Penduga Maksimum Likelihood untuk Parameter Dispersi Model Poisson-Gamma dalam Konteks Pendugaan Area Kecil

The Maximum Likelihood of Estimating Dispersion Parameter for Poisson-Gamma Model in Small Area Estimation Context

Alfian F. Hadi[1)]
Nusyirwan[1)]
Khairil Anwar Notodiputro[2)]

[1)]Mahasiswa Program Doktor Statistika
Sekolah Pascasarjana Institut Pertanian Bogor
afhadi@unej.ac.id

[2)]Guru Besar Statistika, Departemen Statistika, Institut Pertanian Bogor

Abstract

The Poisson-Gamma (Negative Binomial) distribution is considered to be able to handle overdispersion better than other distributions. Estimation of the dispersion parameter, ϕ, is thus important in refining the predicted mean when the Empirical Bayes (EB) is used. In GLM's sense dispersion parameter (ϕ) have effects at least in two ways, (i) for Exponential Dispersion Family, a good estimator of ϕ gives a good reflection of the variance of Y, (ii) although, the estimated β doesnt depend on ϕ, estimating β by maximizing log-likelihood bring us to Fisher's information matrix that depends on its value. Thus, ϕ does affect the precision of β, (iii) a precise estimate of ϕ is important to get a good confidence interval for β. Several estimators have been proposed to estimate the dispersion parameter (or its inverse). The simplest method to estimate ϕ is the Method of Moments Estimate (MME). The Maximum Likelihood Estimate (MLE) method, first proposed by Fisher and later developed by Lawless with the introduction of gradient elements, is also commonly used. This paper will discuss the use of those above methods estimating ϕ in Empirical Bayes and GLM's of Poisson-Gamma model that is applied on Small Area Estimation.

Keywords: Small Area Estimation, Empirical Bayes, Poisson-Gamma, Negative Binomial, dispersion parameter, MLE, MME.

PENDAHULUAN

Latar Belakang

Dalam pendugaan area kecil (*small area estimation*), berbagai metode telah dikembangkan khususnya menyangkut metode yang berbasis model (*model-based area estimation*). Metode tersebut adalah penduga prediksi tak bias linier terbaik empirik atau *empirical best linear unbiased prediction* selanjutnya disebut *EBLUP*, Bayes empirik atau *empirical Bayes* disingkat *EB*, dan Bayes hierarkhi atau *Hierarchical Bayes* yang disingkat *HB*. Metode *EBLUP* merupakan metode untuk data kontinu sedangkan *EB* dan *HB* adalah metode untuk data biner atau cacahan.

S ebaran poisson mempunyai peran yang penting dalam metode *empirical bayes*. Hal ini disebabkan antara lain oleh dapat ditemukannya rataan poisson tanpa pendugaan sebaran prior secara explisit. Namun model poisson mempunyai keterbatasan yakni pada kesamaan nilai tengah dan ragamnya, sehingga umumnya dijumpai fenomena overdispersi. Penanganan overdispersi, seringkali dilakukan melalui pendekatan model campuran Model Poisson-Gamma (*Negative Binomial).* Model ini telah dikenal luas untuk menangani pengaruh acak dengan overdispersi secara lebih baik dari pendekatan/distribusi yang lain.

Model binomial negatif (atau secara umum pada keluarga sebaran eksponential berdispersi) memuat suatu parameter ϕ yang disebut parameter dispersi. Dalam pemodelan GLM, ϕ berperan dalam sedikitnya dua, (i) pada keluarga sebaran eksponensial berdispersi, nilai ragam Y proportional terhadap nilai parameter ϕ, artinya penduga ϕ merefleksikan ragam Y (ii) meskipun, nilai dugaan parameter regresi, β tidak bergantung pada ϕ, tetapi pendugaan β dengan MLE dilakukan melalui matriks turunan kedua, informasi Fisher, dan tergantung pada ϕ . Sehingga ϕ menentukan presisi penduga β, (iii) penduga ϕ yang baik diperlukan untuk mendapatkan selang kepercayaan yang baik bagi β.

Beberapa metode pendugaan ϕ telah diusulkan. Diantaranya adalah metode momen (MME), metode yang cukup sederana untuk menduga ϕ. Metode Maximum Likelihood (MLE) adalah yang paling umum dipakai, pertama kali diusulkan oleh Fisher dan kemudian dikembangkan oleh Lawles melaui gradient elements. Pendugaan parameter dispersi, ϕ, sangat penting dalam memperbaiki penduga, khususnya bila menggunakan Bayes Empirik. Pendugaan ini berperan mendapatkan ϕ yang akan dgunakan sebagai hiperparameter.

Permasalahan

Meskipun pendugaan ϕ terpisah dari pendugaan β, namun perannya tidak dapat diabaikan. Dalam konteks penduga area kecil, menarik untuk dievaluasi bagaimana pengaruh pendugaan parameter dispersi ini terhadap performa penduga komposit Bayes Empiriknya, bukan pada parameter dispersi itu sendiri. Tulisan ini akan membicarakan penggunaan kedua metode di atas untuk menduga parameter dispersi, ϕ dalam skema Bayes Empirik dan model Poisson-Gamma yang digunakan pada penduga area kecil.

Simulasi dilakukan untuk (i) membandingkan metode momen dan maksimum likelihood untuk pendugaan parameter dispersi model poison-gamma dengan dan tanpa peubah tambahan (*auxiliary variable*) dan (ii) membandingkan performa penduga Bayes Empirik pada konteks penduga area kecil yang dibangun dari pendugaan parameter dispersi distribusi negatif binomial melalui kedua metode.

SMALL AREA ESTIMATION

Area kecil didefinisikan sebagai subpopulasi yang memiliki ukuran contoh yang kecil sehingga pendugaan langsung tidak dapat menghasilkan pendugaan yang teliti (Rao, 2003). Area kecil dapat berupa kota, kabupaten, kecamatan, desa/kelurahan, kelompok suku, kelompok jenis kelamin atau kelompok umur.

Small area estimation (SAE) merupakan suatu teknik statistika yang digunakan untuk menduga parameter-parameter area kecil. Teknik ini digunakan dengan memanfaatkan data dari hasil survei domain besar seperti data sensus atau data Survey Sosial Ekonomi Nasional (SUSENAS) untuk menduga peubah yang menjadi perhatian pada area kecil.

Pendugaan langsung *(direct estimation)* adalah pendugaan dengan berdasarkan penerapan model penarikan sampel. Pendugaan ini tidak mampu memberikan ketelitian yang baik jika ukuran contoh dalam area kecil dan statistik yang diperoleh akan memiliki ragam yang besar bahkan terkadang pendugaan ini tidak mampu dilakukan karena sampel tidak mewakili populasi.

SAE dikembangkan sebagai teknik pendugaan alternatif yang mampu mengatasi semua masalah diatas, yaitu dengan pendugaan tidak langsung *(indirect estimation)*. Pendugaan ini bersifat meminjam kekuatan dari pengamatan contoh area yang berdekatan dengan memanfaatkan informasi tambahan yakni dari data sensus atau survei berskala nasional (Rao, 2003). Proses pendugaan tidak langsung merupakan pendugaan pada suatu domain dengan cara menghubungkan informasi pada area tersebut dengan area lain melalui model yang tepat. Hal ini berarti bahwa dugaan tersebut mencakup data dari domain lain.

Small Area Estimation Model

Salah satu model dasar area kecil (Rao, 2003) yaitu *Basic area level (type A) model*, yaitu model yang didasarkan pada ketersediaan data pendukung yang hanya ada untuk level area tertentu. Misalkan $x_i = (x_{1i},...,x_{p1})^T$ dan parameter yang akan diduga θ_i, diasumsikan mempunyai hubungan dengan x_i. Data pendukung tersebut digunakan untuk membangun model: $\theta_i = x_i^T \beta + b_i v_i$, dengan i=1,…,m dan v_i~N$(0, \sigma_v^2)$ sebagai pengaruh acak yang diasumsikan normal serta $\beta = (\beta_1,...,\beta_p)^T$ adalah vektor koefisien regresi berukuran $p \times 1$. Sedangkan b_i adalah konstanta bernilai positif yang diketahui. Untuk melakukan inferensi mengenai θ_i didapatkan dengan mengasumsikan bahwa model penduga langsung y_i tersedia, yaitu: $y_i = \theta_i + e_i$, dimana i=1,…,m dengan sampling error e_i~N$(0, \sigma^2{}_{ei})$ dan $\sigma^2{}_{ei}$ diketahui. Pada akhirnya, kedua model digabungkan dan menghasilkan model gabungan: $y_i = x_i^T \beta + b_i v_i + e_i$, dimana i=1,…,m. Model tersebut merupakan bentuk khusus dari model linier campuran *(generalized linear mixed model)* yang terdiri dari pengaruh tetap *(fixed effect)*, yaitu β dan pengaruh acak *(random effect)* yaitu v_i (Rao, 2003, Kurnia & Notodiputro 2006).

MODEL POISSON-GAMMA

Model poisson adalah model peluang standar untuk data cacahan. Model ini akan mengalami keterbatasan dalam rataan dan ragam ketika digunakan untuk pendugaan parameter tunggal. Umumnya, data cacahan (seperti data jumlah) mengalami overdispersi. Oleh karena itu, dikembangkan suatu formulasi poisson yang mengakomodasi ragam ekstra dari pengamatan data contoh. Maka, diperkenalkan model dua tahap untuk data cacahan, yang dikenal dengan model campuran poisson-gamma.

Model poisson-gamma dimana y_i berdistribusi poisson dengan parameter θ_i, sedangkan θ_i sendiri berdistribusi gamma dengan parameter-parameter yang bersesuaian dengan nilai tengah dan keragaman total y_i. Untuk menentukan parameter-parameter sebaran gamma bagi θ_i marilah kita perhatikan model poisson-gamma bagi area level yang digunakan yaitu:

$$y_i = X'\beta + 1v_i + e_i = \theta_i + e_i$$

$$\theta_i = X'\beta + v_i$$

Dengan $y_i \sim Poisson(\theta_i)$, $E(y_i) = X'\beta$, $Var(y_i) = E(y_i) = X'\beta$

$E(e_i) = 0; \quad Var(e_i) = \sigma^2{}_{e\,i}; \quad E(v_i) = 0; \quad Var(v_i) = \sigma^2$

$$\theta_i \sim gamma\left(\frac{\mu_i^2}{\sigma^2}, \frac{\sigma^2}{\mu_i}\right); ;\ E(\theta_i) = \mu_i\ Var(\theta_i) = \sigma^2$$

Empirical Bayes

Dasar perkembangan pendekatan statistik Bayes adalah hukum Bayes yang dibuat oleh Thomas Bayes. Hukum ini diperkenalkan oleh Richard Price tahun 1763, dua tahun setelah Thomas Bayes wafat. Tahun 1774 dan 1781, Laplace memberikan analisis lebih rinci dan lebih relevan untuk statistik Bayes sekarang (Gill, 2002). Metode Bayes akan sangat sulit digunakan dan kadang sangat sensitif karena membutuhkan penaksiran peluang tertentu yang sulit untuk ditaksir. Maka diperkenalkan metode *Empirical Bayes (EB)* dengan mengasumsikan bahwa prior tidak diketahui, selanjutnya data digunakan untuk memperoleh dugaan parameter prior. Rao (2003) menyatakan bahwa metode *EB* dan *HB (Hierarchical Bayes)* yang cocok digunakan dalam menangani data biner dan cacahan pada pendugaan area kecil.

Metode *EB* dalam konteks pendugaan area kecil secara ringkas adalah:

1. Mendapatkan fungsi kepekatan peluang akhir (posterior) dari parameter area kecil yang menjadi perhatian.
2. Menduga parameter model dari fungsi kepekatan peluang marginal
3. Menggunakan fungsi kepekatan peluang posterior dugaan untuk membuat inferensi parameter area kecil yang menjadi perhatian.

Pendugaan Parameter Dispersi

Untuk mendapat penduga bayes empirik, parameter-parameter pada sebaran prior (*hyperparameter*) harus diduga. Pendugaan parameter dispersi, ϕ, sangat penting dalam memperbaiki penduga Bayes Empirik. Pendugaan ini berperan mendapatkan ϕ yang akan digunakan sebagai hiperparameter. Lebih dari itu, pada kasus tertentu, nilai dugaan ϕ dapat langsung menjadi salah satu komponen dalam pembobot komposit penduga area kecil.

Bagian ini akan memperkenalkan dua metode penduga parameter dispersi, yaitu MME dan MLE.

Method of Moments Estimate (MME)

Untuk sebuah sebaran binom negatif, ragam σ^2, rataan μ dan parameter dispersi ϕ memiliki hubungan $\sigma^2 = \mu + \frac{\mu^2}{\phi}$. Berdasarkan hubungan ini, MME dikembangkan dan diduga dengan, $\hat{\phi} = \hat{\alpha} = \dfrac{\bar{y}}{s^2 - \bar{y}}$ dimana $\bar{y}$ dan s^2 adalah momen contoh takbias pertama dan kedua. Perhatikan bahwa menduga ϕ hanya mungkin jika $s^2 > \bar{y}$ karena $\phi > 0$. Untuk mendapat penduga ϕ dengan baik melalui MME, sangat penting untuk mengetahui ragam karena perubahan sedikit saja pada nilai ragam mengakibatkan variasi besar nilai ϕ. Masalah ini makin besar jika ukuran contoh makin kecil (Zang, et all, 2004).

Maximum Likelihood Estimator (MLE)

Pada kasus tertentu, solusi bagi penduga maksimum likelihood tidak dapat dijumpai dalam bentuk tertutup (*close form*). Namun pendugaan dapat diperoleh secara numerik. Fungsi log-likelihood akan mencapai maksimum jika vektor gradien sama dengan nol. Atau dengan kata lain kita akan memaksimumkannya melalui deferensial tehadap ϕ_j dan β_j, $\partial l/\partial\phi_j=0$ dan $\partial l/\partial\beta_j=0$, sehingga kita akan memperoleh parameters ϕ_j dan β_j yang memenuhi kondisi ini berlaku untuk data set yang kita perhatikan. McCullagh and Nelder (1989) menjelaskan algoritna Iterative (re)Weighted Linear Regression (metode scoring) yang dapat digunakan untuk memperoleh dugaan parameter dalam Model Linier Terampat.

Untuk menduga nilai parameter ϕ_j (yang tidak diketahui) dalam binomial negatif dapat digunakan metode skoring yang merupakan modifikasi dari algoritma Newton–Raphson untuk mencari akar-persamaan. Dan solusi dari persamaan model linier Binomial Negatif dapat diperoleh dari algoritma Newton–Raphson klasik (Dobson 1990). Dan untuk itu diperlukan nilai awal untuk ϕ_j dan β_j.

Binomial negatif adalah sebuah sebaran dengan sebuah parameter tambahan ϕ pada fungsi ragam. SAS dengan PROC GENMOD menduga ϕ dengan maximum likelihood , menurut McCullagh & Nelder (1989) atau Lawless (1987).

METODE BAYES EMPIRIK TANPA PEUBAH PENJELAS BAGI MODEL POISSON-GAMMA

Model Poisson-Gamma merupakan model yang sering digunakan untuk mengakomodasi permasalahan overdispersi (ragam melebihi rataan) pada model Poisson. Dua tahapan model Poisson-Gamma adalah :

$$y_i \overset{ind}{\sim} Poisson(\theta_i),\ i = 1,\ldots,m$$

$$\theta_i \overset{iid}{\sim} gamma\left(\alpha, \alpha^{-1}\right) \qquad (2)$$

dengan y_i adalah banyaknya pengamatan pada area ke-*i*, θ_i adalah nilai harapan dan ragam y, dan *m* menyatakan jumlah area, sedangkan α merupakan parameter prior yang belum diketahui. Sebagai prior diasumsikan bahwa $\theta_i \overset{iid}{\sim} gamma\left(\alpha, \alpha^{-1}\right)$ dengan $E(\theta_i) = 1$, $Var(\theta_i) = 1/\alpha$ Berdasarkan kedua asumsi tersebut maka didapatkan sebaran posterior untuk θ_i yaitu $\theta_i | y_i, \alpha \overset{ind}{\sim} gamma\left(\alpha + y_i, \frac{\mu_y}{\mu + \alpha}\right)$ serta penduga Bayes bagi θ_i dan ragam posterior bagi θ_i adalah :

$$\hat{\theta}_i^B(\alpha) = E\left(\theta_i | y_i, \alpha\right) = (\alpha + y)\left(\frac{\mu}{\mu + \alpha}\right)$$

dan

$$Var\left(\theta_i | y_i, \alpha\right) = g_{1i}\left(\alpha, y_i\right) = (\alpha + y)\left(\frac{\mu}{\mu + \alpha}\right)^2$$

Penduga Bayes ini mensyaratkan terlebih dulu diketahui nilai parameter sebaran prior. Permasalahannya, seringkali informasi mengenai parameter prior belum diketahui. Pendekatan Bayes empirik atau *empirical Bayes* (*EB*) dapat digunakan untuk mengatasinya. Pada metode ini, informasi parameter prior dapat diperoleh dengan memaksimumkan fungsi sebaran marginal $y_i | \alpha \overset{iid}{\sim}$ binomial negatif, meski bentuk tertutup untuk bagi penduga parameternya tidak ada (Clayton & Kaldor 1987), kita dapat memanfaatkan model null binomial negatif, untuk menduga α.

Zang, et al. (2006) menggunakan penduga momen sederhana untuk memperoleh dugaan parameter dispersi binomial-negatif yaitu:

$$\hat{\phi} = \hat{\alpha} = \frac{\bar{y}}{s^2 - \bar{y}}$$

Dengan mensubstitusi $\hat{\alpha}$ diperoleh penduga *EB* bagi θ_i yaitu

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^B\left(\hat{\alpha}\right) = \hat{\gamma}_i \hat{\theta}_i + \left(1 - \hat{\gamma}_i\right)\hat{\theta}_{synt.}$$

dengan $\hat{\gamma}_i = \bar{y} / \left(\bar{y} + \hat{\alpha}\right)$, $\hat{\theta}_i = y_i$ sebagai penduga langsung dari θ_i, y_i menyatakan banyaknya pengamatan, $\hat{\theta}_{synt.} = \bar{y}$ adalah penduga sintetik (Rao, 2003).

METODE BAYES LINIER EMPIRIK BAGI MODEL POISSON-GAMMA

Metode Bayes linier empirik (*Empirical Linear Bayes*/*ELB*) merupakan suatu metode yang menghindari adanya asumsi sebaran pada metode Bayes empirik (Rao 2003). Metode ini hanya menggunakan momen pertama dan kedua dalam menentukan penduga liniernya. Secara umum, model dua tahap pada pendugaan Bayes linier adalah :

Dengan model linier poisson-gamma, $y_i \sim Poisson(\theta_i)$, dan prior distribution $\theta_i \sim gamma\left(\frac{\mu_i^2}{\sigma^2}, \frac{\sigma^2}{\mu_i}\right)$, maka diperoleh posterior Bayes $gamma\left(\alpha + y_i, \frac{X'\beta}{X'\beta + \alpha}\right)$ Nilai tengah posteriornya adalah $(\alpha + y_i)\frac{\mu_i}{\mu_i + \alpha}$, melalui manipulasi nilai tengah diperoleh penduga bayes:

$$\mu_i + \left(\frac{\mu_i}{\mu_i + \alpha}\right)(y_i - \mu_i) = \mu_i + (1 - \gamma^*{}_i)(y_i - \mu_i) = \gamma_i y_i + (1 - \gamma_i)X'\beta\,,$$

diperoleh pembobot bagi penduga bayes adalah:

$$1 - \gamma^*{}_i = \left(\frac{\mu_i}{\mu_i + \alpha}\right) = \gamma_i = \left(\frac{X'\beta}{X'\beta + \alpha}\right)$$

Penduga Empirical bayes diperoleh dengan menduga α dan β melalui *Generalized Linear Mixed Model*, dengan memanfaatkan sebaran marginalnya yaitu *Negative-Binomial*. Dimana $\hat{\beta}$ adalah penduga parameter regresi binomial-negatif sedangkan $\hat{\alpha}$ adalah penduga bagi dispersion parameter distribusi *negative-binomial* (=$\hat{\phi}$). Bila model yang digunakan adalah linier dan model poisson-gamma (binom-negatif) dalam fungsi link logaritmik maka parameter regresi yang diperoleh perlu dieksponensialkan terlebih dahulu.

Dengan denikian penduga bayes adalah: $\hat{\theta}_i^{EB} = \hat{\gamma}_i \hat{\theta}^L{}_i + (1 - \hat{\gamma}_i)\hat{\theta}^{TL}{}_i$, dengan penduga langsung bagi θ adalah y_i dan penduga taklangsungnya adalah $X'\hat{\beta}$.

PENDEKATAN *JACKKNIFE* UNTUK PENDUGA *MSE*($\hat{\theta}_i^{EB}$)

Pendekatan *jackknife* merupakan metode yang sering digunakan dalam survei karena konsepnya yang sederhana (Jiang, Lahiri dan Wan 2002). Metode ini diperkenalkan Tukey pada tahun 1958 dan berkembang menjadi suatu metode yang dapat mengkoreksi bias suatu penduga, yaitu dengan menghapus observasi ke-i untuk i=1,…,m dan selanjutnya melakukan pendugaan parameter.

Langkah-langkah pendekatan *jackknife* dalam menduga *MSE* dugaan *empirical Bayes* adalah sebagai berikut (Kurnia & Notodiputro. 2006):

1. Anggap bahwa $\hat{\theta}_i^{EB} = k_i(y_i, \underline{\hat{\beta}}, \hat{\alpha})$, $\hat{\theta}_{i-1}^{EB} = k_i(y_i, \underline{\hat{\beta}_{-1}}, \hat{\alpha}_{-1})$, lalu $\hat{M}_{2i} = \frac{m-1}{m}\sum_{l}^{m}(\hat{\theta}_i^{EB} - \hat{\theta}_{i-1}^{EB})^2$
2. Dengan mencari $\underline{\hat{\beta}}_{-1}$ dan α_{-1} yang merupakan penduga kemungkinan maksimum yang diperoleh dari data ke-1 yang dihapus, maka dihitung $\hat{M}_{1i} = g_{1i}(\underline{\hat{\beta}}, \hat{\alpha}, y_i) - \frac{m-1}{m}\sum_{i=m}^{m}[g_{1i}(\underline{\hat{\beta}_{-1}}, \hat{\alpha}_{-1}, y_i) - g_{1i}(\underline{\hat{\beta}}, \hat{\alpha}, y_i)]$
3. Penduga Jackknife bagi kuadrat tengah galat penduga Bayes empirik diberikan oleh $ktg_J(\hat{\theta}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i}$

METODOLOGI

Skenario Simulasi

Simulasi dilakukan dengan dua skenario, (i) skenario tanpa peubah penjelas, dan (ii) dengan peubah penjelas. Dengan parameter sebagaimana tabel 1. Dengan penetapan ini, perlu digarisbawahi bahwa nilai Xβ semakin besar dari area 1 ke area 20. Dua hal perlu diperhatikan menyangkut hal ini (i) Xβ adalah nilai harapan Y_i sehingga nilai yi dan Xβ-dugaan akan membesar pula, (ii) Xβ adalah keragaman total yi, artinya keragaman yi menaik dari area 1 ke area 20.

Tabel 1. Parameter-parameter dalam simulasi

Area	X0	X1	mu_i	Area	X0	X1	mu_i	β-0	β-1
1	1	0	1.7214	11	1	3	2.7702	1.7214	0.3496
2	1	0	1.7214	12	1	4	3.1198		
3	1	0	1.7214	13	1	4	3.1198		
4	1	1	2.071	14	1	5	3.4694		
5	1	1	2.071	15	1	5	3.4694		
6	1	1	2.071	16	1	5	3.4694		
7	1	2	2.4206	17	1	6	3.819		
8	1	2	2.4206	18	1	6	3.819		
9	1	3	2.7702	19	1	7	4.1686		
10	1	3	2.7702	20	1	7	4.1686		

Tahapan Simulasi

Pembangkitan data untuk kedua skenario simulasi sama, berbeda pada saat *fitting* model. Model tanpa X *difit* dengan model null, model dengan X dengan regresi.

1. Tetapkan X, β, dan σ^2 ; dengan $\mu_i = X'\beta$

2. Bangkitkan $\theta_i \sim gamma\left(\frac{\mu_i^2}{\sigma^2}, \frac{\sigma^2}{\mu_i}\right); \alpha = \frac{\mu_i^2}{\sigma^2}$; kemudian $y_i \sim Poisson(\theta_i)$

3. Menduga Parameter Dispersi untuk model tanpa X dengan metode momen (MME) dan maksimum likelihood (MLE). Dengan Proc Genmod model null.

4. Menduga Parameter Dispersi untuk model dengan X dengan MME dan MLE. Parameter regresi diduga dengan maksimum likelihood, menggunakan Proc Genmod.

5. Menentukan penduga *Empirical Bayes* bagi langkah 3 dan 4

6. Menghitung KTG/ MSE jackknife bagi penduga bayes.

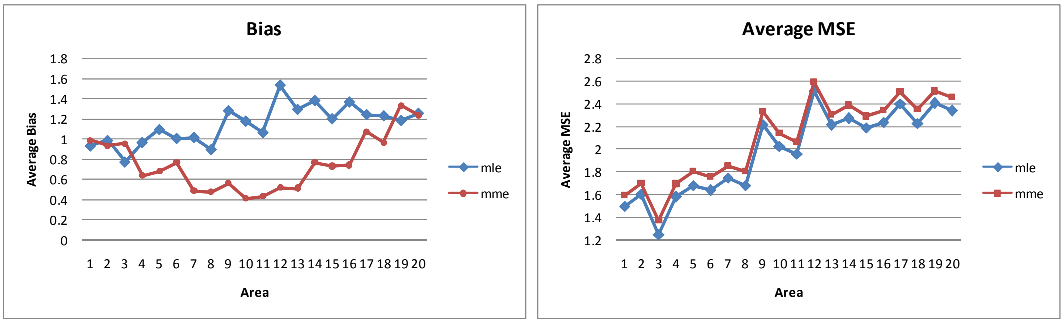
7. Evaluasi simulasi dilakukan dengan memeriksa MSE dan Bias, serta dengan membandingkan dua statistika yaitu Mean Absolute Relative Error (MARE) dan Average Relative Root Mean Square Error (RRMSE). MARE mengukur beda absolute antara parameter penduganya sedangkan RRMSE menghitung keragaman penduga.

MARE = $\frac{|\theta - \hat{\theta}|}{\theta}$, RRMSE = $\frac{\sqrt{MSE}}{\hat{\theta}}$.

HASIL DAN DISKUSI

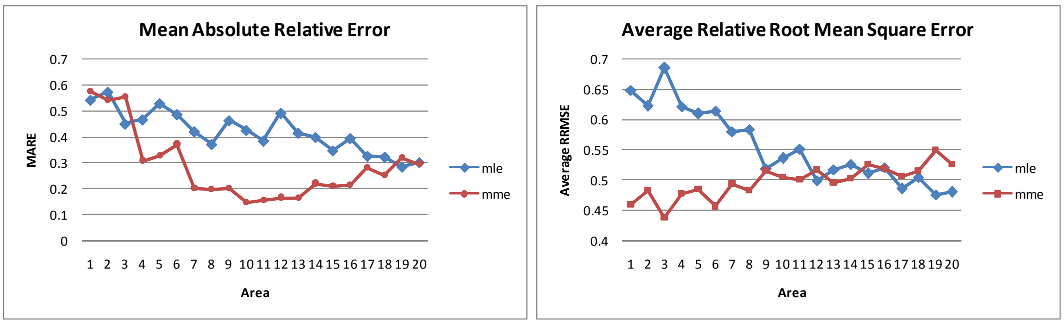
Tanpa Peubah Penjelas

Pada skenario ini peubah X yang merupakan pembentuk nilai tengah dan ragam bagi Y diabaikan. Sehingga pendugaan baik itu penduga langsung maupun sintetiknya diperoleh tanpa informasi tambahan yang seharusnya ada. Meski demikian, perbandingan pada metode pendugaan nilai parameter dispersi sebaran marginal binomial negatif dilakukan pada kondisi yang sama-sama mengabaikan informasi X. Gambar 1 menunjukkan bias MLE tampak naik dengan naiknya nilai harapan dan ragam y. Sedangkan bias dari metode momen relatif rendah pada nilai y yang sedang, pada nilai y yang rendah dan tinggi, bias MME cenderung sama dengan bias MLE.



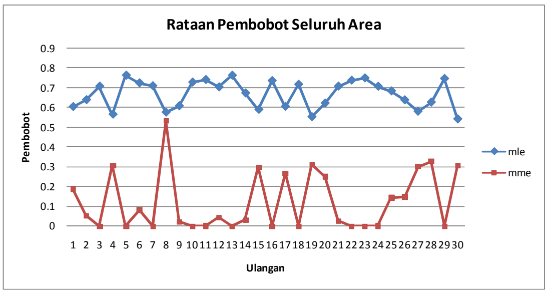
Gambar 1. Rataan Bias dan Rataan MSE bagi model null *EB* Poisson-Gamma

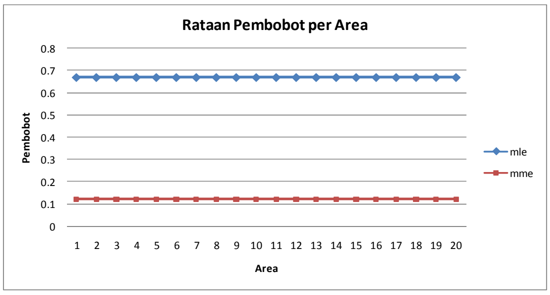
Bila kita perhatikan maka MLE lebih baik karena MSEnya lebih rendah dari MME. Hal ini berarti MLE mempunyai ketelitian yang lebih baik, karena meski biasnya lebih tinggi, variansinya rebih rendah dari MME. Pada RRMSE gambar 2, nilai dugaan MME meningkat pada area-area dengan nilai harapan dan ragam y yang besar. Hal ini sesuai dengan apa yang disebut oleh Zhang, et all. 2002 bahwa perubahan sedikit saja pada nilai ragam mengakibatkan ragam yang besar pada nilai dugaan ϕ.



Gambar 2. MARE dan RRMSE bagi model null Empirical Bayes Poisson-Gamma

Ketelitian metode MLE dibanding MME ini berasal dari pembobot yang lebih besar dari pembobot MME (gambar 3). MME lebih banyak ditentukan oleh penduga sintetik yang dalam kasus ini diperoleh dengan mengabaikan pengaruh X. Sebaliknya, pada MLE bobot bagi penduga langsung lebih besar.

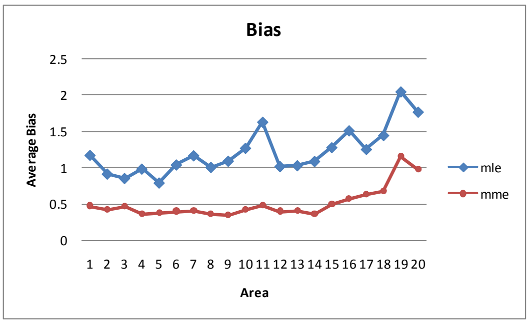


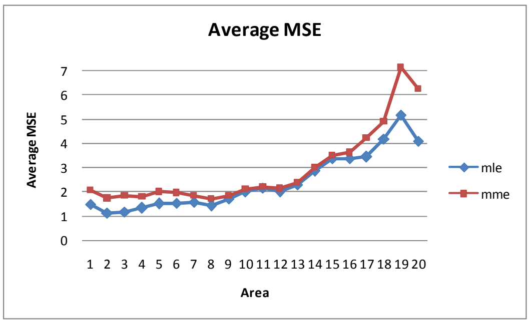


Gambar 3. Rataan Pembobot Seluruh Area dan per Area bagi model null
Empirical Bayes Poisson-Gamma

Dengan Peubah Penjelas

Pada skenario ini peubah X dimodelkan dengan penduga maksimum likelihood dalam model regresi binomiaal-negatif. Nilai penduga bayes empirik ditentukan oleh dua hal (i) pendugaan parameter regresi sebagai penduga tak langsung dan (ii) pembobot komposit bagi penduga langsung dan tak langsung. Namun dalam model poisson-gamma ini keduanya sangat tergantung pada pendugaan parameter regresi.

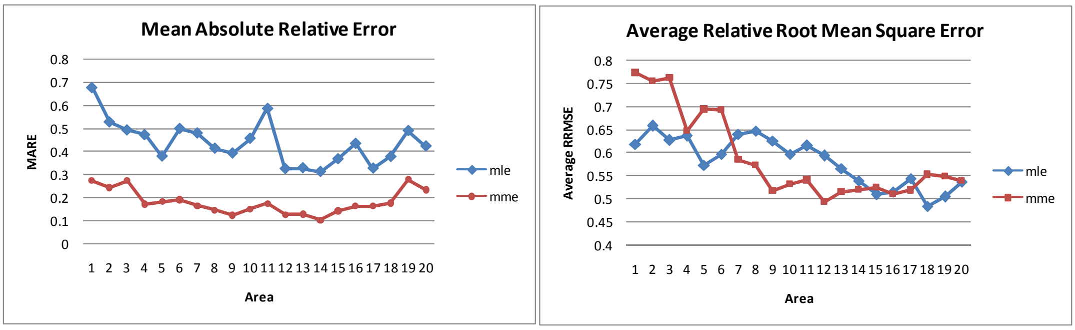




Gambar 4. Rataan Bias dan Rataan MSE bagi Model Linier
Empirical Bayes Poisson-Gamma

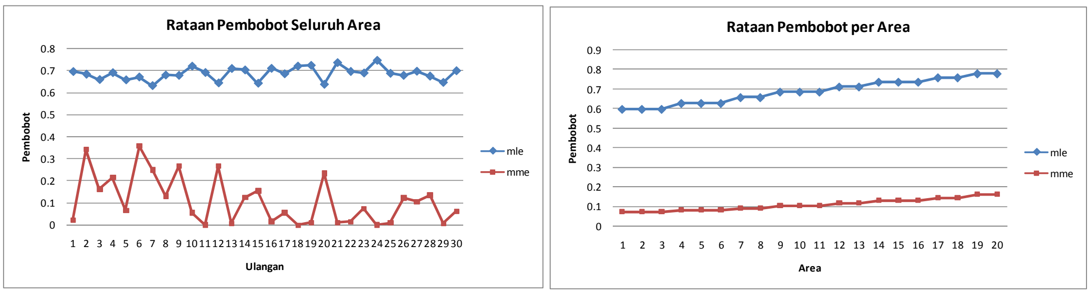
Baik MME maupun MLE tampak memiliki bias yang menaik dengan naiknya nilai harapan dan ragam y. Namun bias dari MLE tampaknya selalu lebih besar dari bias MME (gambar 4). Demikian pula dengan bias relatifnya, tampak MLE selalu menghasilkan bias yang lebih tinggi dari MME.

Namun gambar 5 menunjukkan bahwa pada MSE terjadi yang sebaliknya, MLE tampak lebih baik. Hal ini menunjukkan bahwa ragam dari MLE sangat rendah, karena dengan bias yang lebih tinggi MLE memiliki MSE yang lebih redah. Artinya MLE memberikan penduga area kecil yang lebih baik ketelitiannya.



Gambar 5. Rataan Bias dan Rataan MSE bagi Model Linier Empirical Bayes Poisson-Gamma

Bila kita perhatikan gambar 6, tampak besarnya pembobot komposit pada kedua metode, terlihat bahwa metode momen selalu memberikan pembobot yang rendah. Artinya, penduga area kecilnya akan lebih banyak ditentukan oleh penduga tak langsung yaitu pengaruh X'β.



Gambar 6. Rataan Pembobot Seluruh Area dan per Area bagi Model Linier Empirical Bayes Poisson-Gamma

Bias pada MLE diperkirakan berasal dari penduga langsung, meski secara teoritik penduga ini tak bias namun pada kasus poisson, nilainya sangat rentan terhadap nilai tengah yang mendekati nol.

PENUTUP

Beberapa hal dapat kita catat disini adalah:

1. Secara umum pendugaan parameter dispersi dengan MLE memberikan penduga area kecil yang lebih teliti meskipun tidak sangat tepat, baik itu melibatkan peubah penjelas ataupun tidak.
2. Ketelitian metode MLE dibanding MME pada model null berasal dari pembobot yang lebih besar dari pembobot MME. MME lebih banyak ditentukan oleh penduga sintetik yang dalam kasus ini diperoleh dengan mengabaikan pengaruh X. Sebaliknya, pada MLE bobot bagi penduga langsung lebih besar.
3. Bias pada MLE diperkirakan berasal dari penduga langsung, meski secara teoritik penduga ini tak bias namun pada kasus poisson, nilainya sangat rentan terhadap nilai tengah yang mendekati nol.
4. Penduga lain bagi parameter dispersi, yang mungkin dapat dikaji adalah penduga yang berbasis sisaan model yaitu *Deviance scale* dan *Pearson scale.*

DAFTAR PUSTAKA

Clayton, D.G. and Kaldor, J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43, 671–682.

Gill J. 2002. *Bayesian Methods:A Social and Behavioral Sciences Approach.* Boca Raton: Chapman and Hall.

Dobson, A.J. 1990. An Introduction to generalized linear models. Chapman and Hall, New York.

Jiang, J., P. Lahiri, S. Wan. 2002. A Unified Jackknife Theory For Empirical Best Prediction With M-Estimation. The Annals of Statistics. Vol. 30, No. 6, 1782–1810

Kurnia A, KA Notodiputro. 2006. Penerapan Metode *Jackknife* dalam pendugaan Area Kecil. Forum Statistika dan Komputasi, April 2006, p:12-15.

Kismiantini. 2007. Pendugaan Statistik Area Kecil Berbasis Model Poisson-Gamma [Tesis] Bogor: Institut Pertanian Bogor, Fakultas Matematika dan Pengetahuan Alam.

Lawless, J.F. Negative Binomial and Mixed Poisson Regression. The Canadian Journal of Statistics 15, pp. 209-225, 1987.

McCullagh, P. and J.A. Nelder. 1989. Generalized Linear Models. 2nd ed. Chapman and Hall, London.

Power J. H. & E. B. Moser, 1999. Linear model analysis of net catch data using the negative binomial distribution. Can. J. Fish. Aquat. Sci. 56: 191–200.

Rao JNK. 2003. *Small Area Estimation*. New York: John Wiley & Sons.

Ruoyan, M. 2004. Estimation of Dispersion Parameters in GLMs with and without Random Effects. Mathematical Statistics. Stockholm University Examensarbete 2004:5. http://www.matematik.su.se/matstat.

Wakefield J. 2006. *Disease mapping and spatial regression with count data*. http://www.bepress.com/uwbiostat/paper286.pdf [24 April 2008].

Zang Y, Z. Ye, & D. Lord, 2004. Estimating the Dispersion Parameter of the Negative Binomial Distribution for Analyzing Crash Data Using a Bootstrapped Maximum Likelihood Method. Zachry Department of Civil Engineering. Texas A&M University. Working Paper.