ISBN: 978-602-14387-0-1

KURVA RASIO VARIANSI SPATIAL HIERARCHICAL BAYES SMALL AREA ESTIMATION UNTUK BERBAGAI UKURAN SAMPEL

Siswantining, Titin[1]; Yekti, W[2], Saskya, M[3]
[1,2,3]Universitas Indonesia
titin@sci.ui.ac.id; yekti@ui.ac.id; Saskya-m1@ui.ac.id

Abstract

Area level models such as Fay-Herriot model have been widely used to obtain reliable model-based estimators in small area estimation. However, in the model, two strong assumptions are made. One is that the sampling error variances are customarily assumed to be known, and the other is that the area-specific random effects are assumed to be independent and identically distributed. In this paper, we propose full hierarchical Bayes (HB) models which relax these two strong assumptions by constructing Gaussian conditional autoregressive (CAR) models on the area-specific effects to induce spatial correlation, and assuming the sampling variances unknown. Curve relationship between success proportion to the variance ratio showed a quadratic relationship.

Keywords: Fay-Herriot model, Hierarchical Bayes, CAR, variance ratio.

1. Pendahuluan

Suatu proses yang menggunakan model statistika dalam menghubungkan hasil survei dengan sekelompok variabel prediktor yang diketahui untuk domain (area) kecil, dalam rangka pendugaan tersebut disebut sebagai *small area estimation* (SAE) (Setiawan & Tarumi, 2003).

Survei biasanya digunakan untuk menduga parameter total, mean atau proporsi tetapi juga digunakan untuk menduga parameter dalam subpopulasi (domain) seperti area geografi. Pada prakteknya, SAE juga digunakan untuk menduga parameter pada area geografi. Dalam beberapa aplikasi, domain yang akan diteliti kemungkinan tidak terambil sebagai sampel (ukuran sampelnya nol). (Rao 2003a & 2003b).

Masalah pokok pada SAE adalah menemukan penduga yang handal dari contoh yang berukuran kecil. SAE merupakan teknik statistika pendugaan parameter untuk subpopulasi dalam suatu survei yang besar. Istilah "area kecil" dalam hal ini adalah area geografi kecil seperti desa atau kota, dan area kecil berarti juga domain kecil, area demografi khusus, misalnya desa atau kecamatan, dimana ukuran contoh dalam area ini sangat kecil. Metode pendugaan parameter yang dapat mengatasi hal tersebut disebut

ISBN: 978-602-14387-0-1

pendekatan area kecil (Rao 2003, Sadik 2009). Metode ini dapat menggunakan informasi tambahan yang memungkinkan untuk menghubungkan data survei dengan data administratif yang berperan sebagai data tambahan dengan cara *borrowing strength* (meminjam kekuatan) agar dapat diperoleh penduga SAE tersebut sehingga diperoleh pendugaan yang efisien (Gosh & Rao 1994; Tzavidis 2010).

Berbagai penelitian tentang SAE telah dilakukan yaitu oleh Parsons dan Schenker (2008) yang melakukan dua pendekatan pemodelan pada masalah SAE yaitu prosedur Bayes Berhirarki (*Hierarchical Bayes*) dan model *Generalized Linear Mixed Model* (GLMM). Selain itu, Roy (2007) telah meneliti tentang *Empirical* dan *Hierarchical Bayes* pada SAE (HB SAE). Sunandi (2011) dan Matualage (2012) telah menggunakan SAE spasial. Rumiati (2012) meneliti tentang model Bayes untuk penarikan contoh tidak sama. Siswantining (2013) telah menggunakan HB dan *Empirical Best Linear Unbiased Prediction* (EBLUP) serta *Spatial* EBLUP (SEBLUP) SAE dalam melakukan adaptasi pada statistik pindaian.

Paper ini bertujuan untuk mencari hubungan antara rasio variansi dengan proporsi sukses untuk beberapa ukuran sampel pada setiap area yaitu n = 5, 10, 16, 20; banyaknya area adalah 35, dan pembobotan spasial yang digunakan adalah pembobotan spasial korelasi, variabel tambahan sebanyak 8.

2. Model level Area pada HB SAE

Pendekatan HB SAE merupakan salah satu pendekatan dalam SAE yang memerlukan *prior distribution* (sebaran prior). Model area kecil merupakan model dasar dalam penduga pada area kecil. Pada paper ini hanya membahas model level area dan hanya membahas pembobotan korelasi spasial.

Model Level Area

Model area kecil dengan tipe model level area, peubah penyerta $x_i = (x_{i1}, \dots, x_{ip})'$ hanya tersedia sampai tingkat area dan parameter yang diperhatikan atau parameter yang akan diduga θ_i diasumsikan berhubungan dengan $\mathbf{x}_i$ dengan mengikuti model seperti berikut:

$$\theta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i b_i, i = 1,2, \dots, m \quad (1)$$

ISBN: 978-602-14387-0-1

dimana z_i adalah konstanta positif, $\boldsymbol{\beta}$ adalah vektor dari parameter regresi, dan v_i adalah pengaruh acak (*random effects*) yang diasumsikan mempunyai sebaran normal identik dan saling bebas dengan

$$E(v_i) = 0 \,;\, V(v_i) = \sigma_v^2 \quad (2)$$

Dalam membuat kesimpulan untuk θ_i, diasumsikan bahwa penduga variansi sampling yang diasumsikan mempunyai distribusi normal dengan mean

$E(e_i|\theta_i) = 0$ dan ragam $V(e_i|\theta_i) = \psi_i$.

Jika model pengaruh tetap dan model pengaruh acak digabungkan, maka didapatkan model campuran sebagai berikut :

$$\widehat{\theta}_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i b_i + e_i, i = 1,2, \dots m \quad (3)$$

yang selanjutnya bentuk model tersebut di atas adalah kasus khusus dari *General Linier Mixed Model* (GLMM).

Model HB tanpa pengaruh spasial:

(i) $\hat{\theta}_i | \theta_i, \beta, \sigma_v^2 \ \sim ind\ N(\theta_i, \psi_i)$, $i = 1,\dots, m$

(ii) $\theta_i | \beta, \sigma_v^2 \ \sim ind\ N(x_i^T\beta, b_i^2\sigma_v^2)$, $i = 1,\dots, m$

(iii) $f(\beta) \propto 1$. Jika σ_v^2 tidak diketahui maka (iii) diganti dengan

(iv) $f(\beta, \sigma_v^2) = f(\beta)f(\sigma_v^2) \propto f(\sigma_v^2)$ dengan $f(\sigma_v^2)$ merupakan prior pada σ_v^2.

Pada paper ini menggunakan asumsi bahwa prior pada (iv) adalah β dan σ_v^2 dengan $\sigma_v^{-2} \sim G(a,b)$, $a > 0$, $b > 0$. σ_v^2 mempunyai sebaran invers gamma IG(a,b) dengan $f(\sigma_v^2) \propto \exp\left(-\frac{b}{\sigma_v^2}\right)\left(\frac{1}{\sigma_v^2}\right)^{a+1}$ serta model Fay-Herriot dimana $b_i = 1$.

Peubah respon dalam paper ini adalah peubah kategorik yang bersifat biner yaitu hanya memiliki dua nilai yaitu 1 untuk kejadian sukses dan 0 untuk kejadian gagal dimana kejadian sukses pada penelitian ini adalah miskin dan sebaliknya kejadian gagal adalah tidak miskin.

Misalkan p_i merupakan proporsi area lokal ke-i yang memiliki karakteristik tertentu yang diperhatikan. Dalam kerangka sampel yang diambil dalam setiap area lokal ke-i, p_i dapat ditulis sebagai

$$p_i = \frac{\sum_j y_{ij}}{n_i} = \frac{y_i}{n_i} \quad (4)$$

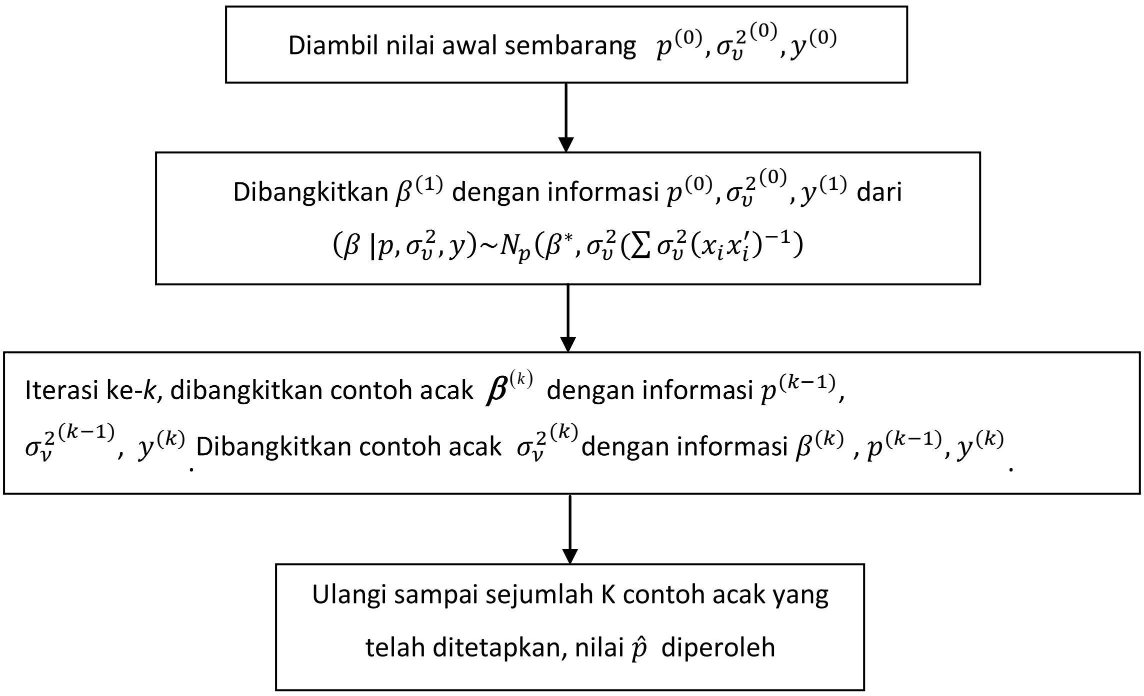
dengan n_i merupakan ukuran contoh dari area lokal ke-i, y_{ij} merupakan nilai

ISBN: 978-602-14387-0-1

nol atau satu, tergantung dari apakah individu ke-j pada area lokal ke-i memenuhi dari karakteristik tertentu yang ingin diperhatikan, dan y_i merupakan jumlah sukses yang terjadi pada suatu area lokal ke-i.

3. Pendugaan parameter proporsi pada HB SAE

Untuk melakukan pendugaan parameter $(p_i, \boldsymbol{\beta}, \sigma_v^2)$ berdasarkan model HB tersebut, digunakan metode *Markov Chain Monte Carlo* (MCMC) dengan menggunakan algoritma Gibbs Sampling dan algoritma Metropolis Hasting. Metode MCMC merupakan metode yang membangkitkan contoh acak dari parameter yang diperhatikan dari suatu sebaran posterior bersyarat. Tahapan pendugaan parameter proporsi menggunakan HB SAE dapat digambarkan seperti pada Gambar 1.



Gambar 1. Bagan algoritma Gibbs untuk menduga p, β dan σ_v^2 pada HB SAE

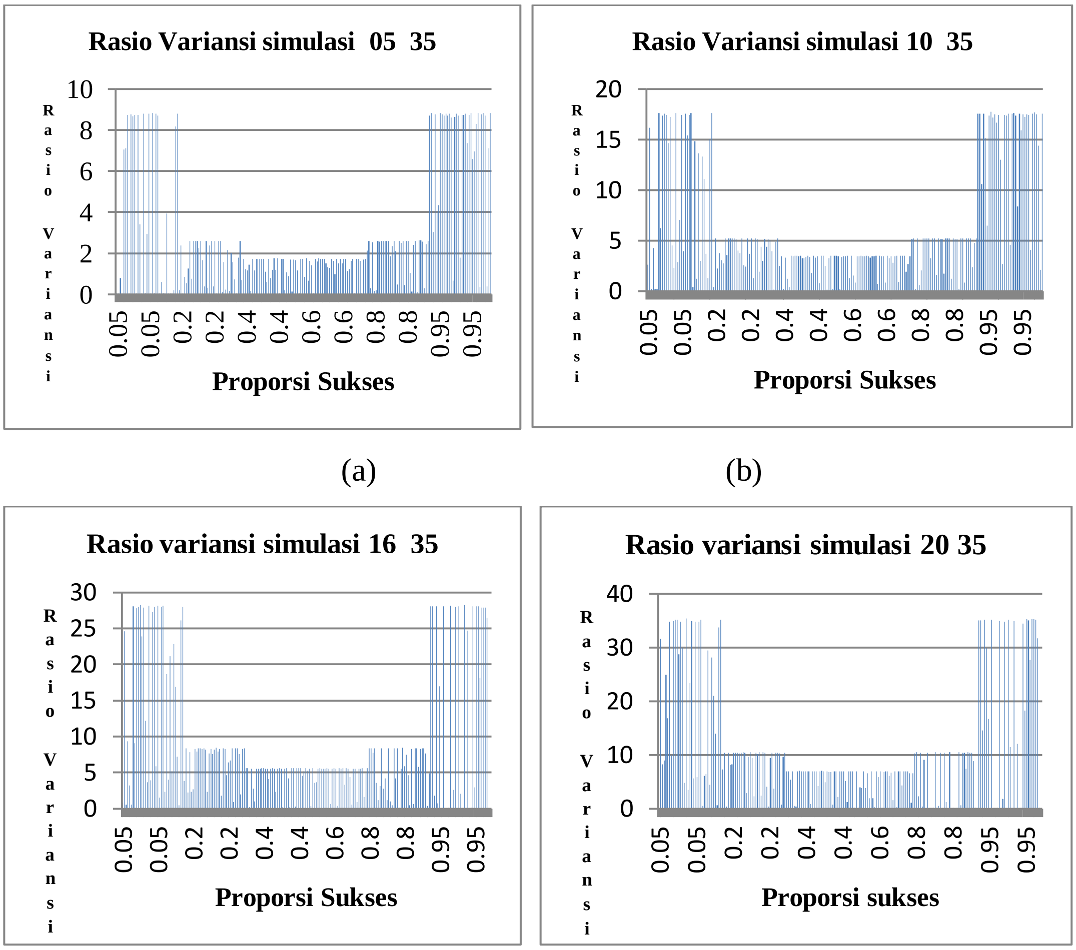
4. Evaluasi hasil pendugaan menggunakan HB SAE

Hasil pendugaan parameter proporsi pada paper ini dievaluasi melalui bias absolut dan rasio variansi. Bias merupakan nilai harapan atau rata-rata dari selisih antara parameter dengan penduganya. Bias absolut merupakan harga mutlak dari nilai bias. Apabila nilai bias absolut kecil berarti penduga tersebut baik. Untuk mengevaluasi variansi penduga SAE digunakan rasio variansi antara variansi penduga langsung DE

ISBN: 978-602-14387-0-1

dengan dengan variansi penduga melalui HB SAE. Jika nilai rasio lebih besar dari 1 berarti penduga tak langsung SAE mempunyai ragam lebih kecil dibandingkan dengan ragam DE.

5. Analisis Data dan Pembahasaan



(a) (b)

(c) (d)

Gambar 2 (a) – (d). Rasio variansi berbagai ukuran sampel n= 5, 10, 16, 20.

Bentuk hubungan antara proporsi sukses dengan rasio variansi dapat dinyatakan dalam persamaan berikut:

Untuk n = 5 maka persamaannya adalah :

Rasio variansi sim 05_35 = 0.6188551 + 8.2565136 * (proporsi sukses) - 5.4280738 * (proporsi sukses kuadrat).

ISBN: 978-602-14387-0-1

Untuk n = 10 maka persamaannya seperti:

Rasio variansi sim 10_35 = 2.748117 + 16.089779 * (proporsi sukses) - 12.468136*(proporsi sukses kuadrat).

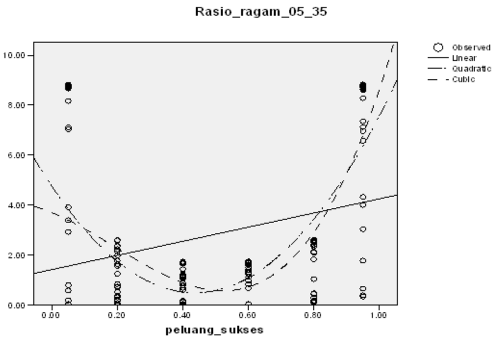
Pada ukuran sampel n = 16 pada setiap area diperoleh persamaan:

Rasio variansi sim 16_ 35 =7.508676+ 11.534759* (proporsi) -13.633853 * (proporsi kuadrat).

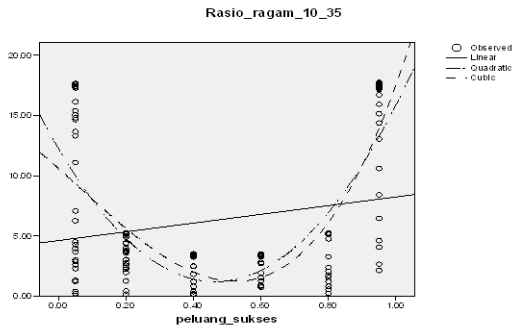
Jika ukuran sampel pada setiap area sebanyak 20 maka persamaannya:

Rasio variansi sim 20_35 = 9.441322 + 15.69451 * (proporsi sukses) – (19.087543)* (proporsi sukses kuadrat).

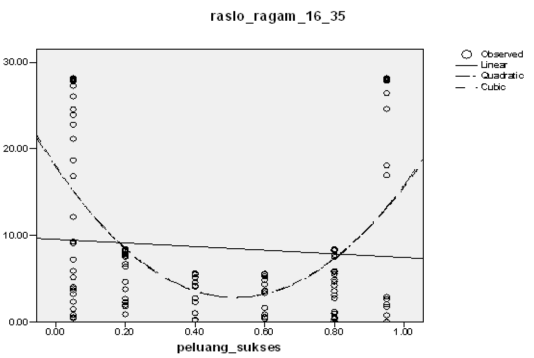
Plot hasil pendugaan rasio ragam (variansi) dengan proporsi sukses untuk masing-masing sampel menggunakan Maple dapat digambarkan seperti Gambar 3 berikut.



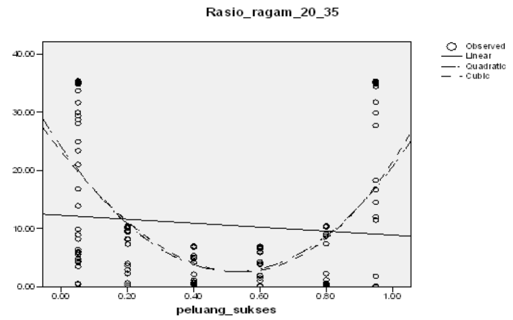
(a)



(b)



(c)

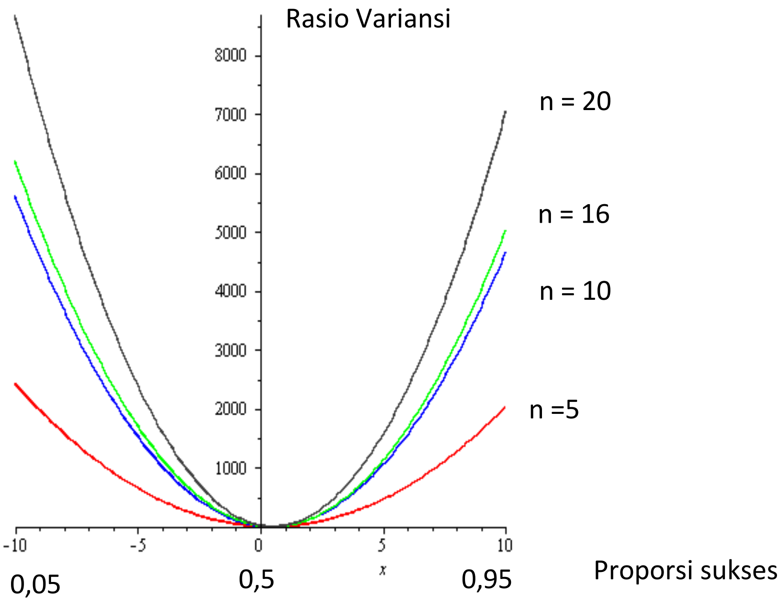


(d)

Gambar 3. (a) – (d). Kurva proporsi sukses untuk n= 5, 10, 16, 20.

ISBN: 978-602-14387-0-1

Gabungan kurva rasio variansi dari empat ukuran sampel, yaitu kurva untuk n = 5, 10, 16, dan 20 seperti pada Gambar 4 berikut.



Gambar 4. Kurva gabungan untuk n = 5, 10, 16, 20

Rasio variansi akan mempunyai nilai minimum pada kisaran proporsi 0.4 – 0.6 apabila dilihat dari Gambar 3 dan Gambar 4. Secara teori dapat dibuktikan bahwa variansi proporsi akan mempunyai nlai minimum pada proporsi sukses 0,5.

Bukti:

$$Var\ p = Var(\frac{y}{n})$$

$$Var\ y = n\ p\ (1-p)$$

$$Var\ p = \frac{1}{n^2} n\ p(1-p) = \frac{1}{n}(p-p^2)$$

Suatu fungsi akan optimum jika diferensialnya sama dengan nol.

$1-2p=0$ berarti $p=\frac{1}{2}$. (terbukti)

Apabila dilihat dari kurva dan dari pembuktian secara teori maka rasio variansi mempunyai nilai minimum pada saat proporsi sukses sama dengan poporsi gagal.

6. Diskusi

Dari hasil analisis data menunjukkan bahwa hubungan antara proporsi sukses dengan rasio variansi adalah bentuk kuadratik. Hal ini dapat dikatakan bahwa rasio variansi akan bernilai tinggi apabila proporsi suksesnya kecil sekali (pada paper ini adalah 0,05) atau besar sekali (pada paper ini adalah 0,95). Hasil pendugaan menggunakan HB SAE akan menghasilkan rasio variansi besar (atau berarti bahwa variansi HB SAE kecil) apabila jumlah sampelnya membesar (pada paper ini adalah n = 20). Rasio variansi akan menghasilkan nilai terendah apabila proporsi sukses bernilai sama dengan proporsi gagal.

Ucapan terimakasih

Ucapan terimakasih disampaikan kepada DRPM UI yang telah mendanai penelitian ini (No kontrak: 1377/H2.PPKP/HKP.05.00/2013).

Daftar Pustaka

Ghosh, M. & J.N.K. Rao. (1994). Small Area Estimation: An Appraisal. *Statistical Science*. Vol. 9, N0. 1, 255-93.

Matualage, D. (2012). Metode Prediksi Tak Bias Linier Terbaik Empiris Spasial Pada Area Kecil Untuk Pendugaan Pengeluaran Per Kapita (Studi Kasus : Kabupaten Jember Provinsi Jawa Timur). [tesis]. Bogor: Institut Pertanian Bogor.

Parsons VL, Schenker N. (2008). County-Level Small Area Estimation Using the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS). Section Survey Research Methods – JSM 2008.

Rao JNK. (2003a). *Small Area Estimation*. New Jersey: John Wiley & Sons.

Rao, J.N.K. (2003b). Some New Developments in Small Area Estimation. *JIRSS* (2003) Vol. 2, No. 2, pp 145-169.

Rao, J.N.K. (2012). Small Area Estimation: Methods and Applications. Paper presented at the Seminar "*Applications of Small Area Estimation Techniques in the Social Sciences*", October 3 – 5, 2012, Iberoamerican University, Mexico City.

Rumiati AT. (2012). Model Bayes Untuk Pendugaan Area Kecil Dengan Penarikan Contoh Berpeluang Tidak Sama Pada Kasus Respon Binomial dan Multinomial. Aplikasi: Pendugaan Indeks Pendidikan Level Kecamatan Di Jawa Timur. [disertasi]. Bogor: Sekolah Pascasarjana, Institut Pertanian Bogor.

Sadik, K. (2009). Metode Prediksi Tak-Bias Linear Terbaik Dan Bayes Berhirarki Untuk Pendugaan Area Kecil Berdasarkan Model State Space. [disertasi]. Bogor: Sekolah Pascasarjana Institut Pertanian Bogor.

Setiawan, A. and T. Tarumi. (2004). Small Geographic Area Estimation in WinBUGS

ISBN: 978-602-14387-0-1

with Two Approaches Prediction. *Journal of the Faculty of Environmental Science and Technology*. Okayama University. Vo1.9. No.1, pp.9-17.

Siswantining, T. A. Saefuddin, K.A. Notodiputro, N. Nuryartono, W. Mangku. (2012). Adaptation of Hierarchical Bayes SAE to Spatial Satscan. IOSR Journal of Mathematics. Vol. 2, Issue 4 (Sept. – Oct. 2012), PP 01 – 08.

Siswantining, T. (2013). Geoinformatika pada kasus area kecil dan penerapannya pada pendeteksian kantong-kantong kemiskinan di Jember. [disertasi]. Bogor: Sekolah Pascasarjana, Institut Pertanian Bogor.

Sunandi E. (2011). Model Spasial Bayes Dalam Pendugaan Area Kecil Dengan Peubah Respon Biner (kasus : Pendugaan Proporsi Keluarga Miskin di Kabupaten Jember Jawa Timur). [tesis]. Bogor: Sekolah Pascasarjana, Institut Pertanian Bogor.

Zhou, Q.M and Yong You. (2008). Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey. SSC Annual Meeting, May 2008. Proceedings of the Survey Methods Section

Tzavidis N. (2010). What is poverty mapping. UK: Social Statistics & Centre for Cencus and Survey Research, University of Manchester.