

Sementara itu Model NL dan BLrs terlihat berkinerja hampir sama. Kedua model tersebut tingkat peluang *noncoverage* untuk 95 persen *credible interval* tidak memuat *true value* sekitar 4-5 persen. Sementara itu, untuk model NLrs peluang nilai parameter proporsi pengangguran yang tidak bisa tercakup dalam 95 persen *credible interval* hanya sebanyak 3,5 persen atau dengan kata lain model ini mampu memprediksi proporsi pengangguran tingkat kecamatan sekitar 96 persen. Dengan kriteria perbandingan model di atas terutama keakuratan untuk memprediksi *true value*, dapat disimpulkan bahwa model yang relatif lebih baik untuk memprediksi besarnya proporsi pengangguran tingkat kecamatan di Provinsi Riau adalah model NLrs. Secara umum berarti bahwa model NLrs mampu untuk memberikan hasil prediksi yang paling mendekati nilai parameter proporsi pengangguran.

E. Prediksi Proporsi Pengangguran 2014

Dari hasil simulasi sederhana di atas diperoleh bahwa model yang terbaik untuk memprediksi proporsi pengangguran tingkat kecamatan di Provinsi Riau adalah model NLrs. Dengan demikian model ini yang akan digunakan untuk memprediksi tingkat atau proporsi pengangguran tingkat kecamatan pada tahun 2014 di Provinsi Riau berdasarkan data Sakernas 2014. Sakernas dilakukan dengan pengambilan sampel dua tahap berstrata desa dan kota secara independen. Dalam penelitian ini hanya akan dilakukan prediksi pengangguran kecamatan pada strata desa pada tahun 2014.

Tabel 5. terlihat bahwa semua parameter untuk memprediksi proporsi pengangguran sudah diestimasi dengan akurat dengan nilai rasio MC *error* yang sangat kecil (dibawah 0,05%), hal ini bisa menjadi indikasi konsistensi dengan rantai MCMC yang konvergen [24][27]. Bahkan Ntzoufras [27] berani mengatakan bahwa jika nilai ini lebih kecil dari 0,1% maka bisa langsung diasumsikan bahwa kekonvergenan telah tercapai. Meskipun demikian, ternyata pada model ini juga

parameter variabel penyerta yang diharapkan memberikan tambahan informasi pada proses estimasi model SAE HB terlihat belum signifikan menerangkan variasi dari proporsi pengangguran terbukti pada alpha 0,05 *credible interval* parameter mengandung nilai nol yang menandakan bahwa variabel tidak signifikan. Keadaan ini menandakan bahwa pemodelan pengangguran strata desa tingkat kecamatan ini pun hanya dipengaruhi oleh faktor random antar kecamatan.

Tabel 5. Hasil Estimasi Parameter Model NLrs untuk Sampel Sakernas 2014 (Strata Desa)

Node	Mean (β)	Sd (β)	MC error	2,50 %	media n	97,50 %	Rasio MC Error
beta[1]	-0.0060	0.0094	0.0001	-0.025	-0.0058	0.012	0.0103
beta[2]	0.0011	0.0086	0.0001	-0.016	0.0013	0.017	0.0104
beta[3]	0.0084	0.0085	0.0001	-0.008	0.0082	0.026	0.0091
beta[4]	0.0117	0.0084	0.0001	-0.003	0.0111	0.030	0.0093
beta[5]	-0.0025	0.0076	0.0001	-0.019	-0.0021	0.012	0.0092
beta[6]	-0.0070	0.0076	0.0001	-0.022	-0.0072	0.009	0.0098
beta0	-3.5280	0.2716	0.0028	-4.129	-3.5020	-3.069	0.0103
sigma	0.9574	0.2800	0.0028	0.520	0.9201	1.628	0.0101

Dengan demikian prediksi pengangguran berdasarkan data Sakernas 2014 dengan model NLrs menjadi kurang lengkap karena konsep *borrowing strength* yang seharusnya diharapkan didapatkan dari variabel penyerta yang digunakan ternyata tidak signifikan. Sehingga konsep *borrowing strength* dalam model SAE untuk memprediksi pengangguran (strata desa) pada tahun 2014 ini hanya dipengaruhi oleh faktor random pada setiap kecamatan yang berasal dari distribusi variabel pengamatan seluruh kecamatan di Provinsi Riau.

Dari tabel di atas setidaknya bisa ditulis persamaan untuk memprediksi proporsi pengangguran tingkat kecamatan adalah:

$$\text{logit}(\theta_i) = -3,528 - 0,006x_1 + 0,001x_2 + 0,008x_3 + 0,012x_4 - 0,003x_5 - 0,007x_6 + v_i$$

Dengan model tersebut maka diprediksi proporsi pengangguran tiap kecamatan di Provinsi Riau pada tahun 2014. Dalam penelitian ini sampel Sakernas (strata desa) hanya mencakup 79 kecamatan sementara 71 kecamatan yang lainnya tidak terdapat sampel. Untuk itu melalui *WinBugs* dilakukan estimasi *predictive distribution* untuk kecamatan yang tidak terkena sampel dengan logika perlakuan yang sama dengan kasus pada *missing value* [27]. Meskipun nilai dari parameter regresinya tidak signifikan secara statistik, tetapi karena tujuan model adalah untuk memprediksi proporsi pengangguran maka variabel-variabel tersebut tetap diikutsertakan dengan harapan untuk meningkatkan akurasi nilai prediksi.

Tabel 6 menunjukkan beberapa contoh hasil prediksi pengangguran kecamatan strata desa Provinsi Riau khususnya di Kabupaten Kuantan Singingi (1401) dan Indagiri Hulu (1402) pada tahun 2014.

Tabel 6. Hasil Prediksi Pengangguran Strata Desa Beberapa Kecamatan di Kabupaten Kuansing (1401) dan Indragiri Hulu (1402) Tahun 2014 dengan Model NLrs

Id_Kec	Proporsi pengangguran	SD	MC error	2.50%	Median	97.50 %	Rasio MC error
1401010	0.0288	0.0266	0.0003	0.0015	0.0214	0.1002	0.0099
1401011	0.0505	0.0596	0.0006	0.0032	0.0327	0.2072	0.0094
1401012	0.0904	0.0928	0.0009	0.0067	0.0622	0.3450	0.0097
1401013	0.0803	0.0858	0.0007	0.0058	0.0550	0.3194	0.0086
1401020	0.0795	0.0413	0.0005	0.0285	0.0698	0.1878	0.0111
1401021	0.0910	0.0532	0.0005	0.0289	0.0774	0.2309	0.0097
1401030	0.0590	0.0650	0.0006	0.0046	0.0403	0.2248	0.0096
1401031	0.0677	0.0722	0.0007	0.0044	0.0460	0.2634	0.0099
1401040	0.0299	0.0212	0.0002	0.0050	0.0246	0.0842	0.0096
1401050	0.0995	0.0492	0.0005	0.0365	0.0885	0.2259	0.0107
1401051	0.0673	0.0349	0.0003	0.0237	0.0592	0.1551	0.0090
1401052	0.0440	0.0254	0.0003	0.0116	0.0385	0.1083	0.0100
1401053	0.0500	0.0595	0.0006	0.0027	0.0327	0.2073	0.0105
1401060	0.0476	0.0243	0.0002	0.0161	0.0425	0.1093	0.0099
1401061	0.0478	0.0249	0.0003	0.0157	0.0427	0.1105	0.0105
1402010	0.0755	0.0805	0.0008	0.0056	0.0521	0.2897	0.0097
1402011	0.0621	0.0673	0.0007	0.0043	0.0413	0.2420	0.0097
1402020	0.0397	0.0385	0.0004	0.0021	0.0288	0.1466	0.0099
1402021	0.0360	0.0357	0.0004	0.0046	0.0254	0.1306	0.0115
1402022	0.1154	0.0763	0.0008	0.0324	0.0942	0.3240	0.0104

KESIMPULAN DAN SARAN

Kesimpulan

Terdapat beberapa kesimpulan yang dapat diambil dari hasil dan pembahasan dalam penelitian ini yaitu:

1. Penarikan sampel dengan jumlah sampel yang kecil cenderung membuat hasil estimasi langsung proporsi pengangguran kecamatan sangat berfluktuasi dan tidak stabil. Hal ini berbeda jika proporsi pengangguran dihitung untuk tingkat kabupaten. Dengan jumlah sampel yang cukup memadai estimasi langsung untuk tingkat kabupaten cenderung lebih stabil dan secara rata-rata mendekati nilai *true value*-nya.
2. Variabel-variabel dari Podes 2014 belum bisa memberikan pengaruh yang signifikan terhadap prediksi proporsi pengangguran tingkat kecamatan. Dengan demikian prediksi pengangguran di kecamatan hanya mendasari pada distribusi proporsi dalam satu provinsi (faktor random) belum bisa mengakomodir konsep *borrowing strenght* antar kecamatan yang berasal dari variabel penyerta.
3. Meskipun model BLrs memiliki nilai DIC yang lebih besar dibanding model NLrs, tetapi pada kasus ini model NLrs terbukti lebih baik untuk memprediksi proporsi pengangguran tingkat kecamatan di Provinsi Riau dilihat dari validasi dengan *Overall Absolute Average Relative Deviation* (OAARD) dan *noncoverage rate* terhadap nilai *true value*nya yang paling kecil diantara model yang lain (*95% credible interval fails to cover true value*).

Saran

Dalam penelitian ini masih banyak permasalahan yang belum dikaji secara mendalam serta terdapat beberapa keterbatasan didalamnya, oleh karena itu penulis memberikan saran sebagai berikut:

1. Untuk mendapatkan pemodelan yang lebih baik dalam memprediksi proporsi pengangguran perlu untuk mencari variabel penyerta selain variabel dari Podes. Hal ini untuk mengantisipasi adanya korelasi yang sangat kecil antara variabel podes dengan proporsi pengangguran sehingga dalam model tidak signifikan mempengaruhi prediksi.
2. Dalam menyimpulkan model NLrs sebagai model terbaik dibandingkan model FH, NL, dan BLrs masih perlu diteliti kembali mengingat simulasi yang dilakukan hanya sebanyak 10 kali. Dalam penelitian selanjutnya disarankan untuk memperbanyak jumlah simulasi. Selain itu perlu juga dicobakan pada desain sampling yang berbeda untuk mendapatkan hasil yang lebih komprehensif.
3. Perlu untuk membandingkan keempat model HB SAE yang telah dibahas dengan model SAE yang lain seperti model HB dengan tranformasi *arcsine* untuk penduga varian sampling ataupun pendekatan binomial dengan koreksi sampling yang digunakan.

DAFTAR PUSTAKA

- [1] Agresti, A. dan Hitchcock D. (2004), *Bayesian Inference for Categorical Data Analysis. Technical report, Departement of Statistics, University of Florida.*
- [2] Arora, V. dan Lahiri P. (1997), "On the Superiority of the Bayesian Method over BLUP in Small Area Estimation Problems", *Statistical Sinica*, Vol. 7, hal. 1053-1063.
- [3] Chandra, H., Chambers, R., dan Salvati, N. (2009), "*Small Area Estimation of Proportions in Business Surveys*", Working Papers, No.15-09, The University of Wollongong, Australia.
- [4] Congdon, P.D. (2010), *Applied Bayesian Hierarchical Methods*, CRC Press, Boca Raton.
- [5] Cowles, M.K., dan Carlin, B.P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review", *Journal of the American Statistical Association*, Vol. 91, hal. 883-904.
- [6] Cho, M., Eltinge, J.L., Gershunskaya, J., dan Huff, L. (2014), "Evaluation of Generalized Variance Functions in the Analysis of Complex Survey Data", *Journal of Official Statistics*, Vol. 30, hal. 63-90.
- [7] Farrell, P.J. (1995), "Hierarchical Bayes Approach for Estimating Small Area Proportions", *Proceedings of the Survei Research Methods Section*, hal. 63-68.
- [8] Farrell, P.J. (2000), "Bayes Inference for Small Area Proportions", *The Indian Journal of Statistics*, Vol. 62, hal. 402-416.
- [9] Frerich, R.R. (2004), *Equal Probability of Selection*, in Rapid Surveys (unpublished). Department of Epidemiology, UCLA, hal. 1-16.
- [10] Gelman, A. (2006), "Prior Distribution for Variance Parameters in Hierarchical Models". *Bayesian Analysis* 1, Vol. 1, No. 3, hal. 515-534.

- [11] Gelman, A., Carlin, J.B., Rubin., D.B. (2004). *Bayesian Data Analysis 2nd*, CRC Press, Boca Raton.
- [12] Ghosh, M., dan Rao, J.N.K., (1994), "Small Area Estimation: an Appraisal". *Statistical Science*, Institute of Mathematical Statistics, Vol. 9, hal. 55-76.
- [13] Ha, N.S., Lahiri, P., dan Parsons, V. (2014), "Methods and Results for Small Area Estimation using Smoking Data from The 2008 Nasional Health Interview Survey", *Statistics in Medicine*, Vol. 33, hal. 3932-3945.
- [14] Ha, N.S. (2013), *Hierarchical Bayes Estimation of Small Area Means Using Complex Survey Data*. Disertasi P.h.D, University of Maryland.
- [15] Hajarisman, N. (2013), *Pemodelan Area Kecil untuk Menduga Angka Kematian Bayi Melalui Pendekatan Model Regresi Poisson Bayes Berhierarchy Dua Level*, Disertasi, FMIPA IPB, Bogor.
- [16] Hawala, S., dan Lahiri, P. (2012), "A Hierarchical Bayes Estimation of Poverty Rates", *Proceedings of American Statistical Association, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association, Fortcoming.
- [17] Jiang, J., dan Lahiri, P. (2006). "Mixed Model Prediction and Small Area Estimation", *Test*, Vol. 15, 1, hal. 1-96.
- [18] Kish, L. (1965). *Survey sampling*, Jhon Wiley, New York.
- [19] Lahiri, P., Rao, J.N.K. (1995), "Robust Estimation of Mean Squared Error of Small Area Estimators", *Journal of the American Statistical Association*, Vol. 90, Hal. 758-766.
- [20] Li, H. (2006), *Bayesian Hierarchical Models for Spatial Count Data With Application to Fire Frequency in British Columbia*, Thesis, University of Victoria, Columbia.
- [21] Liu, B., Diallo, M. (2013), "Parametric Bootstrap Confidence Interval for Survey-Weighted Small Area Proportions. *Proceedings of Survey Research Methods Section of American Statistical Association*, ASA Section on Survey Research Methods, hal. 109-121.
- [22] Liu, B., Lahiri, P., dan Kalton, G. (2014), "Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions", *Survey Methodology*, Vol. 40, hal. 1-13.
- [23] Liu, B. (2009). *Hierarchical Bayes Estimation and Empirical Best Prediction of Small Area Proportions*, Disertasi P.h.D, University of Maryland.
- [24] Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D. (2013), *The Bugs Book: A Practical Introduction to Bayesian Analysis*, CRC Press, Boca Raton.
- [25] Malec, D., Sedransk, J., Moriarity, C., dan LeClere, F. (1997), "Small Area Inference for Binary Variables in National Health Interview Survey, *Journal of the American Statistical Association*, Vol. 92, hal. 815-826.
- [26] Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., dan Van de Kerckhove, W. (2007), "Hierarchical Bayes Small Area Estimates of Adult Literacy using Unmatched Sampling and Linking Models", *Proceedings of American Statistical Association, Section on Survey Research Methods*.
- [27] Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, Canada: Jhon Wiley, New York.
- [28] Pereira, L.N., Mendes, J.M., Coelho, P.S. (2014), "Model-Based Estimation of Unemployment Rates in Small Area Portugal", *Communication in Statistic-Theory and Methods*, Vol. 42:7. hal. 1325-1342.
- [29] Peters, J., dan Eachus, J.I. (1995), "Achieving Equal Probability of Selection under Various Random Sampling Strategies", *Paediatric and Perinatal Epidemiology*, Vol. 9, hal. 219-224.
- [30] Rao, J.N.K. (2003), *Small Area Estimation*, John Wiley & Son, New York.
- [31] Siswantining, T., Saefuddin, A., Notodiputro, K., A., Nuryantono, N., dan Mangku, I.W. "Adaptation of Hierarchical Bayes SAE to Spatial Satscan", *IOSR Journal of Mathematics (IOSRJM)*, Vol. 2, Hal. 1-8.
- [32] Wulansari, I.K. (2015), *Pendugaan Statistik Area Kecil dengan Metode Hierarchical Bayes (HB) pada Data Respon Biner Menggunakan Gibbs Sampling*, Tesis, Universitas Padjajaran, Bandung.
- [33] You, Y. (2008), "An Integrated Modeling Approach to Unemployment Rate Estimation for Subprovincial Areas of Canada, *Survey Methodologi*, Vol. 34, hal. 19-27.
- [34] You, Y., dan Chapman, B. (2006), "Small Area Estimation using Area Level Models and Estimated Sampling Variances", *Survey Methodology*, Vol. 32, hal. 97-103.
- [35] You, Y., dan Rao, J.N.K. (2002), "Small Area Estimation Using Unmatched Sampling and Linking Models, *Canadian Journal of Statistics*, Vol. 30, hal. 3-15.
- [36] You, Y., Zhou, Q.M. (2011), "Hierarchical Bayes Small Area Estimation under a Spatial Model With Application to Health Survey Data", *Survey Methodology*, Vol. 37, hal. 25-37.
- [37] Xie, D., Raghunathan, T.E., dan Lpkowski, J.M. (2006), "Estimation of The Proportion of Overwight Individuals in Small Areas-a Robust Extension of The Fay-Herriot Model", *Statistics In Medicine*, Vol. 26, hal. 2699-2715.
- [38] Zang, X., Holt, J.B., Yun, S., Lu, H., Greenlund, K.J., dan Croft, J.B. (2015), "Validation of Multilevel Regression and Poststratification Methodology for Small Area Estimation of Health Indicators from the Behavioral Risk Factor Surveillance System", *American Journal of Epidemiology*, hal. 1-11.