Jurnal Matematika UNAND
Vol. **VII** No. **4** Hal. 15 – 21
ISSN : 2303–291X
©Jurusan Matematika FMIPA UNAND

PENDUGAAN ANGKA PENGANGGURAN DI KABUPATEN PADANG PARIAMAN MENGGUNAKAN *SMALL AREA ESTIMATION* DENGAN PENDEKATAN *HIERARCHICAL BAYES* (HB) LOGNORMAL

MIA MAULIANI, MAIYASTRI, RITA DIANA

Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus Unand Limau Manis, Padang, Indonesia
email: miamauliani@ymail.com

Abstrak. Informasi mengenai Tingkat Pengangguran Terbuka (TPT) 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 menjadi sangat penting. Penelitian ini bertujuan untuk medapatkan TPT pada level kecamatan menggunakan metode SAE dengan pendekatan Hierarchical Bayes (HB) lognormal. Hasil estimasi TPT dengan penduga HB menghasilkan nilai standar error yang lebih kecil daripada penduga langsung, sehingga penduga HB lebih baik dibandingkan penduga langsung.

Kata Kunci: Tingkat Pengangguran Terbuka, *Hierarchical Bayes*, Lognormal, *Small Area Estimation*

Diterima	:	*29 November 2018*
Direvisi	:	*3 Desember 2018*
Dipublikasikan	:	*30 Desember 2018*

1. Pendahuluan

Permasalahan pengangguran menjadi permasalahan serius yang terjadi di hampir seluruh wilayah termasuk di Kabupaten Padang Pariaman. Masalah pengangguran tidak hanya terkait dengan masalah ekonomi, tetapi berkaitan erat dengan masalah-masalah sosial. Informasi mengenai Tingkat Pengangguran Terbuka (TPT) 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 menjadi sangat penting. Untuk itu, diperlukan metode estimasi yang dapat memberikan tingkat akurasi yang lebih baik yaitu dengan mengkombinasikan antara data survei dengan data pendukung lainnya. Salah satu metode yang sering digunakan adalah *Small Area Estimation* (SAE). *Small Area Estimation* (SAE) adalah sebuah metode untuk memenuhi permintaan akan statistik pada *small area* yang akurat ketika hanya tersedia sampel yang sangat kecil untuk area tersebut atau bahkan untuk daerah yang tidak terambil sebagai sampel [1].

Terdapat beberapa metode yang digunakan dalam SAE, salah satunya dengan pendekatan hierarchical bayes (HB). Model HB memiliki keuntungan yaitu dapat

mengatasi permasalahan pemodelan hirarki untuk jumlah data yang sedikit dan tidak seimbang baik pada tingkat bawah maupun pada tingkat yang lebih tinggi [2]. Berdasarkan hasil uji kesesuaian distribusi terhadap data tingkat pengangguran Kabupaten Padang Pariaman menggunakan uji Kolmogorov Smirnov, diperoleh nilai statistik Kolmogorov Smirnov untuk distribusi lognormal adalah 0,12017 dan *p-value* adalah 0,94222. Karena *p-value* lebih besar dari 0,05 maka dapat disimpulkan bahwa tingkat pengangguran di Kabupaten Padang Pariaman mengikuti distribusi lognormal. Oleh karena itu, tingkat pengangguran level kecamatan di Kabupaten Padang Pariaman dapat diduga dengan pendekatan Hierarchical Bayesian Lognormal.

2. Kajian Pustaka

2.1. *Distribusi Lognormal*

Distribusi Lognormal (LN) adalah distribusi kontinu positif, dengan fungsi kepadatan peluang:

$$f(y|\mu,\sigma^2) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2\right), y>0, \mu>0, \sigma>0,$$

dengan μ adalah parameter lokasi (*location*) dan σ adalah parameter skala (*scale*).

2.2. *Metode Bayes*

Misalkan data observasi $y = (y_1, y_2, \cdots, y_n)^T$ mempunyai distribusi tertentu dengan himpunan parameter $\beta = (\beta_1, \beta_2, \cdots, \beta_k)^T$ yang merupakan variabel *random*, maka distribusi *posterior* dari parameter β atau $f(\beta|y)$ dapat dinyatakan dalam bentuk:

$$f(\beta|y) = \frac{f(y|\beta)f(\beta)}{f(y)},$$

dengan $f(y) = \int_{\Omega\beta_1} \cdots \int_{\Omega\beta_k} f(y|\beta)f(\beta)d\beta_k \cdots d\beta_1$ adalah *normalizing constant*, $f(\beta)$ adalah distribusi *prior* dari parameter β yang berisi informasi yang diketahui tentang parameter β sebelum pengamatan dilakukan. $f(y|\beta) = \prod_{i=1}^{n} f(y_i|\beta)$ adalah fungsi *likelihood* data yang berisi informasi sampel.

2.3. *Metode HB dalam SAE*

Model HB yang digunakan pada penelitian ini adalah model lognormal dengan variabel prediktor berbasis area. Model tersebut didefinisikan sebagai [3]:

(i) *Sampling* model
$y_i|\theta - \log normal(\theta, \sigma^2_{[y]})$, $\sigma^2_{[y]}$ adalah ragam dari distribusi lognormal.

(ii) *Linking* model
$\theta = x\beta + v$, dimana $v - N(0, \sigma^2_{[v]}$ dan $\sigma^2_{[v]} = \frac{1}{\tau_{[v]}}$.

(iii) β dan $\sigma^2_{[v]}$ saling independen dengan $\tau_{[v]} \sim unif(0, 100)$ dan $\beta - N(\mu_{[\beta]}, \sigma^2_{[\beta]})$.

3. Data dan Hasil

Sumber data yang digunakan dalam penulisan ini menggunakan mikro data Sakernas 2016, data Podes 2014 dan data pada publikasi Kabupaten Padang Pariaman dalam angka 2016 yang dihasilkan oleh Badan Pusat Statistik. Variabel respon dalam penelitian ini adalah tingkat pengangguran di Kabupaten Padang Pariaman Provinsi Sumatera Barat yang dihitung berdasarkan data Sakernas 2016 melalui metode estimasi langsung. Variabel prediktor yang digunakan adalah Rasio SLTA/Sederajat (X_1), Persentase keluarga pertanian (X_2), Rasio industri mikro kecil (IMK) (X_3), Persentase buruh tani dalam setiap anggota keluarga (X_4) dan Kepadatan penduduk (X_5).

3.1. *Estimasi* Small Area Estimation *Menggunakan Pendekatan HB Lognormal*

Jika Y adalah variabel *random* yang berdistribusi lognormal atau $y_i - LN(\theta, \sigma^2_{[y]})$ dimana $\theta = x\beta + v$, $\sigma^2_{[y]} = \dfrac{1}{\tau_{[y]}}$ dengan $\theta - N(x\beta, \sigma^2_{[v]})$, $v - N(0, \sigma^2_{[v]})$ dan $\sigma^2_{[v]} = \dfrac{1}{\tau_{[v]}}$ maka model *Small Area Estimation* dengan pendekatan HB menggunakan data tersebut juga berdasarkan distribusi lognormal, dengan prior pada model ini yaitu β_j, $\tau_{[v]}$, dan $\tau_{[y]}$. Fungsi likelihoodnya sebagai berikut :

$$
\begin{aligned}
f(y|\beta, \tau_{[v]}, \tau_{[y]}) &= \prod_{i=1}^{n} f(y_i|\beta, \tau_{[v]}, \tau_{[y]}), \\
&= \prod_{i=1}^{n} \left(\left(\frac{\tau_{[y]}^{1/2}}{y_i (2\pi)^{1/2}} \exp\left[-\frac{\tau_{[y]}}{2} (\ln(y_i) - \theta)^2 \right] \right) \times \frac{\tau_{[v]}^{1/2}}{2\pi^{1/2}} \exp\left[-\frac{\tau_{[v]}}{2} (\theta_i - x_{ij}\beta_j)^2 \right] \right), \\
&= \frac{\tau_{[y]}^{1/2} 2\pi^{n/2}}{\prod_{i=1}^{n} y_i} \exp\left[-\frac{\tau_{[y]}}{2} \sum_{i=1}^{n} (\ln(y_i) - \theta)^2 \right] \times \frac{\tau_{[v]}^{n/2}}{2\pi^{n/2}} \exp\left(-\frac{\tau_{[v]}}{2} \sum_{i=1}^{n} \left(\theta_i - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 \right).
\end{aligned}
$$

Tipe distribusi prior yang digunakan adalah conjugate, dengan distribusi prior $\beta_j - N(\mu_{[\beta]_j}, \sigma^2_{[\beta]_j})$ dan $\tau_{[y]} \sim Gamma(a_{(\tau_{[y]})}, b_{(\tau_{[y]})})$ dan $\tau_{[v]} \sim unif(0, 100)$. Kombinasi antar fungsi *likelihood* dan distribusi *prior* gabungan akan membentuk distribusi *posterior* gabungan dari seluruh parameter yang akan di estimasi. Bentuk proposional distribusi posterior sebagai berikut.

$$
\begin{aligned}
f(\beta, \tau_{[v]}, \tau_{[y]}|y) \propto & \frac{\tau_{[y]}^{n/2}}{\prod_{i=1}^{n} y_i} \exp\left(-\frac{\tau_{[y]}}{2} \sum_{i=1}^{n} (\ln(y_i) - \theta)^2 \right) \\
& \times \tau_{[y]}^{n/2} \exp\left(-\frac{\tau_{[v]}}{2} \sum_{i=1}^{n} \left(\theta_i - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 \right) \\
& \times \tau_{[\beta]_j}^{1/2} \exp\left(-\frac{\tau_{[\beta]_j}}{2} (\beta_j - \mu_{[\beta]_j})^2 \right) \\
& \times \tau_{[y]}^{a_{(\tau_{[y]})} - 1} \exp\left(-b_{(\tau_{[y]})} \times \tau_{[y]} \right).
\end{aligned}
$$

Proses penyelesaian distribusi *posterior* marginal setiap parameter diatas mengandung integral dengan dimensi yang cukup tinggi dan kompleks sehingga untuk menyelesaikannya digunakan pendekatan numerik melalui MCMC dan *Gibbs Sampling*. Proses tersebut dilakukan melalui pengambilan sampel secara berulang melalui bentuk *full conditional posterior*. bentuk distribusi *full conditional posterior* setiap parameter yang akan diestimasi adalah sebagai berikut.

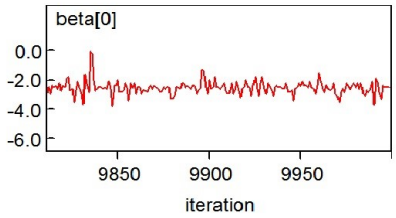
(a) Distribusi *full conditional posterior* untuk parameter $\beta_j f(\beta_j|y,\beta_{\setminus j},\tau_{[v]},\tau_{[y]}) \propto (\tau_{[v]}^{n}\cdot\tau_{[\beta]_j})^{1/2}\exp\left(\frac{1}{2}\left(\tau_{[v]}\sum_{i=1}^{n}\left(\theta_i-\sum_{j=1}^{k}x_{ij}\beta_j\right)^2\right)+\tau_{[\beta]_j}(\beta_j-\mu_{[\beta]_j})^2\right)\right)$.

(b) Distribusi *full conditional posterior* untuk $\tau_{[y]}$ $f(\tau_{[y]}|y,\beta,\tau_{[v]}) \propto \tau_{[y]}^{a_{\tau_{[y]}}+\frac{n}{2}-1}\exp\left(\ln\frac{1}{\prod_{i=1}^{n}y_i}+\left(\tau_{[y]}\left(-\frac{1}{2}\sum_{i=1}^{n}\ (\ln(y_i)-\theta)^2-b_{\tau_{[y]}}\right)\right)\right)$.

(c) Distribusi *full conditional posterior* untuk $\tau_{[v]}$ $f(\tau_{[v]}|y,\beta,\tau_{[y]}) \sim gamma\left(\frac{n}{2}+1,\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\theta_i-\sum_{j=1}^{k}x_{ij}\beta_j\right)^2\right)\right)$.

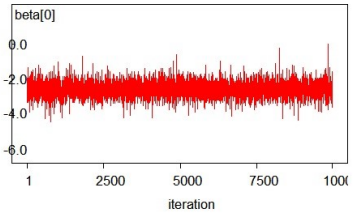
Proses estimasi model *Small Area Estimation* dengan pendekatan *Hierarchical Bayesian* (HB) Lognormal dilakukan menggunakan MCMC dan algoritma *Gibbs Sampling* dengan iterasi sebanyak 10.000 kali dan thin 3.000. Berdasarkan sampel sebanyak 10.000 tersebut, hasil estimasi yang diperoleh telah memenuhi sifat MCMC yang dapat dilihat dari MCMC *diagnostic* plot pada Gambar 1, 2, 3 dan 4.

(1) *Trace plot* untuk estimasi β_0 pada gambar 1 menunjukkan nilai *random* dengan pola yang stasioner. Selain itu, serial nilai sampel yang dihasilkan juga menunjukkan pola random yang stabil dalam suatu domain tetap. *Serial* plot untuk estimasi β_0 ditunjukkan pada Gambar 2.

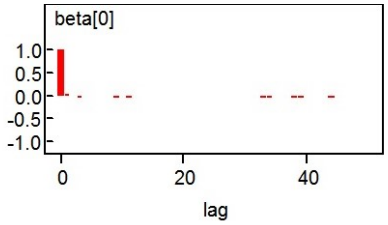


Gambar 1. *plot trace* untuk 10.000 iterasi pada parameter β_0

(2) *Autocorrelation plot* untuk estimasi β_0 dan β_1 pada Gambar 3 memperlihatkan hanya nilai autokorelasi pertama sampel yang tinggi, ini menunjukkan bahwa sampel baik dilakukan dan nilai autokorelasi selanjutnya mendekati nol, ini menunjukkan bahwa iterasi satu dengan yang lainnya saling bebas. Ini berarti bahwa sampel yang dibangkitkan dalam proses MCMC memiliki sifat yang *random*.

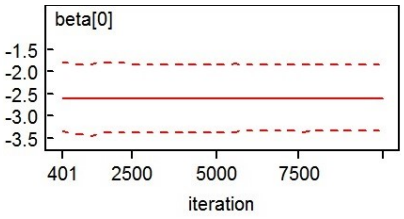


Gambar 2. *Serial plot* untuk 10.000 iterasi pada parameter β_0



Gambar 3. *Autocorrelation plot* untuk 10.000 iterasi pada parameter β_0

(3) *Quantiles plot* pada Gambar 4 menujukkan *ergodic mean*, hasil estimasi parameter yang dihasilkan telah mencapai nilai yang stabil dan berada dalam *credible interval*



Gambar 4. *Quantiles plot* untuk 10.000 iterasi pada parameter β_0

Dari distribusi posterior diatas, dapat dicari distribusi posterior marginal dari masing-masing parameter dengan melakukan perhitungan menggunakan metode MCMC dan *Gibbs Sampling* melalui bentuk *full conditional posterior*. Banyaknya Update yang dijalankan adalah 10.000 dengan thin 3000.

Pendugaan dilakukan dengan bantuan Winbugs. Parameter hasil pendugaan disajikan pada Tabel 1.

Hasil estimasi parameter β model *Small Area Estimation* dengan pendekatan *Hierarchical Bayesian* (HB) Lognormal untuk data Tingkat Pengangguran Terbuka (TPT) dapat dilihat pada Tabel 1. Tabel tersebut memperlihatkan bahwa dengan nilai$\beta_0 = -2,600000$ dan $\beta_1 = 0,007224$ secara statistik mempunyai pengaruh yang

Tabel 1. Nilai Estimasi Parameter Model *Small Area Estimation* dan *Credible Intervalnya* untuk Data TPT Kabupaten Padang Pariaman 2015 menggunakan Pendekatan *Hierarchical Bayesian* (HB) Lognormal

Parameter	Nilai Estimasi	Standar Deviasi	MC error	*Credible Interval*
β_0	-2,600000	0,376700	0,004230	$-3{,}349000 \leq \beta_0 \leq -1{,}848000$
β_1	0,007224	0,003279	0,000031	$0{,}000693 \leq \beta_1 \leq 0{,}013750$
β_2	-0,001022	0,004917	0,000054	$-0{,}010930 \leq \beta_2 \leq 0{,}008735$
β_3	-0,000025	0,001272	0,000015	$-0{,}002573 \leq \beta_3 \leq 0{,}002516$
β_4	0,001599	0,006006	0,000069	$-0{,}010530 \leq \beta_4 \leq 0{,}013710$
β_5	-0,000537	0,000410	0,000004	$-0{,}001352 \leq \beta_5 \leq 0{,}000302$
$\tau_{[v]}$	0,0588	0,0404	0,000453	$0{,}0110500 \leq \tau_{(v)} \leq 0{,}164700$
$\tau_{[y]}$	0,005894	0,015140	0,000130	$0{,}000000 \leq \tau_{(y)} \leq 0{,}050570$

signifikan terhadap Tingkat Pengangguran Terbuka (TPT) di Kabupaten Padang Pariaman. Hal ini ditunjukkan dari nilai β_0 dan β_1 berada dalam *credible interval* yang tidak melewati titik nol.

Selanjutnya akan dilihat perbandingan hasil estimasi langsung dan hasil estimasi metode *Hierarchical Bayesian* (HB) Lognormal dengan melihat nilai standar *error*.

Tabel 2. Perbandingan antara Penduga Langsung dengan Penduga HierarchicalBayesian (HB) Lognormal

Kecamatan	Penduga Langsung		Penduga HB	
	$\hat{y}$	Standar Error	$\hat{y}$	Standar Error
VII Koto Sungai Sariak	0,0562	0,0027	0,0572	0,0007
Batang Anai	0,0442	0,0022	0,0460	0,0007
Lubuk Alung	0,0659	0,0008	0,0649	0,0004
Sintuk Toboh Gadang	0,0753	0,0056	0,0746	0,0012
Ulakan Tapakis	0,0403	0,0043	0,0408	0,0007
Nan Sabaris	0,0566	0,0027	0,0576	0,0007
2 X 11 Enam Lingkung	0,0732	0,0071	0,0724	0,0012
Enam Lingkung	0,0686	0,0016	0,0697	0,0006
2 X 11 Kayu Tanam	0,0567	0,0057	0,0580	0,0010
Patamuan	0,1009	0,0047	0,0999	0,0012
Padang Sago	0,1206	0,2302	0,1226	0,0100
V Koto Kp Dalam	0,0684	0,0032	0,0677	0,0007
V Koto Timur	0,1122	0,0032	0,1107	0,0011
Sungai Limau	0,0632	0,0034	0,0633	0,0008
Batang Gasan	0,091	0,0085	0,0927	0,0019
Sungai Geringging	0,0705	0,0021	0,0713	0,0007
IV Koto Aur Malintang	0,09	0,0024	0,0891	0,0008

Berdasarkan Tabel 2, hasil estimasi Tingkat Pengangguran Terbuka (TPT) menggunakan model *Small Area Estimation* Pendekatan *Hierarchical Bayesian* (HB) Lognormal mempunyai nilai standar *error* yang lebih kecil dibandingkan dengan estimasi TPT dengan pendugaan langsung. Sehingga dapat dikatakan

bahwa metode Small Area Estimation Pendekatan *Hierarchical Bayesian* (HB) Lognormal lebih baik daripada dengan pendugaan langsung.

4. Kesimpulan

Berdasarkan hasil penelitian diperoleh Hasil estimasi Tingkat Pengangguran Terbuka (TPT) menggunakan model *Small Area Estimation* Pendekatan *Hierarchical Bayesian* (HB) Lognormal mempunyai nilai standar *error* yang lebih kecil dibandingkan dengan estimasi TPT dengan pendugaan langsung. Sehingga dapat dikatakan bahwa model *Small Area Estimation* Pendekatan *Hierarchical Bayesian* (HB) Lognormal lebih baik daripada dengan pendugaan langsung.

Daftar Pustaka

[1] Johnson, N.L., Kotz, S., dan Balakrishnan, N. 1995. *Continues Univariate Distributions*. Volume 1 2^{nd} edition. John Wiley and Sons, New York.

[2] Raudenbush, S.W. dan A.S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2^{nd} edition, Sage Publications, Thousand Oaks.

[3] Rao, J.N.K. 2003. *Small Area Estimation*. John Wiley and Sons, New York.