PERBANDINGAN REGRESI BINOMIAL NEGATIF DAN REGRESI M-KUANTIL PADA DATA *EXCESS ZEROES* DALAM KONTEKS *SMALL AREA ESTIMATION*

SKRIPSI

Oleh
Mohammad Zulfi
NIM 121810101061

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016

PERBANDINGAN REGRESI BINOMIAL NEGATIF DAN REGRESI M-KUANTIL PADA DATA *EXCESS ZEROES* DALAM KONTEKS *SMALL AREA ESTIMATION*

SKRIPSI

Diajukan Guna Melengkapi Tugas Akhir Matematika Dan Memenuhi Salah Satu Syarat
Untuk Menyelesaikan Program Studi Matematika (S1) Dalam Gelar Sarjana Sains

Oleh
Mohammad Zulfi
NIM 121810101061

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Kedua orang tua yang memberikan seluruh semangat dan doanya selama perjalanan studi;
2. Dr. Alfian Futuhul Hadi, S.Si, M.Si selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si., M.Si selaku Dosen Pembimbing Anggota yang telah membimbing secara intensif dalam menyempurnakan tugas akhir ini;
3. Almamater Jurusan Matematika FMIPA Universitas Jember;
4. keluarga besar BATHIC'S 12 yang telah menemani perjuangan studi di UNIVERSITAS JEMBER;
5. keluarga besar CTH yang telah membantu di setiap kesusahan;
6. Desi Ratna Fainita yang telah memberikan motivasi selama ini;
7. adik-adikku Dwi, Salsa, Mahrita, Yanti dan Rency yang telah banyak memberikan support dan bantuannya;
8. semua pihak yang membantu penulis dalam penyelesaian tugas akhir.

MOTTO

"Bukan tentang seberapa besar yang kita dapat, tetapi seberapa besar usaha kita untuk mendapatkannya"
(Anonim)

"jika ragu dalam melakukan sesuatu, sebaiknya tanya kepada diri sendiri, apa yang kita inginkan esok hari dari apa yang telah kita lakukan sebelumnya"
(Jonh Lubbock)

"Sesungguhnya dimana ada kesulitan disitu ada kelapangan"
(Al-Insyirah, Ayat 5)*)

"Dan memberinya rezeki dari pintu yang tidak diduga-duga olehnya. Barangsiapa yang bertawakal kepada Allah, maka Tuhan akan mencukupkan kebutuhannya. Bahkan sesungguhnya Allah pelaksana semua peraturan-Nya. Allah juga telah menjadikan segala-galanya serba berukuran"
(Ath-Thalaq, Ayat 3)*)

---

*Departemen Dalam Negeri Republik Indonesia. 1978. *Al-Qur'an & Tafsir, Huruf Arab & Latin.* Bandung : Firma SUMATRA.

PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Mohammad Zulfi

NIM : 121810101061

menyatakan dengan sebenarnya bahwa karya ilmiah yang berjudul "Perbandingan Regresi Binomial Negatif dan Regresi M-kuantil pada Data *Excess Zeroes* dalam Konteks *Small Area Estimation*" adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi manapun, dan bukan karya jiplakan. Saya bertanggungjawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik apabila ternyata dikemudian hari pernyataan ini tidak benar.

Jember, Juni 2016

Yang menyatakan,

Mohammad Zulfi

NIM 121810101061

SKRIPSI

PERBANDINGAN REGRESI BINOMIAL NEGATIF DAN REGRESI M-KUANTIL PADA DATA *EXCESS ZEROES* DALAM KONTEKS *SMALL AREA ESTIMATION*

Oleh

Mohammad Zulfi

NIM 121810101061

Pembimbing

Dosen Pembimbing Utama : Dr. Alfian Futuhul Hadi, S.Si, M.Si.

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si, M.Si

PENGESAHAN

Skripsi berjudul "**Perbandingan Regresi Binomial Negatif dan Regresi M-kuantil pada Data *Excess Zeroes* dalam Konteks *Small Area Estimation***" telah diuji dan disahkan pada:

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Tim Penguji:

Dosen Pembimbing Utama,	Dosen Pembimbing Anggota,
Dr. Alfian Futuhul Hadi, S.Si, M.Si. NIP. 197407192000121001	Dian Anggraeni, S.Si, M.Si. NIP. 198202162006042002
Penguji I,	Penguji II,
Dr. Mohammad Fatekurohman, S.Si., M.Si. NIP. 196906061998031001	Drs. Rusli Hidayat, M.Sc. NIP. 196610121993031001

Mengesahkan

Dekan,

Drs. Sujito,Ph.D.

NIP. 196102041987111001

RINGKASAN

Perbandingan Regresi Binomial Negatif dan Regresi M-kuantil pada Data *Excess Zeroes* dalam Konteks *Small Area Estimation*; Mohammad Zulfi, 121810101061; 2016; 30 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

Small Area Estimation (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi. Teknik pendugaan ini memanfaatkan data dari domain besar untuk menduga peubah yang menjadi perhatian pada domain yang lebih kecil. Pendugaan sederhana area kecil yang didasarkan pada penerapan model desain penarikan contoh (*design-based*) disebut sebagai pendugaan langsung (*direct estimation*). Pendugaan ini tidak mampu memberikan ketelitian yang cukup bila ukuran contoh dalam area yang menjadi perhatian berukuran kecil, sehingga statistik yang diperoleh akan menjadi ragam yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei. Oleh karena itu, dikembangkan teknik pendugaan alternatif untuk meningkatkan keefektifan ukuran contoh dan menurunkan galat baku yakni penduga tak langsung (*indirect estimation*).

Salah satu contoh penggunaan SAE adalah pendugaan proporsi rumah tangga miskin di kabupaten Sampang dengan memanfaatkan hasil survei proporsi rumah tangga miskin propinsi Jawa Timur. Propinsi Jawa Timur terdiri dari 38 kabupaten dan kota. Pengambilan sampel pada survei skala propinsi ini tentu merepresentasikan kabupaten dan kota yang ada. Jika pada survei ini terambil 200 sampel acak dan 10 di antaranya terambil dari kabupaten Sampang. Di waktu yang sama Pemerintah Kabupaten Sampang akan melakukan pendugaan proporsi survei rumah tangga miskin di kabupaten Sampang. Hal ini dapat dibantu dengan hasil survei skala propinsi dengan memanfaatkan teknik SAE. Sebab 10 sampel yang berasal dari kabupaten Sampang belum cukup merepresentasikan 14 kecamatan yang ada di kabupaten ini

Terdapat beberapa metode pendugaan pada SAE yang telah dikembangkan khususnya menyangkut metode berbasis model (*model-based estimation*) sebagai alternatif pendugaan langsung. Metode tersebut adalah *Empirical Best Linier Unbiassed Prediction* (EBLUP), *Empirical Bayes* (EB) dan *Hierarchical Bayes* (HB). Metode EB dan HB adalah yang lebih umum digunakan untuk menangani data kontinu, biner maupun cacahan (Kismiantini, 2007).

Beberapa penelitian tentang SAE telah dilakukan dengan beberapa model regresi diantaranya adalah model Binomial Negatif dan model M-kuantil. Namun yang sering terjadi adalah data yang mengalami *excess zeroes*. Hal ini menyebabkan overdispersi atau varians dan rataannya tidak sebanding. Anggraini (2014) menunjukkan bahwa model Binomial Negatif memberikan estimasi area yang lebih baik dibandingkan dengan model Poisson. Sedangkan Oktarin (2015) dalam penelitiannya menyimpulkan bahwa model M-kuantil memberikan estimasi area yang lebih baik dibandingkan dengan model Poisson. Namun data yang digunakan pada kedua penelitian tersebut merupakan data *count* yang berdistribusi Poisson. Pada faktanya *excess zeroes* sangat sering terjadi pada suatu pengamatan atau survei. *Excess zeroes* merupakan suatu kejadian dimana data pengamatan memiliki nilai 0 yang berlebih.

Penelitian ini merupakan perbandingan estimasi area antara model Binomial Negatif dan model M-kuantil dengan data mengalami *excess zeroes*. Data pada penelitian ini 50% bagiannya mengandung nilai 0. Artinya data yang sudah didapatkan menunjukkan adanya *excess zeroes*. Pada pendugaan area yang dilakukan baik dengan 20 area, 35 area dan 50 area menunjukkan bahwa model M-kuantil lebih baik dalam menduga nilai 0 dibandingkan dengan model Binomial Negatif. Namun secara keseluruhan model Binomial Negatif lebih baik dalam menduga area dibandingkan dengan model M-kuantil. Hal ini ditunjukkan dengan nilai *Mean Square Error* (MSE) pada model Binomial Negatif lebih kecil dibandingkan nilai MSE pada model M-kuantil.

PRAKATA

Puji syukur penulis panjatkan atas kehadirat Allah SWT, yang telah melimpahkan rahmat, taufiq, dan hidayah-Nya sehingga skripsi yang berjudul "Perbandingan Regresi Binomial Negatif dan Regresi M-kuantil pada Data *Excess Zeroes* dalam Konteks *Small Area Estimation*" dapat terselesaikan. Skripsi ini disusun guna untuk melengkapi tugas akhir dan salah satu syarat untuk menyelesaikan program sarjana (S1) Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Jember. Sholawat serta salam senantiasa tercurahkan atas junjungan Nabi Muhammad SAW. Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Dr. Alfian Futuhul Hadi, S.Si, M.Si. selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si, M.Si selaku Dosen Pembimbing Anggota yang telah menyisihkan sedikit waktunya untuk membimbing penulis selama menyusun tugas akhir ini;
2. Dr. Mohammad Fatekurohman, S.Si., M.Si. selaku Dosen Penguji I dan Drs. Rusli Hidayat, M.Sc. selaku Dosen Penguji II yang telah banyak memberikan kritik dan saran dalam penulisan skripsi ini;
3. teman-teman Bathic' 12 yang telah memberikan motivasinya;
4. Umah, Fifit, Dwindah serta rekan-rekan statistik yang lainnya;
5. Desi Ratna Fainita yang banyak memberikan motivasi selama ini;
6. Adik-adikku Salsa, Rency, Dwi, Mahrita dan Yanti yang banyak memberikan bantuan selama ini;
7. semua pihak yang tidak dapat disebutkan satu per satu.

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

Halaman

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Statistik area kecil *(Small Area Estimation)* merupakan suatu teknik statistika untuk menduga parameter-parameter sub populasi yang ukuran contohnya kecil. Teknik pendugaan ini memanfaatkan data dari domain besar untuk menduga peubah yang menjadi perhatian pada domain yang lebih kecil. Pendugaan sederhana area kecil yang didasarkan pada penerapan model desain penarikan contoh (*design-based*) disebut sebagai pendugaan langsung (*direct estimation*). Pendugaan ini tidak mampu memberikan ketelitian yang cukup bila ukuran contoh dalam area yang menjadi perhatian berukuran kecil, sehingga statistik yang diperoleh akan menjadi ragam yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei. Oleh karena itu, dikembangkan teknik pendugaan alternatif untuk meningkatkan keefektifan ukuran contoh dan menurunkan galat baku yakni penduga tak langsung (*indirect estimation*), pendugaan ini bersifat meminjam kekuatan dari pengamatan contoh area yang berdekatan dengan memanfaatkan informasi tambahan yakni dari data sensus dan catatan administratif (Rao, 2005).

Statistik area kecil (SAE) menjadi perhatian serius dalam beberapa tahun ini. Hal tersebut ditunjukkan dengan meningkatnya usaha untuk memperoleh penduga area kecil yang dapat diandalkan. Pentingnya statistik area kecil semakin dirasakan pula seiring dengan banyaknya negara yang mengubah sistem ketatanegaraannya dari sentralisasi ke desentralisasi (Fransisco, 2003). Salah satunya adalah Indonesia yang menerapkan sistem otonomi daerah. Pada sistem ini, pemerintah pusat akan memindahkan berbagai rencana dan program sosial ekonomi ke pemerintah daerah (yakni propinsi, kabupaten/kota, kecamatan, desa/kelurahan), sehingga statistik hasil survei yang dirancang untuk level nasional diharapkan dapat digunakan untuk memperoleh informasi pada area yang lebih kecil. Hal ini sangat membantu terhadap survei yang dilakukan di skala propinsi atau kabupaten untuk negara seperti Indonesia.

Indonesia memiliki jumlah penduduk sangat besar yaitu 248,8 juta jiwa, maka diperlukan sebuah survei atau sensus untuk menyediakan data representatif mengenai jumlah penduduk, ekonomi, sosial, wilayah, lingkungan dan pendidikan. Sensus dan survei memiliki peranan penting untuk pengambilan keputusan yang berbasis pada sebuah data. Tujuan utama dari survei ialah untuk mengumpulkan data dimana yang diselidiki adalah parameter suatu populasi. Data yang diperoleh dari hasil survei merupakan data perkiraan atau *estimated value* (Supranto, 2000). Secara luas survei tidak hanya digunakan untuk menduga parameter populasi, tetapi untuk menduga keragaman sub populasi atau domain. Domain didefinisikan sebagai daerah geografik, sosiodemografi dan sebagainya. Sampel yang diperlukan untuk survei dilakukan dengan data skala nasional dan tidak semua area (kecil) menjadi sampel seperti kabupaten atau kota bahkan mungkin level kecamatan dan desa/kelurahan. Oleh karena itu, dibutuhkan suatu metode pendugaan yang memperoleh informasi pada level area (kecil) tersebut.

Beberapa penelitian mengenai SAE telah dilakukan oleh beberapa orang diantaranya adalah Anggraini (2014). Penelitian ini memberikan kesimpulan bahwa pendekatan Bayes empirik pada model campuran Poisson-Gamma (Binomial Negatif) menghasilkan nilai pendugaan *small area* yang lebih baik apabila dalam model memperhatikan nilai dugaan parameter dispersi yang menyebabkan overdispersi dibandingkan dengan model Poisson yang mengabaikan overdispersi.

Model Binomial Negatif merupakan model yang memuat suatu parameter ϕ, dengan ϕ merupakan parameter dispersi. Metode pendugaan ϕ telah banyak di usulkan salah satunya yaitu *Method of Moments Estimate* (MME). Metode ini merupakan metode yang cukup sederhana untuk menduga ϕ. Estimasi dari parameter dispersi ini sangat diperlukan dalam memperbaiki penduga Bayes empirik. Pendugaan ini berperan untuk mendapatkan ϕ yang akan digunakan sebagai *hyperparameter (*Hadi,A.F.,dkk, 2008).

Penelitian yang lain juga dilakukan oleh Oktarin (2015) tentang penanganan overdispersi menggunakan regresi M-kuantil dengan *resampling bootstrap* untuk SAE. Penelitian ini menyimpulkan bahwa model Regresi M-kuantil menghasilkan nilai pendugaan yang lebih baik daripada model Poisson pada pendugaan area kecil.

Hal ini ditunjukkan dengan nilai rata-rata MSE *resampling bootstrap* yang menunjukkan model Regresi M-Kuantil lebih kecil dibandingkan model Poisson.

Metode Regresi Kuantil memiliki kelebihan yaitu dapat mengetahui perlakuan antara variabel bebas dan variabel tak bebas pada bagian awal, tengah, atau akhir dari data. Regresi M-kuantil merupakan bagian dari Regresi Kuantil, dimana perlakuan antara variabel bebas dan variabel tak bebas dapat dilihat pada setiap kuantil yang diinginkan. Sehingga diperoleh model terbaik dari beberapa model yang telah didapatkan (Dewi, 2013).

Kedua penelitian diatas menggunakan data yang sama yaitu data simulasi yang mengalami overdispersi dengan 20 unit level area. Simulasi adalah proses yang diperlukan untuk operasionalisasi suatu model untuk meniru tingkah laku sistem yang sesungguhnya. Sementara itu overdispersi merupakan keadaan dimana ragam dari respon lebih besar daripada meannya. Namun pada kenyataannya data yang kita peroleh tidak hanya mengalami overdispersi tetapi terjadinya *excess zeroes,* yaitu munculnya nilai nol yang berlebihan pada data pengamatan.

Peneliti tertarik untuk meneliti model Regresi Binomial Negatif dan model Regresi M-kuantil pada *small area estimation* menggunakan data simulasi pada data *excess zeroes*. Pada penelitian ini akan dilakukan simulasi dengan jumlah unit level area yang bervariasi yaitu 20, 35 dan 50 unit level. Sehingga dapat diketahui performa kedua model ini dengan jumlah level unit area yang berbeda.

1.2 Rumusan Masalah

Permasalahan yang dibahas dalam penelitian ini adalah bagaimana pendekatan model Regresi Binomial Negatif dan model Regresi M-kuantil dalam konteks *Small Area Estimation* pada data yang mengalami *excess zeroes* serta mengetahui perbandingan nilai *Mean Square Error* (MSE) dari kedua model regresi tersebut.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini menggunakan model level area pada *Small Area Estimation* (SAE). Data yang digunakan merupakan data yang berdistribusi *Zero Inflated Poisson* (ZIP).

1.4 Tujuan

Adapun tujuan yang ingin dicapai pada penelitian ini adalah mendapatkan model yang baik pada pendugaan area kecil dengan menggunakan pendekatan model Regresi Binomial Negatif dan Regresi M-kuantil. Selain itu dapat mengetahui performa dari kedua model ini pada data pengamatan yang mengalami *excess zeroes*.

1.5 Manfaat

Manfaat yang dapat diambil dari penelitian ini adalah dapat menambah wawasan tentang analisis statistika yaitu *Small Area Estimation* (SAE) mengenai pendugaan parameter menggunakan model Binomial Negatif dan M-kuantil serta mengetahui metode yang memiliki *Mean Square Error* (MSE) terkecil diantara keduanya pada data *excess zeroes*.

\

BAB 2 TINJAUAN PUSTAKA

2.1 *Small Area Estimation* (SAE)

Suatu area dikatakan kecil apabila sampel yang di ambil pada area tersebut tidak mencukupi untuk melakukan suatu pendugaan langsung dengan hasil dugaan yang akurat (Rao, 2003). *Small Area Estimation* (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter pada area kecil dengan memanfaatkan informasi dari dalam dan luar area (Longford, 2005). Permasalahan yang dapat terjadi dalam SAE adalah bagaimana menghasilkan dugaan parameter yang baik untuk ukuran sampel yang kecil pada suatu domain. Selain itu permasalahan yang dapat terjadi adalah bagaimana menduga *Mean Square Error* (MSE) dari dugaan parameter tersebut. Permasalahan tersebut dapat diatasi dengan cara meminjam informasi dari dalam area, luar area maupun dari luar survei.

Salah satu contoh penggunaan SAE adalah pendugaan proporsi rumah tangga miskin di kabupaten Sampang dengan memanfaatkan hasil survei proporsi rumah tangga miskin propinsi Jawa Timur. Propinsi Jawa Timur terdiri dari 38 kabupaten dan kota. Pengambilan sampel pada survei skala propinsi ini tentu merepresentasikan kabupaten dan kota yang ada. Jika pada survei ini terambil 200 sampel acak dan 10 di antaranya terambil dari kabupaten Sampang. Di waktu yang sama Pemerintah Kabupaten Sampang akan melakukan pendugaan proporsi survei rumah tangga miskin di kabupaten Sampang. Hal ini dapat dibantu dengan hasil survei skala propinsi dengan memanfaatkan teknik SAE. Sebab 10 sampel yang berasal dari kabupaten Sampang belum cukup merepresentasikan 14 kecamatan yang ada di kabupaten ini.

Pendugaan parameter pada suatu domain dalam SAE dilakukan dengan menggunakan pendugaan langsung (*direct estimation*) dan tidak langsung (*indirect estimation*). Pendugaan langsung merupakan pendugaan pada suatu domain berdasarkan data sampel dari domain tersebut dengan pendekatan yang digunakan adalah pendekatan berbasis rancangan (*design-based*). Pendugaan ini tidak mampu memberikan ketelitian yang cukup bila ukuran contoh dalam area yang menjadi perhatian berukuran kecil, sehingga statistik yang diperoleh memiliki ragam yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei. Oleh sebab itu dikembangkan model pendugaan tak langsung (*indirect estimation*). Pendugaan tidak langsung merupakan pendugaan pada suatu domain dengan cara menghubungkan informasi tambahan pada area tersebut. Metode dengan memanfaatkan informasi tambahan tersebut secara statistik memiliki sifat "meminjam kekuatan" (*borrowing strength*) informasi dari hubungan antara peubah respon dengan informasi yang ditambahkan. Dengan demikian, pendugaan tidak langsung mencakup data dari domain yang lain. Pendugaan SAE merupakan bentuk komposit dari penduga langsung dan penduga tak langsung. Prosedur SAE pada dasarnya memanfaatkan kekuatan informasi dari area sekitarnya (*neighbouring area*) dan sumber data di luar statistikanya yang diperoleh melalui pembentukan model yang tepat untuk meningkatkan efektifitas ukuran sampel (Kurnia, 2009).

Terdapat beberapa metode pendugaan pada SAE yang telah dikembangkan khususnya menyangkut metode berbasis model (*model-based estimation*) sebagai alternatif pendugaan langsung. Metode tersebut adalah *Empirical Best Linier Unbiassed Prediction* (EBLUP), *Empirical Bayes* (EB) dan *Hierarchical Bayes* (HB). Metode EB dan HB adalah yang lebih umum digunakan untuk menangani data kontinu, biner maupun cacahan (Kismiantini, 2007).

2.2 Model Area Kecil

Pada SAE terdapat dua jenis area kecil yaitu model level area dan model level unit.

2.2.1 Model Level Area

Model level area sangat penting jika data level unit tidak tersedia. Model tersebut didasarkan pada ketersediaan data pendukung yang hanya ada pada level area tertentu, misalkan $x_i = \left(x_1, x_2, \ldots, x_{pi}\right)^T$, dan parameter yang akan digunakan adalah θ_i yang diasumsikan mempunyai hubungan dengan data pendukung x_i. Data pendukung tersebut digunakan untuk membangun model linier :

$$\theta_i = x_i^T \beta + \; b_i v_i \,, \qquad \text{dengan } i = 1,2,3, \ldots, n \qquad (2.1)$$

dengan $\beta = \left(\beta_1, \beta_2, \ldots., \beta_p\right)^T$ meupakan $p \times l$ vector regresi dari parameter dan

$v_i^{iid} \sim N(0, \sigma_v^2)$ sebagai pengaruh acak yang diasumsikan *independent and identically distributed* (iid) dan menyebar normal. Kesimpulan mengenai θ_i dapat diketahui dengan mengasumsikan bahwa model penduga langsung y_i telah tersedia, yaitu :

$$y_i = \theta_i + e_i, \qquad \text{dengan } i = 1,2,3, \dots, n \qquad (2.2)$$

dan error sampling $e^{iid} \sim N(0, \sigma_e^2)$ dengan σ_e^2 diketahui. Dengan menggabungkan model (2.1) dan (2.2) maka diperoleh model :

$$y_i = x_i^T \beta + b_i v_i + e_i, \qquad \text{dengan } i = 1,2,3, \dots, n \qquad (2.3)$$

dengan b_i dan v_i saling bebas. Model (2.3) ini merupakan kasus khusus dari model campuran linier dikenal pula sebagai model Fay-Herriot dalam literatur area kecil (Rao, 2003).

2.2.2 Model Level Unit

Model level unit merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon. Model level unit mengasumsikan bahwa ada peserta unit $X_i = (x_{1ij}, x_{2ij}, \dots, x_{pij})^T$ ada untuk masing-masing populasi j dalam data masing-masing area kecil i namun kadang cukup dengan rata-rata populasi x_i diketahui saja. Selanjutnya peubah perhatian y_{ij} dianggap berkaitan dengan x_{ij} sehingga didapatkan model regresi tersarang:

$$y_i = x_i^T \beta + b_i v_i + e_i, \quad \text{dengan } i = 1,2, \dots, m; j = 1,2, \dots, N;$$

dengan asumsi $v_i^{iid} \sim N(0, \sigma_v^2)$ dan $e^{iid} \sim N(0, \sigma_e^2)$ serta v_i saling bebas (Rao,2003).

2.3 Regresi Binomial Negatif

Percobaan negatif binomial terdiri atas beberapa usaha dan setiap usaha dengan dua kemungkinan hasil yang dapat diberi nama sukses atau gagal dan dilakukan sampai tercapai sejumlah sukses tertentu. Model Regresi Binomial Negatif memiliki kegunaan yang sama dengan model Regresi Poisson yaitu untuk menganalisis hubungan antara suatu variabel respon data *count* dengan satu atau lebih variabel acak penjelas, tetapi model Regresi Binomial Negatif telah fleksibel dibandingkan dengan model Poisson karena mean dan variansi dari model Binomial

Negatif tidak harus sama. Model Regresi Binomial Negatif yang dibangun memiliki sebaran Binomial Negatif dengan parameter μ dan a. Sehingga sebaran Y_i menjadi:

$$f(y;\mu,a)=\frac{\Gamma\left(y_i+\frac{1}{a}\right)}{y_i!\,\Gamma\left(\frac{1}{a}\right)}\left(\frac{\frac{1}{a}}{\frac{1}{a}+\mu_i}\right)^{\frac{1}{a}}\left(\frac{\mu_i}{\frac{1}{a}+\mu_i}\right)^{y_i}$$

Distribusi Binomial Negatif $f(y;\mu,a)$ mempunyai rataan dan variansi

$$E(Y_i)=\mu_i$$

$$Var(Y_i)=\mu_i+a\mu_i^2$$

Y_i dalam Binomial Negatif adalah variabel yang berupa data *count* sehingga Y_i merupakan bilangan non-negatif, maka ekspektasi dari Y_i juga tidak mungkin negatif. Karena ruang nilai untuk $\mu_i=x_i\beta$ adalah bilangan riil pada interval $(-\infty,\infty)$, membuat model regresi tidak dapat digunakan untuk menganalisis data *count*. Untuk mengatasinya digunakan sebuah fungsi penghubung yang menghubungkan antara *fitted value* (μ_i) dengan prediktor linier $(x_i\beta)$. Hilbe (2011) menyatakan bahwa model Binomial Negatif pada umumnya menggunakan fungsi penghubung logaritma atau log *link* sebagai berikut

$$\ln\mu_i=x_i\beta$$

$$\ln\{E[Y_i|x_i]\}=\ln\mu_i=x_i\beta \qquad \text{untuk } i=1,2,\dots,n$$

$$\mu_i=\exp(x_i\beta)$$

Sehingga model regresi Binomial Negatif untuk memodelkan data *count* dapat dituliskan sebagai berikut:

$$f(y;\mu,a)=\frac{\Gamma\left(y_i+\frac{1}{a}\right)}{y_i!\,\Gamma\left(\frac{1}{a}\right)}\left(\frac{\frac{1}{a}}{\frac{1}{a}+\exp(x_i'\beta)}\right)^{y_i}\left(\frac{\exp(x_i'\beta)}{\frac{1}{a}+\exp(x_i'\beta)}\right)^{\frac{1}{a}}$$

2.4 Regresi M-Kuantil

Regresi Kuantil merupakan suatu metode pendekatan dalam analisis regresi yang menduga berbagai fungsi kuantil dari suatu distribusi *Y* sebagai fungsi dari *X* (Koenker dan Bassett, 1978). Metode Regresi Kuantil memiliki kelebihan yaitu dapat mengetahui perlakuan antara variabel bebas dan variabel tak bebas pada bagian awal, tengah atau akhir dari data. Regresi M-kuantil merupakan bagian dari

Regresi Kuantil, dimana perlakuan antara variabel bebas dan variabel tak bebas dapat dilihat pada setiap kuantil. Metode ini dapat dibentuk beberapa model untuk beberapa kuantil yang diinginkan. Sehingga diperoleh model terbaik dari beberapa model yang telah didapatkan (Dewi, 2013).

2.4.1 Regresi M-kuantil untuk Respon Kontinu

Menurut Chambers dan Tzavidis (2013) model regresi klasik meringkas perilaku rataan dari variabel acak Y pada setiap titik dalam himpunan kovariat X. Hal ini memberikan gambaran tidak lengkap dari suatu distribusi. Regresi Kuantil meringkas perilaku dari bagian yang berbeda dari distribusi bersyarat dari y pada setiap titik di x. Model regresi linier untuk kuantil dengan kuantil ke- q ditulis sebagai berikut :

$$Q_y(q|x_j) = x_i^T \beta_j \tag{2.4}$$

Dengan q bernilai $0 < q < 1$. Estimasi dari regresi ke-q parameter β_q diperoleh dengan meminimalkan

$$\sum_{j=1}^{n} |y_j - x_i^T \beta_j| \{(1-q)I(y_j - x_i^T \beta_j \leq 0) + qI(y_j - x_i^T \beta_j > 0)\}$$

M-kuantil order q untuk kepadatan bersyarat y diberikan himpunan kovariat x , $f(y|x)$ didefinisikan sebagai solusi $MQ_y(q|x;\psi)$ dari persamaan pendugaan $\int \psi_q\{y - MQ_y(q|x;\psi)\} f(y|x)dy - 0$, dimana ψ_q menunjukkan suatu fungsi pengaruh asimetris, yang merupakan turunan dari fungsi kerugian asimetris ρ_q. Menurut Chambers dan Tzavidis (2006) jika M-kuantil digunakan untuk mengestimasi daerah kecil dengan unit ke-j, maka persamaan (2.4) menjadi:

$$MQ_y(q|x;\psi) = x_i^T \beta_j \ , \tag{2.5}$$

2.4.2 Regresi M-Kuantil untuk data cacahan

Pendekatan yang sering digunakan dalam pemodelan dengan hasil rataan diskrit sebagai fungsi *predictor* yaitu dengan menggunakan GLM dengan mengasumsikan bahwa variabel respon mengikuti distribusi yang merupakan anggota dari keluarga distribusi eksponensial dan menggunakan fungsi link yang sesuai. Untuk data cacahan, distribusi yang tepat adalah Poisson dan fungsi link logaritma.

Pada data cacahan ditentukan spesifikasi kontinu yang tepat (dalam q) pada $MQ_y(q|x;\psi)$. Spesifikasi yang jelas untuk data cacahan adalah spesifikasi log-linier yaitu dengan mengganti (2.5) oleh

$$MQ_y(q|x;\psi) = t_j \exp(x_i^T \beta_j) \tag{2.6}$$

dengan t_j adalah syarat offset. Untuk estimasi β_q menggunakan persamaan estimasi yang didefinisikan sebagai berikut

$$\psi(\beta) = n^{-1} \sum_{j=1}^{n} \{(\psi(r_j) w(x_j) \frac{1}{\sigma(\mu_j)} \mu_j' - a(\beta)\} = 0 \tag{2.7}$$

dimana r_j adalah residu Pearson, $E[Y_i] = \mu_j$, turunan terhadap β adalah μ_i' , $Var[Y_i] = \sigma^2(\mu_j)$ dan $a(\beta) = n^{-1} \sum_{j=1}^{n} E[\psi(r_j)] w(x_j) \frac{1}{\sigma(\mu_j)} \mu_j'$ untuk memastikan konsistensi Fisher dari estimator (Chambers dan Tzavidis, 2013).

2.5 Estimasi Area Kecil Oleh Model Regresi Binomial Negatif

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 (binomial negatif). 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. Poisson-gamma bagi area level yang digunakan yaitu:

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

$$\boldsymbol{\theta_i = X'\beta + v_i}$$

dengan $y_i \sim Poisson(\theta_i)$, $E[y_i] = X'\beta$, $Var(y_i) = E[y_i] = X'\beta$

$E[e_i] = 0$; $Var(e_i) = \sigma_{e_i}^2$; $E[v_i] = 0$; $Var(v_i) = \sigma^2$

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

Konsep pendugaan area kecil dengan model Binomial Negatif merupakan komposisi dari pendugaan langsung dan pendugaan tak langsung pada model ini. Model dari pendugaan ini dirumuskan dengan

$$\hat{\theta}_{BN} = \pi\hat{y}_D + (1-\pi)\hat{y}_{BN}$$

dengan $\hat{y}_D$ merupakan penduga langsung serta $\hat{y}_{BN}$ merupakan penduga regresi. Sementara itu nilai π diperoleh dari $\pi = \frac{\mu}{\mu+\exp(\phi)}$ dengan ϕ merupakan parameter dispersi (Anggraeni, 2014).

2.6 Estimasi Area Kecil Oleh Model Regresi M-Kuantil

Konsep kunci penerapan metode M-kuantil pada data dengan struktur grup yaitu dengan mengidentifikasi keunikan 'koefisien M-kuantil' yang terkait dengan setiap data yang diamati. Koefisien tersebut kemudian dirata-rata dalam beberapa cara yang cocok. Pengamatan yang lebih membuat grup untuk mendefinisikan level grup koefisien M-Kuantil, yang dapat digunakan untuk mengkarakterisasi distribusi dari $y|x$ dalam grup.

Pada kasus y kontinu, koefisien M-kuantil untuk pengamatan j hanya didefinisikan sebagai solusi unik q_j untuk persamaan $y_j = \widehat{MQ_y}(q|x;\psi)$. Namun, untuk data cacahan persamaan tersebut tidak memiliki solusi ketika $y_j = 0$. Menurut Chambers *et al* (2013), untuk mengatasi masalah tersebut dengan menggunakan definisi sebagai berikut

$$\widehat{MQ_y}(q|x;\psi) = \begin{cases} k(x_j), & y_j = 0 \\ y_j, & y_j = 1,2,\ldots \end{cases}$$

Kemungkinannya yaitu $k(x_j) = \widehat{MQ_y}(q|x;\psi)$ dimana q_{min} menunjukkan q-value terkecil dari q-value yang digunakan untuk menentukan nilai q_j dari unit yang diamati. Namun, secara tidak langsung menyatakan bahwa $q_j = q_{min}$ setiap kali $y_j = 0$ terlepas dari nilai x_j. Untuk mengatasi masalah tersebut yaitu dengan menggunakan definisi q_j untuk kasus Bernoulli. Ini berarti bahwa pengamatan dengan nilai $y_1 = 0$ sesuai dengan q-value terkecil dari pada yang lain dengan $y_2 =$

0 ketika $\widehat{MQ_{y1}}(0.5|x_1;\psi) > \widehat{MQ_{y2}}(0.5|x_2;\psi)$. Untuk mendefinisikan yaitu dengan cara mengatur

$$k(x_j) = \min\{1-\epsilon, \left[\widehat{MQ_y}(0.5|x_j;\psi)\right]^{-2}\}$$

dimana $\epsilon > 0$ adalah konstanta positif kecil. Kemudian koefisien M-kuantil untuk unit j adalah q_j, dimana

$$\widehat{MQ_y}(q_j|x_j;\psi) = \begin{cases} \min\{1-\epsilon, \dfrac{1}{\exp(x_j^T \beta_{0.5})}\}, & y_j = 0 \\ y_j, & y_j = 1,2,\ldots \end{cases}$$

Koefisien M-kuantil di area khusus d yaitu $\hat{\theta}_d$ dapat didefinisikan sebagai nilai rata-rata dari koefisien sampel M-kuantil di area d dengan syarat terdapat sampel pengamatan di daerah kecil d. Adapun nilai pendugaan SAE dengan menggunakan model M-kuantil adalah sebagai berikut

$$\hat{\theta} = N^{-1}(\hat{y}_D + \widehat{MQ_y}(q_j|x_j;\psi))$$

dimana $\hat{y}_D$ merupakan penduga langsung serta $\widehat{MQ_y}(q_j|x_j;\psi) = \exp\{x^T\beta_{0.5}\}$ (Tzavidiz, *et al* , 2014)

2.7 Overdispersi

Masalah yang sering terjadi pada regresi Poisson adalah adanya overdispersi. Menurut Hilbe (2011), overdispersi pada regresi Poisson terjadi ketika ragam dari respon lebih besar daripada meannya. Jika pada data cacah terjadi overdispersi namun tetap digunakan model Poisson, maka estimasi parameter koefisien regresinya tetap konsisten tetapi tidak efisien karena berdampak pada nilai standart error yang menjadi *under estimate* sehingga kesimpulannya menjadi tidak valid. Overdispersi pada regresi Poisson dapat dideteksi dengan nilai Pearson chi-squared (χ^2) atau devian yang dibagi dengan derajat bebasnya. Jika kedua nilai ini lebih besar dari 1, maka dikatakan terjadi overdispersi pada data.

Overdispersi memiliki dampak yang sama dengan pelanggaran asumsi homokedastisitas dalam model regresi linier dan menyebabkan nilai devians model menjadi sangat besar sehingga model yang dihasilkan menjadi kurang tepat. Salah satu cara untuk mengatasi adanya kasus overdispersi dalam regresi Poisson adalah dengan mengganti asumsi distribusi Poisson dengan distribusi lain yang lebih fleksibel.

2.8 *Mean Square Error* (MSE)

Mean square error adalah metode untuk mengevaluasi metode peramalan dengan masing-masing kesalahan atau sisa dikuadratkan. Kemudian dijumlahkan dan dibagi dengan jumlah observasi. Pendekatan ini mengatur kesalahan peramalan yang besar karena kesalahan-kesalahan itu dikuadratkan. Suatu teknik yang menghasilkan kesalahan moderat mungkin lebih baik untuk salah satu yang memiliki kesalahan kecil tapi kadang-kadang menghasilkan sesuatu yang sangat besar.

Selain menggunakan metode MSE, masih terdapat metode yang lainnya untuk mengevaluasi suatu metode peramalan. Salah satunya adalah dengan uji signifikansi koefisien β. Namun dalam konteks SAE, metode evaluasi yang digunakan adalah MSE. Hal ini disebabkan pada persamaan pendugaan SAE tidak muncul nilai β, melainkan nilai $\hat{\theta}$ yang merupakan nilai pendugaan area pada SAE. Adapun persamaan MSE dirumuskan sebagai berikut

$$MSE(y) = \frac{\sum_{i=1}^{n}\left(\theta_i - \hat{\theta}_i\right)^2}{n}$$

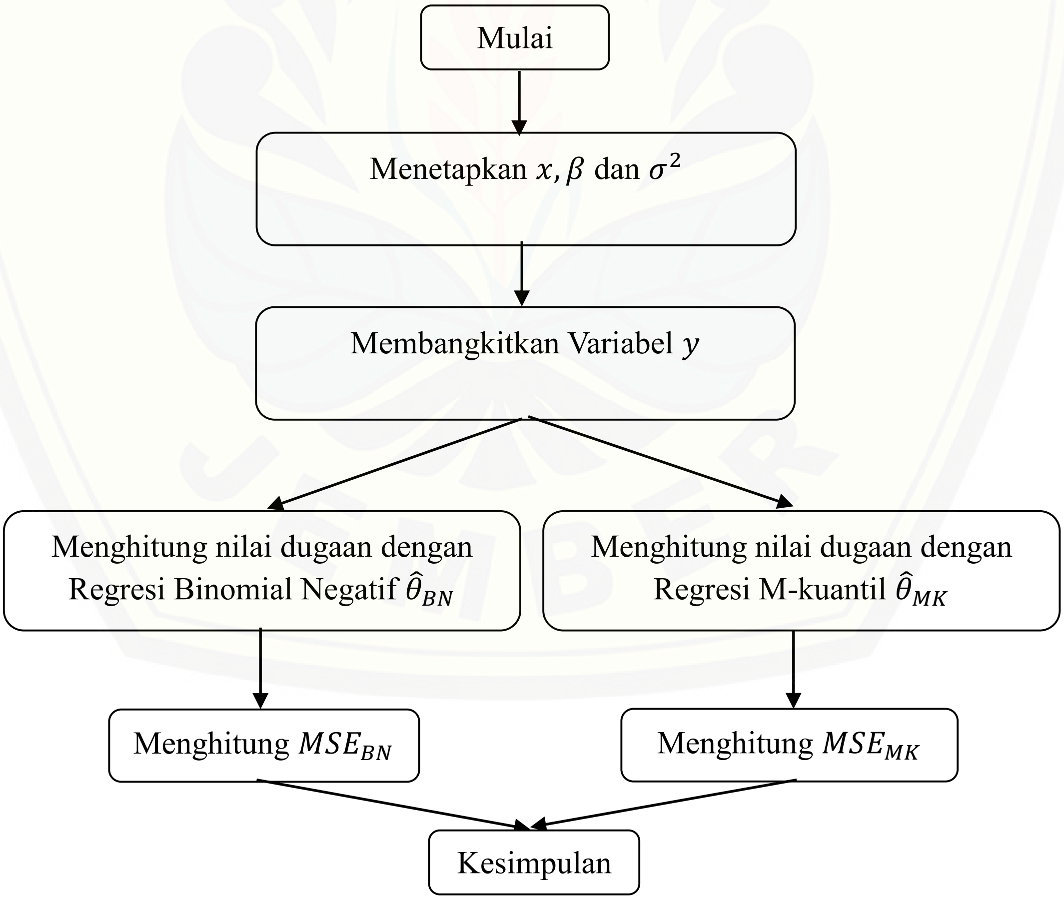
BAB 3. METODOLOGI PENELITIAN

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data simulasi dengan cara membangkitkan data berdistribusi *Zero Inflated Poisson* pada *Small Area Estimation* (SAE) melalui program R versi 2.13.0 dengan menetapkan variabel X , $\beta = (2{,}0; 0{,}3)$, dan $\sigma^2 = (2{,}0)$. Variabel yang digunakan adalah variabel bebas (X) dan variabel tak bebas (Y). Area kecil yang digunakan dalam data tersebut yaitu area 1 sampai area 20, area 1 sampai area 35 dan area 1 sampai area 50.

3.2 Langkah-langkah Simulasi

Langkah-langkah simulasi dalam penelitian ini disajikan dalam diagram alir pada Gambar 3.2 berikut.



Gambar 3.1 Skema Desain penelitian

Penjelasan dari masing-masing langkah penelitian pada skema desain penelitian di atas adalah sebagai berikut:

1. Menentukan nilai variabel bebas X, koefisien β serta nilai varians (σ^2). Nilai $\mu_i = x'\beta$. Panjang dari X beragam yaitu 20,35 dan 50 sesuai dengan banyaknya level unit area.
2. Menentukan nilai *direct estimation* $(\hat{y}_D)$ menggunakan metode *Zero Inflated Poisson* pada masing-masing level unit area.
3. Melakukan pendugaan parameter untuk model Binomial Negatif dan M-kuantil.
4. Melakukan pendugaan model Binomial Negatif dan M-kuantil pada SAE.
5. Menghitung *Mean Square Error* untuk model Binomial Negatif dan M-kuantil.
6. Kesimpulan yang menjawab rumusan masalah

3.3 Metode Analisis Data

Pengolahan data dalam penelitian ini menggunakan software R versi 2.13.0. Paket yang digunakan adalah `quantreg`. Fitting model menggunakan regresi sehingga fungsi R-nya sebagai berikut:

```
rq.fit(X, Y, tau=0.5, alpha=0.1, ci=FALSE,iid=TRUE,
       interp=TRUE, tcirt=TRUE)
```

dengan keterangan sebagai berikut:

a. `X` merupakan desain matrik.
b. `Y` merupakan variabel respon
c. `tau` merupakan kuantil yang di inginkan, jika `tau` diluar garis (0,1) seluruh proses diperkirakan.
d. `alpha` merupakan kemungkinan nominal *noncoverage* untuk selang kepercayaan 1-alpha adalah kemungkinan cakupan nominal dari selang.

e. `ci` merupakan fungsi logika jika “TRUE” maka menghitung selang kepercayaan untuk parameter dengan menggunakan metode rank inversi Koenker. Jika “FALSE” maka kembali hanya untuk estimasi koefisien.

f. `iid` merupakan fungsi logika jika “TRUE” maka rank inversi didasarkan pada asumsi kesalahan `iid` model, jika “FALSE” maka didasarkan pada asumsi kesalahan `iid`.

g. `interp` digunakan untuk mempertimbangkan interval yang interpolasi antara nilai-nilai parameter tepat dibawah *cutoff* dan nilai diatas *cutoff* ditetapkan.

h. `tcrit` merupakan fungsi logika jika T nilai kritis t yang digunakan jika F maka nilai normal yang digunakan.

BAB 5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan estimasi dengan model Binomial Negatif lebih baik dibandingkan estimasi dengan model M-kuantil pada data *excess zeroes* dalam konteks *Small Area Estimation*. Hal ini ditunjukkan dengan uji nilai *Mean Square Error* (MSE). Namun secara umum nilai MSE estimasi model Binmoial Negatif masih tergolong besar dibandingkan dengan beberapa penelitian sebelumnya. Hal ini disebabkan data pada penelitian ini merupakan data yang mengalami *excess zeroes*. Namun pada penelitian ini estimasi dengan model M-kuantil lebih baik daripada estimasi dengan model Binomial Negatif dalam menduga area bernilai 0.

5.2 Saran

Penelitian ini memberikan nilai MSE yang relatif besar sehingga saran untuk penelitian selanjutnya dapat dilakukan penanganan data *excess zeroes* dalam konteks *Small Area Estimation*.

DAFTAR PUSTAKA

Anggraini, A. 2014. *"*Penduga Momen untuk Parameter Dispersi pada Pendekatan Bayes Empirik Model Campuran Poisson-Gamma dalam Konteks *Small Area Estimation*". Tidak Diterbitkan. Skripsi. Jember: Jurusan Matematika FMIPA Universitas Jember.

Chambers, R. and Tzavidis, N. 2006. M-Quantile Models For Small Area Estimation.. *Biometrika*. **93**, 255-268.

Chambers, R. and Tzavidis, N. 2013. M-Quantile Regresion For Binary Data With Application To Small Area Estimation. *Biometrika*. **12**,1-24.

Dewi, L. A. 2013. "Estimasi Parameter Model Regresi M-kuantil Menggunakan Metode *Iterative Reweighted Least Square"*. Tidak Diterbitkan. Skripsi. Surakarta: Jurusan Matematika FMIPA Universitas Sebelas Maret.

Hadi, A. F.,dkk. 2008. *Penduga Maksimum Likelihood untuk Parameter Dispersi Model Poisson-Gamma dalam Konteks Pendugaan Area Kecil*. Seminar Nasional Matematika dan Pendidikan Matematika.

Hilbe, J. M. 2011. *Negative Binomial Regression Second Edition.* New York: Cambridge University Press.

Keswari,N.M.R.,dkk. 2014. *Perbandingan Regresi Binomial Negatif dan Regresi Generalisasi Poisson dalam Mengatasi Overdispersi*. E-jurnal Matematika Universitas Udayana.

Kismiantini. 2007. *Pendugaan Statistik Area Kecil Dengan Metode Empirical Constrained Bayes*. Seminar Nasional Matematika.

Koenker, R. & Bassett, G. 1978. Regression Quantiles. *Econometrika*. **46,** 33-50.

Kurnia, A. 2009. "Prediksi Terbaik Empirik Untuk Model Transformasi Logaritma di Dalam Pendugaan Area Kecil Dengan Penerapan Pada Data Susenas". Tidak Diterbitkan. Disertasi. Bogor: Program Pascasarjana Institut Pertanian Bogor.

Rao, J. N. K. 2003. *Small Area Estimation*. New Jersey. John Willey & Sons, Inc.

Supranto, J. 2005. *Statistik Teori dan Aplikasi Jilid 1&2*. Jakarta: Penerbit Erlangga.

Lampiran A. Data pengamatan 20 area

No	x_0	x_1	β_0	β_1	μ_i	y_D
1	1	0	2	0,3	2,0	4
2	1	0			2,0	2
3	1	0			2,0	0
4	1	1			2,3	3
5	1	1			2,3	0
6	1	1			2,3	2
7	1	2			2,6	2
8	1	2			2,6	0
9	1	2			2,6	4
10	1	3			2,9	0
11	1	3			2,9	0
12	1	3			2,9	0
13	1	4			3,2	6
14	1	4			3,2	1
15	1	4			3,2	0
16	1	5			3,5	2
17	1	5			3,5	0
18	1	5			3,5	2
19	1	6			3,8	0
20	1	6			3,8	0

Lampiran B. Data pengamatan 35 area

No	x_0	x_1	β_0	β_1	μ_i	y_D
1	1	0	2	0,3	2,0	0
2	1	0			2,0	0
3	1	0			2,0	0
4	1	1			2,3	0
5	1	1			2,3	0
6	1	1			2,3	0
7	1	2			2,6	0
8	1	2			2,6	2
9	1	2			2,6	0
10	1	3			2,9	0
11	1	3			2,9	0
12	1	3			2,9	0
13	1	4			3,2	3
14	1	4			3,2	0
15	1	4			3,2	0
16	1	5			3,5	0
17	1	5			3,5	4
18	1	5			3,5	0
19	1	6			3,8	4
20	1	6			3,8	6
21	1	6			3,8	0
22	1	7			4,1	3
23	1	7			4,1	2
24	1	7			4,1	5
25	1	8			4,4	0
26	1	8			4,4	5
27	1	8			4,4	6
28	1	9			4,7	0
29	1	9			4,7	4
30	1	9			4,7	4
31	1	10			5,0	8
32	1	10			5,0	2
33	1	10			5,0	5
34	1	11			5,3	0
35	1	11			5,3	0

Lampiran C. Data pengamatan 50 area

No	x_0	x_1	β_0	β_1	μ_i	y_D
1	1	0	2	0,3	2,0	0
2	1	0			2,0	0
3	1	0			2,0	3
4	1	1			2,3	5
5	1	1			2,3	0
6	1	1			2,3	2
7	1	2			2,6	1
8	1	2			2,6	0
9	1	2			2,6	0
10	1	3			2,9	0
11	1	3			2,9	0
12	1	3			2,9	0
13	1	4			3,2	0
14	1	4			3,2	1
15	1	4			3,2	0
16	1	5			3,5	7
17	1	5			3,5	0
18	1	5			3,5	0
19	1	6			3,8	0
20	1	6			3,8	0
21	1	6			3,8	5
22	1	7			4,1	0
23	1	7			4,1	9
24	1	7			4,1	3
25	1	8			4,4	7
26	1	8			4,4	0
27	1	8			4,4	1
28	1	9			4,7	0
29	1	9			4,7	3
30	1	9			4,7	4
31	1	10			5,0	0
32	1	10			5,0	1
33	1	10			5,0	5
34	1	11			5,3	3
35	1	11			5,3	0
36	1	11			5,3	0
37	1	12			5,6	0

No	x_0	x_1	β_0	β_1	μ_i	y_D
38	1	12			5,6	5
39	1	12			5,6	3
40	1	13			5,9	5
41	1	13			5,9	5
42	1	13			5,9	1
43	1	14			6,2	5
44	1	14			6,2	3
45	1	14			6,2	0
46	1	15			6,5	0
47	1	15			6,5	0
48	1	15			6,5	7
49	1	16			6,8	6
50	1	16			6,8	0

Lampiran D. Hasil pendugaan langsung dan tak langsung pada 20 area

Estimasi Binomial Negatif		Estimasi M-kuantil	
Direct	Indirect	Direct	Indirect
1,876	1,103	0,338	0,369
0,938	1,103	0,238	0,369
0,000	1,103	0,138	0,369
1,294	1,014	0,265	0,264
0,000	1,014	0,115	0,264
0,862	1,014	0,215	0,264
0,788	0,927	0,192	0,189
0,000	0,927	0,092	0,189
1,576	0,927	0,292	0,189
0,000	0,842	0,069	0,135
0,000	0,842	0,069	0,135
0,000	0,842	0,069	0,135
1,943	0,762	0,346	0,097
0,323	0,762	0,096	0,097
0,000	0,762	0,046	0,097
0,582	0,685	0,123	0,069
0,000	0,685	0,023	0,069
0,582	0,685	0,123	0,069
0,000	0,613	0,000	0,050
0,000	0,613	0,000	0,050

Lampiran E. Hasil pendugaan langsung dan tak langsung pada 35 area

Estimasi Binomial Negatif		Estimasi M-kuantil	
Direct	Indirect	Direct	Indirect
0,000	0,201	-0,016	0,020
0,000	0,201	-0,016	0,020
0,000	0,201	-0,016	0,020
0,000	0,266	0,000	0,028
0,000	0,266	0,000	0,028
0,000	0,266	0,000	0,028
0,000	0,348	0,016	0,398
0,361	0,348	0,074	0,398
0,000	0,348	0,016	0,398
0,000	0,448	0,033	0,055
0,000	0,448	0,033	0,055
0,000	0,448	0,033	0,055
0,883	0,566	0,136	0,077
0,000	0,566	0,050	0,077
0,000	0,566	0,050	0,077
0,000	0,701	0,067	0,108
1,457	0,701	0,182	0,108
0,000	0,701	0,067	0,108
1,763	0,848	0,198	0,151
2,644	0,848	0,256	0,151
0,000	0,848	0,084	0,151
1,599	1,001	0,187	0,211
1,039	1,001	0,158	0,211
2,599	1,001	0,244	0,211
0,000	1,151	0,118	0,294
2,991	1,151	0,261	0,294
3,589	1,151	0,289	0,294
0,000	1,293	0,135	0,411
2,686	1,293	0,249	0,411
2,686	1,293	0,249	0,411
5,901	1,420	0,380	0,573
1,475	1,420	0,209	0,573
3,688	1,420	0,295	0,573
0,000	1,529	0,169	0,800
0,000	1,529	0,169	0,800

Lampiran F. Hasil pendugaan langsung dan tak langsung pada 50 area

Estimasi Binomial Negatif		Estimasi M-kuantil	
Direct	Indirect	Direct	Indirect
0,000	0,656	-0,005	0,017
0,000	0,656	-0,005	0,017
1,184	0,656	0,054	0,017
2,060	0,684	0,100	0,020
0,000	0,684	0,000	0,020
0,824	0,684	0,040	0,020
0,429	0,713	0,025	0,023
0,000	0,713	0,005	0,023
0,000	0,713	0,005	0,023
0,000	0,742	0,011	0,026
0,000	0,742	0,011	0,026
0,000	0,742	0,011	0,026
0,000	0,772	0,016	0,030
0,464	0,772	0,036	0,030
0,000	0,772	0,016	0,030
3,375	0,801	0,162	0,035
0,000	0,801	0,022	0,035
0,000	0,801	0,022	0,035
0,000	0,831	0,028	0,040
0,000	0,831	0,028	0,040
2,500	0,831	0,128	0,040
0,000	0,860	0,033	0,047
4,660	0,860	0,213	0,047
1,557	0,860	0,093	0,047
3,748	0,890	0,179	0,054
0,000	0,890	0,039	0,054
0,535	0,890	0,059	0,054
0,000	0,919	0,045	0,062
1,659	0,919	0,105	0,062
2,212	0,919	0,125	0,062
0,000	0,948	0,050	0,072
0,570	0,948	0,070	0,072
2,853	0,948	0,150	0,072
1,764	0,977	0,116	0,083
0,000	0,977	0,056	0,083
0,000	0,977	0,056	0,083
0,000	1,005	0,062	0,096
3,025	1,005	0,162	0,096

1,815	1,005	0,122	0,096
3,110	1,033	0,167	0,111
3,110	1,033	0,167	0,111
0,622	1,033	0,087	0,111
3,192	1,061	0,173	0,128
1,915	1,061	0,133	0,128
0,000	1,061	0,073	0,128
0,000	1,088	0,079	0,147
0,000	1,088	0,219	0,147
4,583	1,088	0,204	0,147
4,024	1,114	0,204	0,170
0,000	1,114	0,084	0,170

Lampiran G. Perintah R pengamatan 20 area

```
> #SAE 20 unit level AREA
> rm(list=ls(all=TRUE))
> library(MASS)
> X0<-rep(1,20)
> X1<-c(0,0,0,1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6)
> X<-cbind(X0,X1)
> summary(X)
X0 X1
Min. :1 Min. :0.00
1st Qu.:1 1st Qu.:1.00
Median :1 Median :3.00
Mean :1 Mean :2.85
3rd Qu.:1 3rd Qu.:4.25
Max. :1 Max. :6.00
> b0<-2
> b1<- .3
> b<-as.matrix(c(b0,b1))
> miu<-X%*%b
> print(miu)
[,1]
[1,] 2.0
[2,] 2.0
[3,] 2.0
[4,] 2.3
[5,] 2.3
[6,] 2.3
[7,] 2.6
[8,] 2.6
[9,] 2.6
[10,] 2.9
[11,] 2.9
[12,] 2.9
[13,] 3.2
[14,] 3.2
[15,] 3.2
[16,] 3.5
[17,] 3.5
[18,] 3.5
[19,] 3.8
[20,] 3.8
> #membangkitkan Y direct
> library(gamlss)
> n<-length(X1)
> Y<-rZIP(n,miu,0.5)
> Y
```

```
[1] 4 2 0 3 0 2 2 0 4 0 0 0 6 1 0 2 0 2 0 0
> summary(Y)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0 0.0 0.5 1.4 2.0 6.0
> tabel<-data.frame(Y=Y,X0=X0,X1=X1)
> tabel
Y X0 X1
1 4 1 0
2 2 1 0
3 0 1 0
4 3 1 1
5 0 1 1
6 2 1 1
7 2 1 2
8 0 1 2
9 4 1 2
10 0 1 3
11 0 1 3
12 0 1 3
13 6 1 4
14 1 1 4
15 0 1 4
16 2 1 5
17 0 1 5
18 2 1 5
19 0 1 6
20 0 1 6
>
# ESTIMASI M-KUANTIL
> library(quantreg)
> MK<-rq(Y~X1, tau= .5)
> summary(MK, se="iid")

Call: rq(formula = Y ~ X1, tau = 0.5)

tau: [1] 0.5

Coefficients:
             Value    Std. Error t value  Pr(>|t|)
(Intercept)  2.00000  0.95686    2.09018  0.05106
X1          -0.33333  0.27796   -1.19920  0.24600
> #reg mkuantil
> RMK<-rq.fit(X,Y, tau=.5)
> summary(RMK,se='iid')
              Length Class  Mode
coefficients   2     -none- numeric
x             40     -none- numeric
```

```
y               20     -none- numeric
residuals       20     -none- numeric
dual            20     -none- numeric
fitted.values 20       -none- numeric
> y_htRMK<-fitted.values(RMK)
> print(y_htRMK)
               [,1]
 [1,]  2.000000e+00
 [2,]  2.000000e+00
 [3,]  2.000000e+00
 [4,]  1.666667e+00
 [5,]  1.666667e+00
 [6,]  1.666667e+00
 [7,]  1.333333e+00
 [8,]  1.333333e+00
 [9,]  1.333333e+00
[10,]  1.000000e+00
[11,]  1.000000e+00
[12,]  1.000000e+00
[13,]  6.666667e-01
[14,]  6.666667e-01
[15,]  6.666667e-01
[16,]  3.333333e-01
[17,]  3.333333e-01
[18,]  3.333333e-01
[19,] -2.220446e-16
[20,] -2.220446e-16
> bMK<-as.matrix(coef(RMK))
> #KOEFISIEN
> eps<-abs(Y-X%*%bMK)
> #min.koef dgn Y
> #1-eps
> n1<-rep(1,n)
> min1<-n1-eps
> #1/e
> min2<-1/exp(X%*%bMK)
> min<-c(min1,min2)
> min(min)
[1] -4.333333
>
> #koefMK
> eps1<-abs(y_htRMK-X%*%bMK)
> #minkoefdgy
> #1-eps
> n2<-rep(1,n)
> min2<-n2-eps1
> #1/exp
```

```
> min3<-1/exp(X%*%bMK)
> min1<-c(min2,min2)
> minkoef<-min(min1)
> #teta dari rata-rata koef MK
> koef_MK<-c(minkoef,Y)
> teta_MK<-mean(koef_MK)
> #yhat MK
> y_htMK<-1/n*(Y+X%*%bMK*teta_MK)
> y_htMK
               [,1]
 [1,]  3.380952e-01
 [2,]  2.380952e-01
 [3,]  1.380952e-01
 [4,]  2.650794e-01
 [5,]  1.150794e-01
 [6,]  2.150794e-01
 [7,]  1.920635e-01
 [8,]  9.206349e-02
 [9,]  2.920635e-01
[10,]  6.904762e-02
[11,]  6.904762e-02
[12,]  6.904762e-02
[13,]  3.460317e-01
[14,]  9.603175e-02
[15,]  4.603175e-02
[16,]  1.230159e-01
[17,]  2.301587e-02
[18,]  1.230159e-01
[19,] -1.533165e-17
[20,] -1.533165e-17
> #predict MK
> ypred_hat<-1/n*(exp(X%*%bMK))
> ypred_hat
            [,1]
 [1,] 0.36945280
 [2,] 0.36945280
 [3,] 0.36945280
 [4,] 0.26472450
 [5,] 0.26472450
 [6,] 0.26472450
 [7,] 0.18968339
 [8,] 0.18968339
 [9,] 0.18968339
[10,] 0.13591409
[11,] 0.13591409
[12,] 0.13591409
[13,] 0.09738670
```

```
[14,] 0.09738670
[15,] 0.09738670
[16,] 0.06978062
[17,] 0.06978062
[18,] 0.06978062
[19,] 0.05000000
[20,] 0.05000000
> TETAmk<-ypred_hat+y_htMK
> TETAmk
            [,1]
 [1,] 0.70754804
 [2,] 0.60754804
 [3,] 0.50754804
 [4,] 0.52980387
 [5,] 0.37980387
 [6,] 0.47980387
 [7,] 0.38174689
 [8,] 0.28174689
 [9,] 0.48174689
[10,] 0.20496171
[11,] 0.20496171
[12,] 0.20496171
[13,] 0.44341845
[14,] 0.19341845
[15,] 0.14341845
[16,] 0.19279649
[17,] 0.09279649
[18,] 0.19279649
[19,] 0.05000000
[20,] 0.05000000
> #MSE MK
> eMK<-Y-TETAmk
> sseMK<-as.numeric(t(eMK)%*%eMK)
> mseMK<-sseMK/n
> mseMK
[1] 3.744423
>
> #GLM.NB
> NB<-glm.nb(Y~X1,data=tabel)
> summary(NB)

Call:
glm.nb(formula = Y ~ X1, data = tabel, init.theta =
0.8555284164,
    link = log)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-1.4524  -1.2617  -0.5791   0.4787   1.7478

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.7323     0.5271   1.389    0.165
X1           -0.1532     0.1622  -0.944    0.345

(Dispersion parameter for Negative Binomial(0.8555)
family taken to be 1)

    Null deviance: 21.603  on 19  degrees of freedom
Residual deviance: 20.747  on 18  degrees of freedom
AIC: 70.219

Number of Fisher Scoring iterations: 1


              Theta:  0.856
          Std. Err.:  0.597

 2 x log-likelihood:  -64.219
> y_htNB<-fitted.values(NB)
> print(y_htNB)
        1         2         3         4         5
6         7         8
2.0799012 2.0799012 2.0799012 1.7845525 1.7845525
1.7845525 1.5311437 1.5311437
        9        10        11        12        13
14        15        16
1.5311437 1.3137193 1.3137193 1.3137193 1.1271694
1.1271694 1.1271694 0.9671099
       17        18        19        20
0.9671099 0.9671099 0.8297790 0.8297790
> #teta DUGA NB
> psi_NB<-unlist(NB[24])
> psi_NB
    theta
0.8555284
> GAMht_i_NB<-y_htNB/(y_htNB+(exp(psi_NB)))
> direct<-GAMht_i_NB*Y
> direct
        1         2         3         4         5
6         7         8
1.8769476 0.9384738 0.0000000 1.2940386 0.0000000
0.8626924 0.7884850 0.0000000
```

```
        9        10        11        12        13
14        15        16
1.5769701 0.0000000 0.0000000 0.0000000 1.9435147
0.3239191 0.0000000 0.5826442
       17        18        19        20
0.0000000 0.5826442 0.0000000 0.0000000
> indirect<-(1-GAMht_i_NB)*y_htNB
> indirect
        1         2         3         4         5
6         7         8
1.1039348 1.1039348 1.1039348 1.0147925 1.0147925
1.0147925 0.9275017 0.9275017
        9        10        11        12        13
14        15        16
0.9275017 0.8429883 0.8429883 0.8429883 0.7620577
0.7620577 0.7620577 0.6853694
       17        18        19        20
0.6853694 0.6853694 0.6134222 0.6134222
> teta_EB<-direct+indirect
> print(teta_EB)
        1         2         3         4         5
6         7         8
2.9808824 2.0424086 1.1039348 2.3088312 1.0147925
1.8774849 1.7159868 0.9275017
        9        10        11        12        13
14        15        16
2.5044718 0.8429883 0.8429883 0.8429883 2.7055724
1.0859768 0.7620577 1.2680136
       17        18        19        20
0.6853694 1.2680136 0.6134222 0.6134222
> #mse NB
> eNB<-Y-teta_EB
> sse<-as.numeric(t(eNB)%*%eNB)
> mseNB<-sse/n
> mseNB
[1] 1.141315
>
```

Lampiran H. Perintah R untuk 35 area

```
> #SAE 35 unit level AREA
> rm(list=ls(all=TRUE))
> library(MASS)
> X0<-rep(1,35)
> X1<-
c(0,0,0,1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6,7,7,7,8,8,8
,9,9,9,10,10,10,11,11)
> X<-cbind(X0,X1)
> summary(X)
X0 X1
Min. :1 Min. : 0.000
1st Qu.:1 1st Qu.: 2.500
Median :1 Median : 5.000
Mean :1 Mean : 5.343
3rd Qu.:1 3rd Qu.: 8.000
Max. :1 Max. :11.000
> b0<-2
> b1<- .3
> b<-as.matrix(c(b0,b1))
> miu<-X%*%b
> print(miu)
[,1]
[1,] 2.0
[2,] 2.0
[3,] 2.0
[4,] 2.3
[5,] 2.3
[6,] 2.3
[7,] 2.6
[8,] 2.6
[9,] 2.6
[10,] 2.9
[11,] 2.9
[12,] 2.9
[13,] 3.2
[14,] 3.2
[15,] 3.2
[16,] 3.5
[17,] 3.5
[18,] 3.5
[19,] 3.8
[20,] 3.8
[21,] 3.8
[22,] 4.1
[23,] 4.1
```

```
[24,] 4.1
[25,] 4.4
[26,] 4.4
[27,] 4.4
[28,] 4.7
[29,] 4.7
[30,] 4.7
[31,] 5.0
[32,] 5.0
[33,] 5.0
[34,] 5.3
[35,] 5.3
> #membangkitkan Y direct
> library(gamlss)
> n<-length(X1)
> Y<-rZIP(n,miu,0.5)
> summary(Y)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0 0.0 0.0 1.8 4.0 8.0
> tabel<-data.frame(Y=Y,X0=X0,X1=X1)
> tabel
Y X0 X1
1 0 1 0
2 0 1 0
3 0 1 0
4 0 1 1
5 0 1 1
6 0 1 1
7 0 1 2
8 2 1 2
9 0 1 2
10 0 1 3
11 0 1 3
12 0 1 3
13 3 1 4
14 0 1 4
15 0 1 4
16 0 1 5
17 4 1 5
18 0 1 5
19 4 1 6
20 6 1 6
21 0 1 6
22 3 1 7
23 2 1 7
24 5 1 7
25 0 1 8
```

```
26 5 1 8
27 6 1 8
28 0 1 9
29 4 1 9
30 4 1 9
31 8 1 10
32 2 1 10
33 5 1 10
34 0 1 11
35 0 1 11
> # ESTIMASI M-KUANTIL
> library(quantreg)
> MK<-rq(Y~X1, tau= .5)
> summary(MK, se="iid")

Call: rq(formula = Y ~ X1, tau = 0.5)

tau: [1] 0.5

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept) -0.33333  0.77422   -0.43054  0.66960
X1           0.33333  0.12255    2.72006  0.01033
> #reg mkuantil
> RMK<-rq.fit(X,Y, tau=.5)
> summary(RMK,se='iid')
              Length Class  Mode
coefficients   2     -none- numeric
x             70     -none- numeric
y             35     -none- numeric
residuals     35     -none- numeric
dual          35     -none- numeric
fitted.values 35     -none- numeric
> y_htRMK<-fitted.values(RMK)
> print(y_htRMK)
            [,1]
 [1,] -0.3333333
 [2,] -0.3333333
 [3,] -0.3333333
 [4,]  0.0000000
 [5,]  0.0000000
 [6,]  0.0000000
 [7,]  0.3333333
 [8,]  0.3333333
 [9,]  0.3333333
[10,]  0.6666667
[11,]  0.6666667
```

```
[12,]  0.6666667
[13,]  1.0000000
[14,]  1.0000000
[15,]  1.0000000
[16,]  1.3333333
[17,]  1.3333333
[18,]  1.3333333
[19,]  1.6666667
[20,]  1.6666667
[21,]  1.6666667
[22,]  2.0000000
[23,]  2.0000000
[24,]  2.0000000
[25,]  2.3333333
[26,]  2.3333333
[27,]  2.3333333
[28,]  2.6666667
[29,]  2.6666667
[30,]  2.6666667
[31,]  3.0000000
[32,]  3.0000000
[33,]  3.0000000
[34,]  3.3333333
[35,]  3.3333333
> bMK<-as.matrix(coef(RMK))
> #KOEFISIEN
> eps<-abs(Y-X%*%bMK)
> #min.koef dgn Y
> #1-eps
> n1<-rep(1,n)
> min1<-n1-eps
> #1/e
> min2<-1/exp(X%*%bMK)
> min<-c(min1,min2)
> min(min)
[1] -4
>
> #koefMK
> eps1<-abs(y_htRMK-X%*%bMK)
> #minkoefdgy
> #1-eps
> n2<-rep(1,n)
> min2<-n2-eps1
> #1/exp
> min3<-1/exp(X%*%bMK)
> min1<-c(min2,min2)
> minkoef<-min(min1)
```

```
> #teta dari rata-rata koef MK
> koef_MK<-c(minkoef,Y)
> teta_MK<-mean(koef_MK)
> #yhat MK
> y_htMK<-1/n*(Y+X%*%bMK*teta_MK)
> y_htMK
             [,1]
 [1,] -0.01693122
 [2,] -0.01693122
 [3,] -0.01693122
 [4,]  0.00000000
 [5,]  0.00000000
 [6,]  0.00000000
 [7,]  0.01693122
 [8,]  0.07407407
 [9,]  0.01693122
[10,]  0.03386243
[11,]  0.03386243
[12,]  0.03386243
[13,]  0.13650794
[14,]  0.05079365
[15,]  0.05079365
[16,]  0.06772487
[17,]  0.18201058
[18,]  0.06772487
[19,]  0.19894180
[20,]  0.25608466
[21,]  0.08465608
[22,]  0.18730159
[23,]  0.15873016
[24,]  0.24444444
[25,]  0.11851852
[26,]  0.26137566
[27,]  0.28994709
[28,]  0.13544974
[29,]  0.24973545
[30,]  0.24973545
[31,]  0.38095238
[32,]  0.20952381
[33,]  0.29523810
[34,]  0.16931217
[35,]  0.16931217
> #predict MK
> ypred_hat<-1/n*(exp(X%*%bMK))
> ypred_hat
            [,1]
 [1,] 0.02047232
```

```
 [2,] 0.02047232
 [3,] 0.02047232
 [4,] 0.02857143
 [5,] 0.02857143
 [6,] 0.02857143
 [7,] 0.03987464
 [8,] 0.03987464
 [9,] 0.03987464
[10,] 0.05564954
[11,] 0.05564954
[12,] 0.05564954
[13,] 0.07766520
[14,] 0.07766520
[15,] 0.07766520
[16,] 0.10839051
[17,] 0.10839051
[18,] 0.10839051
[19,] 0.15127114
[20,] 0.15127114
[21,] 0.15127114
[22,] 0.21111589
[23,] 0.21111589
[24,] 0.21111589
[25,] 0.29463596
[26,] 0.29463596
[27,] 0.29463596
[28,] 0.41119760
[29,] 0.41119760
[30,] 0.41119760
[31,] 0.57387248
[32,] 0.57387248
[33,] 0.57387248
[34,] 0.80090357
[35,] 0.80090357
> TETAmk<-ypred_hat+y_htMK
> TETAmk
             [,1]
 [1,] 0.003541106
 [2,] 0.003541106
 [3,] 0.003541106
 [4,] 0.028571429
 [5,] 0.028571429
 [6,] 0.028571429
 [7,] 0.056805858
 [8,] 0.113948715
 [9,] 0.056805858
[10,] 0.089511978
```

```
[11,] 0.089511978
[12,] 0.089511978
[13,] 0.214173132
[14,] 0.128458846
[15,] 0.128458846
[16,] 0.176115379
[17,] 0.290401093
[18,] 0.176115379
[19,] 0.350212943
[20,] 0.407355800
[21,] 0.235927229
[22,] 0.398417476
[23,] 0.369846047
[24,] 0.455560333
[25,] 0.413154476
[26,] 0.556011619
[27,] 0.584583047
[28,] 0.546647338
[29,] 0.660933052
[30,] 0.660933052
[31,] 0.954824864
[32,] 0.783396293
[33,] 0.869110579
[34,] 0.970215738
[35,] 0.970215738
> #MSE MK
> eMK<-Y-TETAmk
> sseMK<-as.numeric(t(eMK)%*%eMK)
> mseMK<-sseMK/n
> mseMK
[1] 6.909844
>
> #GLM.NB
> NB<-glm.nb(Y~X1,data=tabel)
> summary(NB)

Call:
glm.nb(formula = Y ~ X1, data = tabel, init.theta =
0.6552237653,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8152  -0.9691  -0.6380   0.2989   1.2634

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) -1.49102    0.61531  -2.423 0.015383 *
X1           0.31803    0.08525   3.730 0.000191 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6552)
family taken to be 1)

    Null deviance: 43.421  on 34  degrees of freedom
Residual deviance: 31.946  on 33  degrees of freedom
AIC: 118.72

Number of Fisher Scoring iterations: 1


              Theta:  0.655
          Std. Err.:  0.308

 2 x log-likelihood:  -112.722
> y_htNB<-fitted.values(NB)
> print(y_htNB)
        1         2         3         4         5
6         7         8
0.2251418 0.2251418 0.2251418 0.3094381 0.3094381
0.3094381 0.4252960 0.4252960
        9        10        11        12        13
14        15        16
0.4252960 0.5845327 0.5845327 0.5845327 0.8033899
0.8033899 0.8033899 1.1041903
       17        18        19        20        21
22        23        24
1.1041903 1.1041903 1.5176145 1.5176145 1.5176145
2.0858306 2.0858306 2.0858306
       25        26        27        28        29
30        31        32
2.8667946 2.8667946 2.8667946 3.9401623 3.9401623
3.9401623 5.4154140 5.4154140
       33        34        35
5.4154140 7.4430204 7.4430204
> #teta DUGA NB
> psi_NB<-unlist(NB[24])
> psi_NB
    theta
0.6552238
> GAMht_i_NB<-y_htNB/(y_htNB+(exp(psi_NB)))
> direct<-GAMht_i_NB*Y
```

```
> direct
        1         2         3         4         5
6         7         8
0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
0.0000000 0.0000000 0.3618202
        9        10        11        12        13
14        15        16
0.0000000 0.0000000 0.0000000 0.0000000 0.8831815
0.0000000 0.0000000 0.0000000
       17        18        19        20        21
22        23        24
1.4577907 0.0000000 1.7630342 2.6445514 0.0000000
1.5599256 1.0399504 2.5998760
       25        26        27        28        29
30        31        32
0.0000000 2.9910001 3.5892001 0.0000000 2.6869007
2.6869007 5.9015647 1.4753912
       33        34        35
3.6884780 0.0000000 0.0000000
> indirect<-(1-GAMht_i_NB)*y_htNB
> indirect
        1         2         3         4         5
6         7         8
0.2015735 0.2015735 0.2015735 0.2665963 0.2665963
0.2665963 0.3483557 0.3483557
        9        10        11        12        13
14        15        16
0.3483557 0.4484116 0.4484116 0.4484116 0.5668769
0.5668769 0.5668769 0.7017707
       17        18        19        20        21
22        23        24
0.7017707 0.7017707 0.8487129 0.8487129 0.8487129
1.0012504 1.0012504 1.0012504
       25        26        27        28        29
30        31        32
1.1518780 1.1518780 1.1518780 1.2934561 1.2934561
1.2934561 1.4204870 1.4204870
       33        34        35
1.4204870 1.5298007 1.5298007
> teta_EB<-direct+indirect
> print(teta_EB)
        1         2         3         4         5
6         7         8
0.2015735 0.2015735 0.2015735 0.2665963 0.2665963
0.2665963 0.3483557 0.7101759
        9        10        11        12        13
14        15        16
```

```
0.3483557 0.4484116 0.4484116 0.4484116 1.4500584
0.5668769 0.5668769 0.7017707
       17        18        19        20        21
22        23        24
2.1595614 0.7017707 2.6117472 3.4932643 0.8487129
2.5611760 2.0412008 3.6011264
       25        26        27        28        29
30        31        32
1.1518780 4.1428781 4.7410781 1.2934561 3.9803568
3.9803568 7.3220517 2.8958781
       33        34        35
5.1089649 1.5298007 1.5298007
> #mse NB
> eNB<-Y-teta_EB
> sse<-as.numeric(t(eNB)%*%eNB)
> mseNB<-sse/n
> mseNB
[1] 0.9319745
>
```

Lampiran I. Perintah R untuk 50 area

```
> #SAE 50 unit level AREA
> rm(list=ls(all=TRUE))
> library(MASS)
> X0<-rep(1,50)
> X1<-
c(0,0,0,1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6,7,7,7,8,8,8
,9,9,9,10,10,10,11,11,11,12,12,12
,13,13,13,14,14,14,15,15,15,16,16)
> X<-cbind(X0,X1)
> summary(X)
X0 X1
Min. :1 Min. : 0.00
1st Qu.:1 1st Qu.: 4.00
Median :1 Median : 8.00
Mean :1 Mean : 7.84
3rd Qu.:1 3rd Qu.:12.00
Max. :1 Max. :16.00
> b0<-2
> b1<- .3
> b<-as.matrix(c(b0,b1))
> miu<-X%*%b
> print(miu)
[,1]
[1,] 2.0
[2,] 2.0
[3,] 2.0
[4,] 2.3
[5,] 2.3
[6,] 2.3
[7,] 2.6
[8,] 2.6
[9,] 2.6
[10,] 2.9
[11,] 2.9
[12,] 2.9
[13,] 3.2
[14,] 3.2
[15,] 3.2
[16,] 3.5
[17,] 3.5
[18,] 3.5
[19,] 3.8
[20,] 3.8
[21,] 3.8
[22,] 4.1
```

```
[23,] 4.1
[24,] 4.1
[25,] 4.4
[26,] 4.4
[27,] 4.4
[28,] 4.7
[29,] 4.7
[30,] 4.7
[31,] 5.0
[32,] 5.0
[33,] 5.0
[34,] 5.3
[35,] 5.3
[36,] 5.3
[37,] 5.6
[38,] 5.6
[39,] 5.6
[40,] 5.9
[41,] 5.9
[42,] 5.9
[43,] 6.2
[44,] 6.2
[45,] 6.2
[46,] 6.5
[47,] 6.5
[48,] 6.5
[49,] 6.8
[50,] 6.8
> #membangkitkan Y direct
> library(gamlss)
> n<-length(X1)
> Y<-rZIP(n,miu,0.5)
> Y
[1] 0 0 3 5 0 2 1 0 0 0 0 0 0 1 0 7 0 0 0 0 5 0 9 3 7 0
1 0 3 4 0 1 5 3 0 0 0 5 3 5 5 1 5 3 0
0 0 7 6 0
> summary(Y)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.00 0.00 0.50 2.00 3.75 9.00
> tabel<-data.frame(Y=Y,X0=X0,X1=X1)
> tabel
Y X0 X1
1 0 1 0
2 0 1 0
3 3 1 0
4 5 1 1
5 0 1 1
```

```
6 2 1 1
7 1 1 2
8 0 1 2
9 0 1 2
10 0 1 3
11 0 1 3
12 0 1 3
13 0 1 4
14 1 1 4
15 0 1 4
16 7 1 5
17 0 1 5
18 0 1 5
19 0 1 6
20 0 1 6
21 5 1 6
22 0 1 7
23 9 1 7
24 3 1 7
25 7 1 8
26 0 1 8
27 1 1 8
28 0 1 9
29 3 1 9
30 4 1 9
31 0 1 10
32 1 1 10
33 5 1 10
34 3 1 11
35 0 1 11
36 0 1 11
37 0 1 12
38 5 1 12
39 3 1 12
40 5 1 13
41 5 1 13
42 1 1 13
43 5 1 14
44 3 1 14
45 0 1 14
46 0 1 15
47 0 1 15
48 7 1 15
49 6 1 16
50 0 1 16
>
#SAE 20 unit level AREA
```

```
rm(list=ls(all=TRUE))
library(MASS)
X0<-rep(1,50)
X1<-
c(0,0,0,1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6,7,7,7,8,8,8
,9,9,9,10,10,10,11,11,11,12,12,12,13,13,13,14,14,14,15,
15,15,16,16)
X<-cbind(X0,X1)
summary(X)
b0<-2
b1<- .3
b<-as.matrix(c(b0,b1))
miu<-X%*%b
print(miu)
#membangkitkan Y direct
library(gamlss)
n<-length(X1)
Y<-
c(0,0,3,5,0,2,1,0,0,0,0,0,0,1,0,7,0,0,0,0,5,0,9,3,7,0,1
,0,3,4,0,1,5,3,0,0,0,5,3,5,5,1,5,3,0,0,0,7,6,0)
summary(Y)
tabel<-data.frame(Y=Y,X0=X0,X1=X1)
tabel

# ESTIMASI M-KUANTIL
library(quantreg)
MK<-rq(Y~X1, tau= .5)
summary(MK, se="iid")
#reg mkuantil
RMK<-rq.fit(X,Y, tau=.5)
summary(RMK,se='iid')
y_htRMK<-fitted.values(RMK)
print(y_htRMK)
bMK<-as.matrix(coef(RMK))
#KOEFISIEN
eps<-abs(Y-X%*%bMK)
#min.koef dgn Y
#1-eps
n1<-rep(1,n)
min1<-n1-eps
#1/e
min2<-1/exp(X%*%bMK)
min<-c(min1,min2)
min(min)

#koefMK
eps1<-abs(y_htRMK-X%*%bMK)
```

```
#minkoefdgy
#1-eps
n2<-rep(1,n)
min2<-n2-eps1
#1/exp
min3<-1/exp(X%*%bMK)
min1<-c(min2,min2)
minkoef<-min(min1)
#teta dari rata-rata koef MK
koef_MK<-c(minkoef,Y)
teta_MK<-mean(koef_MK)
#yhat MK
y_htMK<-1/n*(Y+X%*%bMK*teta_MK)
y_htMK
#predict MK
ypred_hat<-1/n*(exp(X%*%bMK))
ypred_hat
TETAmk<-ypred_hat+y_htMK
TETAmk
#MSE MK
eMK<-Y-TETAmk
sseMK<-as.numeric(t(eMK)%*%eMK)
mseMK<-sseMK/n
mseMK

#GLM.NB
NB<-glm.nb(Y~X1,data=tabel)
summary(NB)
y_htNB<-fitted.values(NB)
print(y_htNB)
#teta DUGA NB
psi_NB<-unlist(NB[24])
psi_NB
GAMht_i_NB<-y_htNB/(y_htNB+(exp(psi_NB)))
direct<-GAMht_i_NB*Y
direct
indirect<-(1-GAMht_i_NB)*y_htNB
indirect
teta_EB<-direct+indirect
print(teta_EB)
#mse NB
eNB<-Y-teta_EB
sse<-as.numeric(t(eNB)%*%eNB)
mseNB<-sse/n
mseNB
```