PEMODELAN *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENANGANI DATA *EXCESS ZEROES* PADA *SMALL AREA ESTIMATION*

SKRIPSI

Oleh

Abdulloh Hadi
NIM. 131810101013

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2020

PEMODELAN *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENANGANI DATA *EXCESS ZEROES* PADA *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

Abdulloh Hadi
NIM. 131810101013

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2020

PERSEMBAHAN

Bismillahirrohmanirrohim

Dengan rahmat Allah SWT yang Maha Pengasih lagi Maha Penyayang, sholawat serta salam terhadap baginda Rasulullah Muhammad SAW. Saya persembahkan skripsi ini sebagai wujud syukur dan rasa terima kasih untuk:

1. Orang tua saya, Bapak Abdul Hadi dan Ibu Khotim, yang selalu memberikan segala bentuk dukungan, motivasi, dan do'anya untuk kemudahan, kelancaran, dan kesuksesan saya.
2. Seluruh Keluarga besar yang juga memberikan segenap dukungan dan doanya untuk saya.
3. Teman-teman ATLAS 13 dan Smokers yang sudah berjuang bersama sampai akhir.
4. Almamater tercinta Jurusan Matematika FMIPA Universitas Jember
5. Serta, beberapa pihak yang tidak bisa saya sebutkan satu persatu, yang telah memberikan dukungannya.

MOTTO

“ *Jagalah shalatmu, ketika engkau kehilangannnnya maka engkau akan kehilangan yang lainnya* “

[1] Umar bin khattab

PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama: Abdulloh Hadi

NIM : 131810101013

menyatakan dengan sesungguhnya bahwa skripsi yang berjudul "Pemodelan *Zero Inflated Negative Binomial* (ZINB) untuk menangani data *excess zeroes* pada *small area estimation"* adalah benar-benar hasil karya sendiri, kecuali kutipan yang telah disebutkan sumbernya, belum pernah diajukan di institusi manapun, dan bukan karya jiplakan. Saya bertanggung jawab 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 jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Januari 2020

Yang menyatakan,

Abdulloh Hadi

NIM 131810101013

SKRIPSI

PEMODELAN *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENANGANI DATA *EXCESS ZEROES* PADA *SMALL AREA ESTIMATION*

Oleh

Abdulloh Hadi

NIM 131810101013

Pembimbing:

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

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

PENGESAHAN

Skripsi berjudul "Pemodelan *Zero Inflated Negative Binomial* (ZINB) untuk menangani data *excess zeroes* pada *small area estimation*" telah diuji dan disahkan pada:

hari, tanggal :

tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Tim Penguji:

Ketua,	Anggota I,
Prof. Drs. I Made Tirta M. Sc., Ph. D.	Dr. Mohamat Fatekurohman, S.Si., M.Si.
NIP. 195912201985031002	NIP. 196906061998031001
Anggota II,	Anggota III,
Dr. Alfian Futuhul Hadi, S.Si., M.Si.	Dian Anggraeni, S.Si., M.Si.
NIP. 197407192000121001	NIP. 198202162006042002

Mengesahkan

Dekan,

Drs. Achmad Sjaifullah, M. Sc., Ph.D

NIP. 195910091986021001

RINGKASAN

PEMODELAN *ZERO INFLATED NEGATIVE BINOMIAL* (ZINB) UNTUK MENANGANI DATA *EXCESS ZEROES* PADA *SMALL AREA ESTIMATION* ; Abdulloh Hadi, 131810101013, 2020: 40 Halaman: Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Small Area Estimation (SAE) merupakan suatu teknik statistika yang digunakan untuk menduga parameter-parameter subpopulasi dengan ukuran sampel kecil. Penelitian tentang SAE telah dilakukan oleh beberapa orang pada kasus overdispersi.Salah satu yang menyebabkan overdispersi adalah terdapat nilai nol yang lebih pada data pengamatan (*excess zeroes*).

Overdispersi pada regresi Poisson terjadi ketika varians dari respon lebih besar dari meannya, Salah satu yang menyebabkan overdispersi adalah terdapat nilai nol yang lebih besar pada data pengamatan (*excess zeroes*). Zulfi(2016) telah meneliti bahwa estimasi dengan model Binomial Negatif lebih baik dibandingkan dengan model M-Kuantil pada data *excess zeroes* dalam konteks SAE. Namun secara umum nilai MSE estimasi model Binomial Negatif masih tergolong besar dibandingkan dengan beberapa peneliti sebelumnya. Hal itu dikarenakan data yang digunakan merupakan data yang mengalami *excess zeroes*. Sehingga pada penelitian ini dilakukan penelitian tenntang overdispersi dengan data yang berdistribusi *Zero-Inflated Poisson* menggunakan model *Zero inflated Negative Binomial* (ZINB) pada konteks *Small Area Estimation*.

Data yang digunakan dalam penelitian ini merupakan data simulasi dengan cara membangkitkan data berdistribusi *Zero-Inflated poisson* pada *Small Area*

Estimation (SAE). Area kecil yang digunakan dalam data berikut adalah 1 sampai 50 area dengan 60%, 70%, 80%, 90%, bernilai nol.

Nilai MSE pada penelitian 50 area dengan 60% nilai nol model ZINB adalah 1.076108, sedangkan untuk model NB adalah 1.076948 menunjukkan bahwa model ZINB memiliki nilai eror lebih kecil daripada nilai model NB. Namun nilai MSE pada penelitian 70%, 80%,dan 90% nilai nol dengan model ZINB berturut-turut memiliki nilai eror lebih besar dibandingkan dengan model NB. Nilai MSE yang telah dihitung pada tabel 4.6 menunjukkan bahwa estimasi model NB lebih baik dari pada estimasi model ZINB.

PRAKATA

Puji syukur kehadirat Allah SWT atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Pemodelan *Zero Inflated Negative Binomial* (ZINB) untuk menangani data *excess zeroes* pada *small area estimation*". Skripsi ini disusun untuk memenuhi salah satu syarat pada program pendidikan strata satu (S1) Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Skripsi ini disusun melalui beberapa tahap, baik dalam bentuk seminar maupun bimbingan intensif. Skripsi ini tidak akan terselesaikan tanpa adanya bantuan dari beberapa pihak, serta kerja keras dan keuletan dari diri pribadi. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terima kasih atas bantuan dan bimbingan dalam penyusunan sripsi ini, terutama kepada yang terhormat:

1. Drs. Achmad Sjaifullah, M. Sc., Ph.D, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember ;
2. Kusbudiono, S.Si., M.Si., selaku Ketua Jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Jember,
3. Dr. Alfian Futuhul Hadi, S.Si., M.Si., selaku Dosen Pembimbin Utama, Dian Anggraeni, S.Si., M.Si., selaku Dosen Pembimbing Angggota, Prof. Drs. I Made Tirta selaku Dosen Penguji I, Dr. Mohamat Fatekurohman, S.Si., M.Si., selaku Dosen Penguji II yang juga telah meluangkan waktu dan pikiran dalam membimbing penulisan skripsi ini sampai terselesaikan;
4. Dosen dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember yang telah menyalurkan ilmunya;

5. Kedua orang tua dan keluarga besar yang selalu mendukung dan memberikan seluruh bantuannya dari awal sampai terselesaikannya skripsi ini;

6. Teman-teman dan semua pihak yang juga telah membantu terselesaikannya skripsi ini.

Penulis menyadari masih banyak kekurangan dalam penulisan skripsi ini. Namun, suatu usaha tidak akan berakhir dan berhasil jika tidak dimulai. Oleh karena itu, penulis mengharapkan kritik dan sarannya demi penyempurnaan skripsi ini. Penulis berharap semoga skripsi ini bermanfaat bagi para pembaca

Jember, Januari 2020

Penulis

DAFTAR ISI

Halaman

DAFTAR GAMBAR

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

Halaman

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Small Area Estimation (SAE) merupakan suatu teknik statistika yang digunakan untuk menduga parameter-parameter subpopulasi dengan ukuran sampel kecil. Teknik pendugaan ini memanfaatkan data dari domain besar untuk menduga parameter pada domain yang lebih kecil yang dapat berupa desa/kelurahan, kecamatan, kabupaten, kelompok suku, maupun kelompok umur. Metode SAE mempunyai konsep dalam pendugaan parameter secara tidak langsung di suatu area yang relatif kecil dalam percontohan survey (*survey sampling*) dimana pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam *small area* berukuran kecil/sedikit, sehingga yang dihasilkan akan memiliki varians yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei (Prasad dan Rao, 1990).

Pendugaan dalam SAE didasarkan pada model area kecil yang membutuhkan informasi tambahan yang memiliki hubungan dengan peubah yang sedang diamati yang disebut juga sebagai variabel penyerta *(auxiliary variabel*). Variabel penyerta ini dapat diperoleh dari survei yang lain dan diharapkan memiliki korelasi dengan peubah yang diamati (Rumiati, 2012).

Sebagai alternatif teknik pendugaan untuk meningkatkan efektivitas ukuran sampel dan menurunkan eror, dikembangkan teknik pendugaan tak langsung (*indirect estimation*) untuk melakukan pendugaan pada area kecil dengan ketelitian yang cukup. Teknik pendugaan ini dilakukan melalui suatu model yang menghubungkan area terkait melalui penggunaan informasi tambahan atau variabel penyerta (*model-based*). Secara statistik metode dengan memanfaatkan informasi tambahan akan mempunyai sifat “meminjam kekuatan” (*borrowing strength*) dari hubungan antara rataan area kecil dan informasi tambahan tersebut, Jika tidak ada hubungan linier antara rataan area kecil dan variabel penyerta maka tidak tepat ‘meminjam kekuatan’ dari area lain dengan menggunakan model linier dalam pendugaan tak langsung. Untuk mengatasi hal tersebut dikembangkan pendekatan nonparametrik. Salah satu pendekatan nonparametrik yang digunakan adalah pendekatan *Kernel-Based* (Mukhopadhay dan Maiti, 2004).

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. Penanganan Overdispersi juga dilakukan oleh Oktarin (2015) menggunakan regresi M-Kuantil dengan *resampling bosstrap* menyimpulkan bahwa model Regresi M-kuantil menghasilkan nilai pendugaan lebih baik dari pada model Poisson pada SAE.

Overdispersi pada regresi Poisson terjadi ketika varians dari respon lebih besar dari meannya, Salah satu yang menyebabkan overdispersi adalah terdapat nilai nol yang lebih besar pada data pengamatan (*excess zeroes*). Zulfi (2016) telah meneliti bahwa estimasi dengan model Binomial Negatif lebih baik dibandingkan dengan model M-Kuantil pada data *excess zeroes* dalam konteks SAE. Namun secara umum nilai MSE estimasi model Binomial Negatif masih tergolong besar dibandingkan dengan beberapa peneliti sebelumnya. Hal itu dikarenakan data yang digunakan merupakan data yang mengalami *excess zeroes*.

Berdasarkan dari penjelasan tersebut peneliti tertarik untuk melakukan penelitian tentang overdispersi dengan data simulasi pada data *excess zeroes* menggunakan model *Zero inflated Negative Binomial* (ZINB) pada konteks *Small Area Estimation*.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut dapat dirumuskan permasalahan sebagai berikut:

a. Bagaimana model *Zero Inflated Negative Binomial* pada konteks *small area estimation*?
b. Bagaimana performa model *Zero-Inflated Negative Binomial* pada data cacah dengan konteks *small area estimation*?

1.3 Batasan Masalah

Model yang digunakan dalam mengatasi overdispersi karena mengalami excess zeroes adalah *Zero Inflated Negative Binomial* dengan menggunakan data berdistribusi *Zero-Inflated Poisson*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah :

a. Memperoleh model *Zero Inflated Negative Binomial* pada *Small Area Estimation.*
b. Mengetahui perfoma model *Zero Inflated Negative Binomial* pada *Small Area Estimation.*

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan pada penelitian ini adalah sebagai berikut:

a. Memberikan wawasan ilmiah berkaitan dengan *Small Area Estimation*.
b. Penelitian ini diharapkan dapat digunakan sebagai informasi tambahan bagi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam khususnya, jurusan Matematika.

BAB 2. TINJAUAN PUSTAKA

2.1. *Small Area Estimation* (SAE)

Area kecil atau *small area* diartikan sebagai bagian dari populasi, baik berdasarkan area geografi maupun sosial-demografi. Suatu daerah disebut area kecil jika di dalam daerah tersebut, contoh yang terambil kurang banyak untuk mendapatkan nilai penduga langsung dengan presisi yang memadai. Pendugaan area kecil merupakan pendugaan parameter suatu area yang lebih kecil dengan memanfaatkan informasi dari luar area, dari dalam area itu sendiri, dan dari luar survei (Rao , 2003).

Pendugaan parameter pada suatu area kecil dapat dilakukan dengan pendugaan langsung maupun pendugaan tidak langsung. Pendugaan langsung merupakan pendugaan berdasarkan data sampel atau design sampling, dimana hasil pendugaan langsung pada suatu area kecil merupakan pendugaan tak bias meskipun memiliki varian yang besar dikarenakan dugaannya diperoleh dari ukuran sampel kecil. Pendugaan tak langsung merupakan pendugaan dengan cara memanfaatkan informasi variabel lain yang berhubungan dengan parameter yang diamati, dimana variabel penyerta terpilih diharapkan bisa memberikan pengaruh pada parameter yang diamati. Model area kecil biasanya menggunakan model linier campuran dalam bentuk $y = X\beta + Zu + e$, dimana X adalah matriks peubah penyerta, Z adalah vektor acak yang biasa dikenal sebagai pengaruh area kecil, dan e adalah vektor dari galat sampel. Dalam kebanyakan aplikasi pendugaan area kecil, digunakan asumsi model linier campuran dan pendugaannya sensitive terhadap asumsi ini. Jika asumsi kelinieran antara rataan area kecil dan variabel penyerta tidak terpenuhi, maka "meminjam kekuatan" dari area lain dengan menggunakan model linier tidak tepat, mukhopaday dan Maiti(2004) menggunakan model,

$$y_i = \theta_i + \epsilon_i$$
$$\theta_i = m(x_i) + u_i$$

dimana i = 1,2…,m menyatakan banyak area kecil. Fungsi m(.) adalah fungsi mulus (*smoothing function*) yang mendefinisikan relasi antara variabel x dan y. θ_i adalah rataan area kecil yang tidak teramati, y_i adalah penduga langsung dari rataan area kecil, u_i galat peubah acak yang berdistibusi independen dan identik dengan E(u_i) = 0 dan var (u_i) = σ_u^2 dan ϵ_i berdistribusi inependen dan identik dengan E(ϵ_i) = 0 dan var(ϵ_i) =D_i.

Small Area Etimation (SAE) memiliki dua jenis model pada pendugaan area kecil adalah *basic area level model* dan *basic unit level model* (Rao , 2003).

2.1.1. *Basic area level model*

mengasumsikan bahwa peubah penjelas yang tersedia hanya ada untuk level area tertentu. Misalkan tersedia vektor peubah penjelas $(x_i = x_{1i}, x_{2i}, \ldots x_{pi})^T$dan parameter θ_i yang akan diduga diasumsikan memiliki hubungan dengan x_i. Peubah penjelas tersebut dimodelkan dengan data pendukung x_i. Data pendukung tersebut digunakan untuk membangun model $\theta_i = x_i^T\beta + b_i v_i$ dengan *i*=1,...,m dan $v_i \sim$ N$(0\,,\sigma_v^2)$, sebagai pengaruh acak yang diasumsikan menyebar normal. Kesimpulan mengenai θ_i, dapat diketahui dengan mengasumsikan bahwa model penduga langsung y_i telah tersedia, yaitu: $y_i = \theta_i + e_i$, dengan *i*=1,...,m menyatakan banyak area kecil dan *sampling error* e_i~N$(0, \sigma_{ei}^2)$, dengan σ_{ei}^2 diketahui.

Kemudian kedua model tersebut digabung sehingga didapatkan model campuran (*mixed model*) $y_i = x_i^T\beta + b_i v_i + e_i$, dengan *i*=1,...,m dan b_i diketahui bernilai positif konstan. Model tersebut merupakan bentuk khusus dari model linier campuran (*general linear mixed model*) yang terdiri dari pengaruh tetap (*fixed effect*) yaitu β dan pengaruh acak (*random effect*) yaitu v_i.

2.1.2. *Basic unit model*

Merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$,dimana x_{ij} merupakan data penyerta untuk unit tertentu telah tersedia untuk masing-masing elemen populasi j di setiap small area. Selanjutnya variabel y_{ij}, diasumsikan berhubungan dengan x_{ij} melalui model

$$y_{ij} = x_{ij}^T\beta + b_i v_i + e_{ij}$$

dengan i=1,...,m; j=1,...N_i, $v_i \sim \mathrm{N}(0\,,\sigma_v^2)$ dan $e_i{\sim}\mathrm{N}(0,\sigma_{ei}^2)$. Dimana $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ yang merupakan data penyerta unit tertentu, p adalah variabel prediktor, dan v = pengaruh acak area yang disumsikan merupakan variabel yang bersifat *iid*.

2.2. Regresi Binomial Negatif (NB)

Percobaan negative binomial terdiri dari atas beberapa usaha dan tiap usaha dengan dua kemungkinan hasil yang dapat diberi nama sukses atau gagal dan dilakukan sampai tercapai jumlah sukses tertentu. Model Regresi *Binomial Negative* memiliki kegunaan yang sama dengan Regresi Poisson yaitu untuk menganalisis hubungan antara satu variabel respon *data count* dengan satu atau lebih variabel acak penjelas, tetapi model Regresi *Binomial Negative* lebih fleksibel dibandingkan dengan model poisson karena mean dan variansi dari model *Binomial Negative* tidak harus sama. Model regresi Negatif Binomial yang dibangun memiliki sebaran Negatif Binomial dengan parameter μ dan α. Sehingga sebaran Y_i menjadi:

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

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

$$E(Y_i) = \mu_i$$

$$Var(Y_i) = \mu_i + \alpha{\mu_i}^2$$

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

$$\ln(\mu_i) = x_i'\beta \qquad (2.1)$$

$$\ln\{E[Y_i|x_i]\} = \ln(\mu_i) = x_i{}'\beta \text{ untuk i=1,2,..,n} \qquad (2.2)$$

$$\mu_i = \exp(x_i{}'\beta) \qquad (2.3)$$

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

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

2.3. *Overdispersi*

menurut Hilbe(2011) menjelaskan bahwa overdispersi pada regresi poisson terjadi ketika varians dari respon lebih besar dari meannya. berikut adalah beberapa hal yang dapat menyebabkan overdispersi, antara lain:

a. terdapat korelasi antar pengamatan.
b. terdapat pelanggaran assumsi distribusi poisson, yaitu Var(y) > E(y).
c. terdapat *excess zeros*.
d. terdapat outlier dalam data,

Overdispersi menyebabkan nilai devian model menjadi sangat besar dan menyebabkan 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.

2.4. *Zero Inflated Models*

Model *Zero Inflated* mempertimbangkan dua sumber hasil nol yang berbeda. Satu sumber dihasilkan dari individu yang masuk dalam proses perhitungan, yang lain dari mereka yang masuk dalam proses perhitungan tapi menghasilkan nol (Hardin & Hilbe, 2007)

Untuk model nol-meningkat, probabilitas mengamati hasil nol sama dengan probabilitas bahwa individu berada dalam kelompok yang selalu nol ditambah probabilitasnya Individu bukan kelompok itu yang menduga probabilitas proses penghitungan menghasilkan sebuah nol. Jika B(0) sebagai probabilitas bahwa hasil biner menghasilkan hasil nol dan jika Pr(0) sebagai probabilitas bahwa penghitungan hasil nol, probabilitas nol Hasil untuk sistem ini kemudian diberikan oleh (Hardin & Hilbe, 2007):

$$\Pr(y = 0) = B(0) + (1 - Z) \Pr(0)$$

Peluang dari perhitungan tidak nol adalah:

$$\Pr(y = k; k > 0) = [1 - B(0)]\Pr(k)$$

2.5. *Excess Zeroes*

*Excess zero*es (kelebihan nol) merupakan salah satu permasalahan yang dialami pada regresi poisson. Variabel respon pada data diskrit mungkin ditemukan data kosong atau bernilai nol. Akan tetapi dalam banyak kasus , data kosong memiliki arti penting pada penelitian yang bersangkutan. Jika data kosong memilki arti penting dalam data diskrit maka data tersebut harus dimasukkan dalam analisis. *Excess zeroes* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskritnya. *Excess zeros* merupakan salah satu penyebab overdispersi (Ariawan, 2012).

2.6. *Zero Inflated Negative Binomial(ZINB)*

Zero inflated Negative Binomial merupakan model yang dibentuk dari distribusi Negative Binomial.

Ariawan (2012) mendefinisikan model regresi *Zero Inflated Negative Binomial (ZINB)* adalah:

$$P(Y_i = y_i|x_i) = \begin{cases} \varphi(\emptyset' x_i) + \{1-\varphi(\emptyset' x_i)\}g(0|x_i) \\ \{1-\varphi(\emptyset' x_i)\}g(y_i|x_i) \end{cases}$$

dimana φ adalah sebuah fungsi dari $z'_i\emptyset; x_i$ yang merupakan sebuah vector kovariat zero inflated dan $\emptyset$ adalah sebuah vektor koefisien *zero inflated* yang akan dijadikan nilai duga. $g(y_i|x_i)$ adalah peluang distribusi *Negative Binomial* yaitu:

$$g(y_i|x_i) = \frac{\Gamma(y_i+\alpha)}{\Gamma(\mu_i)\Gamma(y_i+1)}\left(\frac{\alpha}{\mu_i+\alpha}\right)^{\alpha}\left(\frac{\mu_i}{\mu_i+\alpha}\right)^{y_i}$$

Mean dan varian didefinisikan sebagai berikut:

$$E(y_i|x_i) = \mu_i(1-\varphi_i)$$

dan

$$Var(y_i|x_i) = \mu_i(1-\varphi_i)(1+\mu_i(\varphi_i+\alpha))$$

2.6.1. Regresi *Zero Inflated Negative Binomial (ZINB)*

Model *Zero Inflated Negative Binomial* dapat digunakan untuk mengatasi masalah data yang mengalami overdispersi dan *excess zeroes*. Model Regresi *Zero Inflated Negative Binomial* sama dengan Regresi *Negative Binomial* yaitu untuk menganalisis hubungan antara suatu vaiabel respon data cacah dengan satu atau lebih variabel acak penjelas. Model *Zero Inflated Negative Binomial* dinotasikan dengan $Y_i \sim ZINB(\mu_i, \alpha, \varphi)$ dan untuk memodelkan φ umumya di gunakan model logit, yaitu

$$\varphi = \frac{\exp(\mathrm{Z_i^T}\lambda)}{1+\exp(\mathrm{Z_i^T}\lambda)}$$

Diasumsikan bahwa parameter μ_i dan φ masing-masing berantung pada variabel x_i ,sehingga model regresi ZINB dibagi menjadi dua komponen model yaitu:

a. Model data diskrit untuk μ_i adalah

$$\ln(\mu_i) = x_i^T\beta \quad , \mu_i \geq 0 = 1, \dots, n. \qquad (2.4)$$

b. Model excess zeros untuk φ adalah

$$log(\varphi) = \ln\left(\frac{\varphi}{1-\varphi}\right) = x_i^T\gamma \quad , 0 \leq \varphi \leq 1, i = 1, \dots, n. \qquad (2.5)$$

xi adalah matiks variabel yang memuat himpunan-himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan peluang pada *mean Negative Binomial* pada *Negative Binomial State* sedangkan β dan γ adalah parameter regresi yang akan ditaksir.

2.6.2. Estimasi Regresi *Zero Inflated Negative Binomial (ZINB)*

Estimasi parameter regresi ZINB menggunakan metode Maximum Likeelihood Estimation dengan prosedur Algoritma EM dan newton Rapshon. Metode ini biasanya digunakan untuk menaksir parameter suatu model yang diketahui fungsi densitasnya, sehingga fungsi log-likelihood dari probabilitas ZINB adalah :

$$\log L(\theta|y_i)$$
$$= \begin{cases} \displaystyle\sum_{i=1}^{n} ln\left\{e^{x_i^T,\gamma} + \left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{1/k}\right\} - \sum_{i=1}^{n} ln\left[1+e^{x_i^T\gamma}\right] \\ \displaystyle -\sum_{i=1}^{n} \ln[1+e^{x_i^T\gamma}] + \sum_{i=1}^{n} \ln[\mathrm{r}\left(\frac{1}{k}+y_i\right)] - \sum_{i=1}^{n} \ln[\mathrm{r}(y_i+1)] - \sum_{i=1}^{n} \ln[\mathrm{r}\left(\frac{1}{k}\right)] \\ \displaystyle +y_i\sum_{i=1}^{n} ln\left\{\left(\frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\right)^{1/k}\right\} + 1/k\sum_{i=1}^{n} ln\left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{1/k} \end{cases},$$

dengan i = 1,2,3,…,n. Estimasi dengan maksimum likelihood rasio dihitung dengan memaksimalkan log-likehoodnya. Karena fungsi log-likelihoodnya tidak linier jika tidak digunakan nilai awal yang bagus, sehingga fungsi likelihood ini tidak dapat diselesaikan dengan metode numerik biasa, hingga digunakan algoritma EM (*Expectation Maximization*). Algoritma EM dibagi menjadi dua langkah yaitu:

Misalkan variabel y_i (i=1,2,3,...,n) berkaitan dengan vektor variabel indikator W = $(w_1, w_2, \dots, w_n)^T$ yaitu:

$$w_i = \begin{cases} 1, \text{jika yi berasal dari zero state} \\ 0, \text{jika yi berasal dari binomial state} \end{cases}$$

dengan i=1,2,3,..,n , jika nilai variabel respon y_i = 1,2,…,n maka nilai w_i = 0. Sedangkan jika nilai variabel respon y_i = 0, maka nilai w_i mungkin 0 mungkin 1. Peluang dari w_i dapat dinyatakan :

P (w_i=1) = p_i

P (w_i=0) = 1- p_i

dengan i= 1,2,3,…,n sehingga distribusi dari variabel W adalah w_i ~ Binomial (1, p_i) mempunyai rataan dan variansi yaitu E(w_i)= p_i, dan var(w_i) = w_i(1- p_i). Distribusi gabungan antara y_i dan w_i yang terbentuk yaitu

$$f(w_i, y_i|p_i, \mu_i) = (p_i)^{w_i}(1-p_i)^{(1-w_i)}\left[\frac{\Gamma\left(y_i+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{1/k}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}\right]^{(1-w_i)},$$

didapat persamaan log likehoodnya :

$$\ln L(\beta, \gamma|y_i, w_i) = \sum_{i=1}^{n}\{w_i x_i^T\gamma - \ln[1+\exp(x_i^T\gamma)] + (1-w_i)\ln[g(y_i;\beta,1/k)]\}$$

dimana $g\left(y_i;\beta,\frac{1}{k}\right) = \frac{\Gamma\left(y_i+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{1/k}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}$, dan $\mu_i = e^{x_i^T\beta}$ dengan i=1,2,3,…,n.

1. Tahap expektasi (*E-step*)

Tahap perhitungan ekspektasi dari fungsi ln likelihood (mengganti variabel w_i dengan $w_i^{(m)}$ yang merupakan ekspektasi w_i).

$$w_i^{(m)} = E\left(w_i|y_i, \gamma^{(m)}, \beta^{(m)}\right)$$

$$= \begin{cases} \left[\left(1 + (e^{x_i^T \gamma^{(m,)}})\right)\left[\frac{1}{1+k^{(m)} e^{x_i^T \beta^{(m)}}}\right]^{1/k^{(m)}}\right]^{-1} & ,\text{jika } \mathrm{y_i} = 0 \\ = 0, jika\ \mathrm{y_i} = 1,2, \ldots \end{cases}$$

Sehingga didapat,

$$\log L\left(\gamma^{(m)} \middle| y_{i,} w_i^{(m)}\right) = \sum_{i=1}^{n} \left[w_i^{(m)} x_i^T \gamma - ln(1 + e^{x_i^T \gamma})\right]$$

$$\log L\left(\beta^{(m)} \middle| y_{i,} w_i^{(m)}\right) = \sum_{i=1}^{n} (1 - w_i^{(m)}) \left\{ \frac{\Gamma(\frac{1}{k} + y_i)}{\Gamma(y_i + 1)\Gamma(\frac{1}{k})} \left(\frac{e^{x_i^T \gamma}}{1 + e^{x_i^T \gamma}}\right)^{y_i} \left(\frac{1}{1 + k e^{x_i^T \beta}}\right) \right\}$$

2. Tahap maksimalisasi (*M-step*)

Memaksimalkan β dan γ dari hasil E-step dengan menghitung β^{m+1} dan γ^{m+1} dengan metode *Newton Rapshon* (Hall,2000).

2.7.Metode Bayes Empirik

Metode Bayes merupakan perpaduan antara sebaran prior dan sebaran posterior, yaitu jika dimisalkan dengan sebaran percontohan $X|\theta \sim f(x|\theta)$ dan sebaran prior $\theta \sim \pi(\theta)$diketahui maka sebaran posterior dari θ adalah

$$\pi(\theta|x) = \frac{f(x.\theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{m(x)} \qquad \text{dengan } m(x) = \int f(x|\theta)\pi(\theta)d\theta$$

Suatu sebaran prior dinamakan konjugate bila menghasilkan sebaran posterior yang sama dengan dirinya. Sebaran yang masih dalam keluarga eksponensial mempunyai prior konjugate. Sebaran poisson memiliki prior konjugate Gamma (Abadi, 2011).

Metode Bayes Linier Empirik adalah 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 poison gamma, $y_i \sim Poisson\ (\theta_i)$, dan prior distributionn $\theta_i \sim gamma\left(\frac{\mu_i^2}{\sigma^2}, \frac{\sigma^2}{\mu_i}\right)$, 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 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 Linier Mixed Model*, dengan memanfaatkan sebaran marginal yaitu Negative Binomial. Dimana $\hat{\beta}$ adalah penduga parameter regresi binomial-negatif sedangkan $\hat{\alpha}$ adalah penduga bagi parameter dispersi distribusi binomial-negatif (=$\widehat{\emptyset}$). Bila model yang digunakan adalah linier dan model poisson gamma dalam fungsi link logaritmik maka parameter regresi yang diperoleh perlu dieksponensialkan terlebih dahulu.

Dengan demikian penduga bayes adalah :

$$\hat{\theta}_i^{EB} = \hat{\gamma}_i\hat{\theta}_i^{L} + (1 - \hat{\gamma}_i)\hat{\theta}_i^{TL} \qquad (2.6)$$

dengan penduga langsung bagi θ adalah y_i dan penduga tak langsugnya adalah $X'\hat{\beta}$.

2.8.*Mean Square Eror (MSE)*

Mean Square Eror 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 besar karena kesalahan-kesalahan itu dikuadratkan.

Salah satunya adalah dengan uji signifikasi koefisien β. Namun dalam konteks SAE, metode evaluasi yang digunakan adalah MSE. Hal ini disebabkan pada persamaan pendugaan SAE yang tidak muncul nilai β, melainkan nilai θ yang merupakan nilai pendugaan area pada SAE. Adapun persamaan MSE diumuskan sebagai berikut

$$MSE(y) = \frac{\sum_{i=1}^{n}\left(\theta_i - \hat{\theta}_i\right)^2}{n} \tag{2.7}$$

BAB 3. METODOLOGI PENELITIAN

3.1 Model Penelitian

Model yang digunakan dalam penelitian ini merupakan model area kecil dan metode *Zero Inflated Negative Binomial* (ZINB) dan *Negative Binomial* (NB). Untuk memudahkan dalam proses perhitungan digunakan salah satu software statistika dengan Program R 3.3.3 dan Data yang digunakan dalam penelitian ini merupakan data simulasi berdistribusi *Zero-Inflated Poisson* (ZIP).

3.2 Pembangkitan Data

Data yang digunakan dalam penelitian ini merupakan data simulasi dengan cara membangkitkan data berdistribusi *Zero-Inflated poisson* pada *Small Area Estimation* (SAE). Area kecil yang digunakan dalam data berikut adalah 1 sampai 50 area dengan 60% hingga 90% bernilai nol. Dengan menetapkan nilai $x, \beta = (2{,}0; 0{,}3)$,dan $\sigma^2 = (2{,}0)$. Variabel yang digunakan merupakan variabel tak bebas (Y) dan Variabel bebas (X).

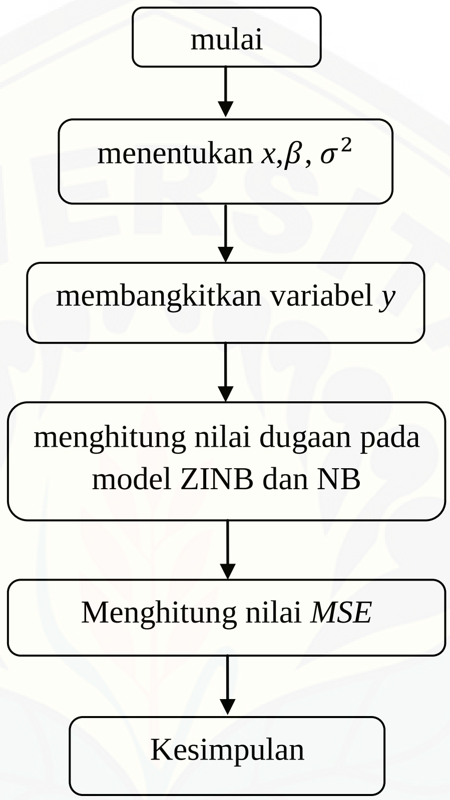
3.3 Langkah-Langkah Simulasi

Adapun langkah-langkah simulasi dalam penelitian ini yaitu :

1. Menentukan nilai variabel bebas x,β,dan σ^2,agar nilai μ yang diperoleh dari $x'\beta$ bertambah besar sehingga nilai keragaman y_i bertambah besar pula.
2. Menentukan nilai penduga lansung (y_D) menggunakan metode ZIP.
3. Melakukan pendugaan parameter dengan regresi ZINB dan regresi NB.
4. Melakukan pendugaan area model ZINB dan model NB pada SAE. Pendugaan area dilakukan dengan menambakan penduga langsung dan penduga tak langsung yang dibangun oleh model regresi.
5. Menghitung nilai MSE.

6. Kesimpulan yang menjawab rumusan masalah.

Skema dari langkah-langkah penelitian dapat dilihat pada gambar 3.2 dibawah ini.



Gambar 3.2 Skema Desain Penelitian

3.4 Struktur dan program fungsi R

Metode yang digunakan untuk menganalisis data pada penelitian ini dengan metode ZINB. Untuk mempermudah dalam proses pehitungan digunakan salah satu software dalam statistik yaitu program R 3.3.3. Paket yang digunakan adalah *pscl.*

```
zeroinfl(formula, data, subset, na.action, weights,
     offset,
    dist = c("poisson", "negbin", "geometric"),
    link = c("logit", "probit", "cloglog", "cauchit",
    "log"),
```

```
control = zeroinfl.control(...),
model = TRUE, y = TRUE, x = FALSE, ...)
```

dengan keterangan fungsi sebagai berikut:

a. *formula* = Deskripsi simbolis dari model.

b. *data* = subset

c. *argumen* = mengendalikan pemrosesan formula melalui model.frame.

d. bobot = vektor angka numerik opsional.

e. *offset* = vektor numerik opsional dengan komponen yang dikenal secara priori untuk dimasukkan dalam prediksi linier model hitungan.

f. *dist* = spesifikasi karakteristik keluarga model hit (link log selalu digunakan). karakter link spesifikasi fungsi link dalam model zero-inflation biner (keluarga bi-nomial).

g. *control* = daftar argumen kontrol yang ditentukan melalui zeroinfl.control. Jika TRUE komponen komponen yang sesuai (kerangka model, re-sponsor, matriks model) dikembalikan.

BAB 5. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa :

a. Model ZINB pada *small area* dtitentukan dengan komposisi pendugaan langsung dan tak langsung dengan penduga langsung bagi θ adalah y_i dan penduga tak langsungnya adalah $X'\hat{\beta}$

b. Estimasi dengan model ZINB dapat mengatasi masalah *exces zeroes* pada penelitian dengan 60% nilai nol. Lebih dari 60% nilai nol model ZINB sudah tidak optimal dalam mengatasi excess zeroes. Hal ini ditunjukkan dengan nilai *Mean Square Eror* (MSE) yang semakin membesar.

5.2. Saran

Pada penelitian ini Model ZINB dalam masalah excess zeros hanya mampu mengatasi tidak lebih dari 60%, sehingga saran untuk penelitian selanjutnya dapat dilakukan penanganan *excess zeros* dengan metode lain seperti *Zero inflated Generalized Poisson* (ZIGP).

.

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.

Ariawan, B., Suparti, dan Sudarno. 2012. Pemodelan Regresi *Zero-Inflated Negative Binomial* (ZINB) Untuk Data Respon Diskrit dengan *Excees Zeroes*. *Jurnal Gaussian* vol. 1(1) : 55-64.

Darsyah, M.Y, Rumiati, A.T, Otok, B.W. (2012). *Small Area Estimation* terhadap Pengeluaran Per Kapita di Kabupaten Sumenep dengan pendekatan *Kernel-Bootstrap*. Prosiding Semnas MIPA UNESA, Surabaya.

Hall, D.B. 2000. "*Zero-Inflated Poisson and Binomial Regression with Random Effects* : A Case Study". Biometrics. Vol.56. pp. 1030-1039.

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

Mukhopadhay P, Maiti T. (2004). *Two Stage Non-Parametric Approach fot Small Area Estimation*. P*roceeding of ASA Section on Survey Research Methods,* hal. 4058-4065.

Oktarin, S. 2015. "Penanganan Overdispersi Menggunakan Regresi M-Kuantil dengan Resampling Bootstrap untuk *Small Area Estimation*". Tidak Diterbitkan. Skripsi. Jember:Jurusan Matematika FMIPA Universitas Jember.

Prasad, N.G.N. dan Rao, J.N.K. (1990). *The Estimation of The Mean Squared Error of The Small Area Estimators*. Journal of American Statistical Association, 85, hal.163-171.

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

Zulfi, M. 2016. “Perbandingan Regresi Binomial Negatif dan Regresi M-Kuantil Pada Data Excees Zeroes dalam Konteks *Small Area Estimation*”. Tidak Diterbitkan. Skripsi. Jember:Jurusan Matematika FMIPA Universitas Jember.

LAMPIRAN

LAMPIRAN A. Hasil Simulasi Data untuk 60% hingga 90% nilai nol.

Area	X0	X1	μ_i	β_0	β_1	σ^2	y_D	y_D	y_D	y_D
1	1	0	2.0	2.0	0.3	2.0	0	0	0	0
2	1	0	2.0				3	0	0	0
3	1	0	2.0				0	0	0	0
4	1	1	2.3				0	0	1	0
5	1	1	2.3				0	0	0	0
6	1	1	2.3				0	1	0	0
7	1	2	2.6				2	0	0	0
8	1	2	2.6				5	0	0	0
9	1	2	2.6				0	0	0	0
10	1	3	2.9				2	3	4	0
11	1	3	2.9				0	0	0	0
12	1	3	2.9				0	0	0	0
13	1	4	3.2				0	0	6	0
14	1	4	3.2				4	2	0	0
15	1	4	3.2				2	6	0	0
16	1	5	3.5				0	0	3	4
17	1	5	3.5				1	0	0	0
18	1	5	3.5				0	1	0	0
19	1	6	3.8				0	1	0	0
20	1	6	3.8				0	4	0	0
21	1	6	3.8				0	3	0	0
22	1	7	4.1				3	0	0	0
23	1	7	4.1				0	0	0	0
24	1	7	4.1				0	0	0	0
25	1	8	4.4				0	0	0	0
26	1	8	4.4				0	0	0	5
27	1	8	4.4				3	0	4	0
28	1	9	4.7				6	0	0	0
29	1	9	4.7				4	0	0	0
30	1	9	4.7				0	0	0	0
31	1	10	5.0				4	0	5	0
32	1	10	5.0				0	0	0	0
33	1	10	5.0				0	0	0	0
34	1	11	5.3				6	8	0	0
35	1	11	5.3				0	0	4	0
36	1	11	5.3				7	5	0	0
37	1	12	5.6				11	0	0	0

38	**1**	**12**	**5.6**				**0**	**0**	**0**	**0**
39	**1**	**12**	**5.6**				**0**	**0**	**3**	**0**
40	**1**	**13**	**5.9**				**0**	**0**	**0**	**0**
41	**1**	**13**	**5.9**				**3**	**8**	**0**	**0**
42	**1**	**13**	**5.9**				**2**	**7**	**2**	**0**
43	**1**	**14**	**6.2**				**0**	**0**	**0**	**5**
44	**1**	**14**	**6.2**				**5**	**8**	**0**	**3**
45	**1**	**14**	**6.2**				**0**	**0**	**0**	**0**
46	**1**	**15**	**6.5**				**0**	**6**	**8**	**0**
47	**1**	**15**	**6.5**				**0**	**0**	**0**	**0**
48	**1**	**15**	**6.5**				**6**	**0**	**0**	**0**
49	**1**	**16**	**6.8**				**7**	**0**	**0**	**0**
50	**1**	**16**	**6.8**				**0**	**0**	**0**	**0**

LAMPIRAN B. Nilai Penduga Regresi Model ZINB dan NB

60% nilai nol		70% nilai nol	
MODEL ZINB	MODEL NB	MODEL ZINB	MODEL NB
0.7646244	0.7680679	0.4291138	0.4101932
0.7646244	0.7680679	0.4291138	0.4101932
0.7646244	0.7680679	0.4291138	0.4101932
0.8396329	0.8410968	0.4829517	0.4639306
0.8396329	0.8410968	0.4829517	0.4639306
0.8396329	0.8410968	0.4829517	0.4639306
0.9216284	0.9210693	0.5435403	0.5247079
0.9216284	0.9210693	0.5435403	0.5247079
0.9216284	0.9210693	0.5435403	0.5247079
1.0112189	1.0086457	0.6117253	0.5934472
1.0112189	1.0086457	0.6117253	0.5934472
1.0112189	1.0086457	0.6117253	0.5934472
1.1090604	1.1045490	0.6884587	0.6711918
1.1090604	1.1045490	0.6884587	0.6711918
1.1090604	1.1045490	0.6884587	0.6711918
1.2158611	1.2095708	0.7748115	0.7591212
1.2158611	1.2095708	0.7748115	0.7591212
1.2158611	1.2095708	0.7748115	0.7591212
1.3323845	1.3245783	0.8719889	0.8585699
1.3323845	1.3245783	0.8719889	0.8585699
1.3323845	1.3245783	0.8719889	0.8585699
1.4594536	1.4505208	0.9813469	0.9710469
1.4594536	1.4505208	0.9813469	0.9710469
1.4594536	1.4505208	0.9813469	0.9710469

1.5979549	1.5884381	1.1044113	1.0982589
1.5979549	1.5884381	1.1044113	1.0982589
1.5979549	1.5884381	1.1044113	1.0982589
1.7488428	1.7394688	1.2428989	1.2421363
1.7488428	1.7394688	1.2428989	1.2421363
1.7488428	1.7394688	1.2428989	1.2421363
1.9131445	1.9048596	1.3987415	1.4048624
1.9131445	1.9048596	1.3987415	1.4048624
1.9131445	1.9048596	1.3987415	1.4048624
2.0919649	2.0859760	1.5741125	1.5889064
2.0919649	2.0859760	1.5741125	1.5889064
2.0919649	2.0859760	1.5741125	1.5889064
2.2864916	2.2843132	1.7714575	1.7970611
2.2864916	2.2843132	1.7714575	1.7970611
2.2864916	2.2843132	1.7714575	1.7970611
2.4980011	2.5015085	1.9935281	2.0324851
2.4980011	2.5015085	1.9935281	2.0324851
2.4980011	2.5015085	1.9935281	2.0324851
2.7278645	2.7393550	2.2434203	2.2987508
2.7278645	2.7393550	2.2434203	2.2987508
2.7278645	2.7393550	2.2434203	2.2987508
2.9775537	2.9998162	2.5246173	2.5998987
2.9775537	2.9998162	2.5246173	2.5998987
2.9775537	2.9998162	2.5246173	2.5998987
3.2486488	3.2850424	2.8410383	2.9404985
3.2486488	3.2850424	2.8410383	2.9404985

80% nilai nol		90% nilai nol	
MODEL ZINB	MODEL NB	MODEL ZINB	MODEL NB
0.4975960	0.4910377	0.1341830	0.09196529
0.4975960	0.4910377	0.1341830	0.09196529
0.4975960	0.4910377	0.1341830	0.09196529
0.5264880	0.5203459	0.1493758	0.10640555
0.5264880	0.5203459	0.1493758	0.10640555
0.5264880	0.5203459	0.1493758	0.10640555
0.5570402	0.5514033	0.1662125	0.12311320
0.5570402	0.5514033	0.1662125	0.12311320
0.5570402	0.5514033	0.1662125	0.12311320
0.5893471	0.5843144	0.1848517	0.14244427
0.5893471	0.5843144	0.1848517	0.14244427
0.5893471	0.5843144	0.1848517	0.14244427
0.6235080	0.6191899	0.2054632	0.16481067
0.6235080	0.6191899	0.2054632	0.16481067
0.6235080	0.6191899	0.2054632	0.16481067
0.6596281	0.6561469	0.2282264	0.19068900
0.6596281	0.6561469	0.2282264	0.19068900
0.6596281	0.6561469	0.2282264	0.19068900
0.6978182	0.6953098	0.2533304	0.22063072
0.6978182	0.6953098	0.2533304	0.22063072
0.6978182	0.6953098	0.2533304	0.22063072
0.7381957	0.7368101	0.2809720	0.25527385
0.7381957	0.7368101	0.2809720	0.25527385
0.7381957	0.7368101	0.2809720	0.25527385
0.7808839	0.7807875	0.3113542	0.29535659

0.7808839	0.7807875	0.3113542	0.29535659
0.7808839	0.7807875	0.3113542	0.29535659
0.8260137	0.8273897	0.3446834	0.34173307
0.8260137	0.8273897	0.3446834	0.34173307
0.8260137	0.8273897	0.3446834	0.34173307
0.8737227	0.8767733	0.3811660	0.39539152
0.8737227	0.8767733	0.3811660	0.39539152
0.8737227	0.8767733	0.3811660	0.39539152
0.9241563	0.9291045	0.4210046	0.45747534
0.9241563	0.9291045	0.4210046	0.45747534
0.9241563	0.9291045	0.4210046	0.45747534
0.9774683	0.9845592	0.4643924	0.52930747
0.9774683	0.9845592	0.4643924	0.52930747
0.9774683	0.9845592	0.4643924	0.52930747
1.0338206	1.0433237	0.5115074	0.61241859
1.0338206	1.0433237	0.5115074	0.61241859
1.0338206	1.0433237	0.5115074	0.61241859
1.0933842	1.1055957	0.5625054	0.70857970
1.0933842	1.1055957	0.5625054	0.70857970
1.0933842	1.1055957	0.5625054	0.70857970
1.1563397	1.1715844	0.6175121	0.81983989
1.1563397	1.1715844	0.6175121	0.81983989
1.1563397	1.1715844	0.6175121	0.81983989
1.2228775	1.2415117	0.6766141	0.94857000
1.2228775	1.2415117	0.6766141	0.94857000

LAMPIRAN C. Nilai Penduga Area Model ZINB dan NB

Area	θ	60%		θ	70%	
		(θ_{ZINB})	(θ_{NB})		(θ_{ZINB})	(θ_{NB})
1	**0**	**0.4944609**	**0.4958986**	0	0.3158276	0.3054577
2	**3**	**1.5544461**	**1.5589659**	0	0.3158276	0.3054577
3	0	0.4944609	0.4958986	0	0.3158276	0.3054577
4	0	0.5247775	0.5253489	0	0.3440563	0.3342922
5	0	0.5247775	0.5253489	0	0.3440563	0.3342922
6	0	0.5247775	0.5253489	1	0.6316532	0.6137271
7	2	1.3498188	1.3493249	0	0.3737353	0.3647342
8	5	2.5410326	2.5401029	0	0.3737353	0.3647342
9	0	0.5556763	0.5554730	0	0.3737353	0.3647342
10	2	1.4259915	1.4238828	3	1.4197649	1.3914091
11	0	0.5870341	0.5861660	0	0.4047565	0.3966727
12	0	0.5870341	0.5861660	0	0.4047565	0.3966727
13	0	0.6187211	0.6173145	0	0.4369827	0.4299619
14	4	2.3872066	2.3817795	2	1.1675306	1.1487725
15	2	1.5029638	1.4995470	6	2.6286264	2.5863936
16	0	0.6506031	0.6487977	0	0.4702482	0.4644223
17	1	1.1155065	1.1124110	0	0.4702482	0.4644223
18	0	0.6506031	0.6487977	0	0.4702482	0.4644223
19	0	0.6825440	0.6804896	1	0.9259579	0.9176621
20	0	0.6825440	0.6804896	1	0.9259579	0.9176621
21	0	0.6825440	0.6804896	4	2.1907464	2.1711191
22	3	2.2458971	2.2391471	3	1.8910377	1.8800823
23	0	0.7144077	0.7122606	0	0.5391103	0.5359871
24	0	0.7144077	0.7122606	0	0.5391103	0.5359871

25	0	0.7460610	0.7439800	0	0.5742638	0.5725959
26	0	0.7460610	0.7439800	0	0.5742638	0.5725959
27	0	0.7460610	0.7439800	0	0.5742638	0.5725959
28	3	2.4438500	2.4380099	0	0.6095810	0.6093975
29	6	4.1103246	4.1005020	0	0.6095810	0.6093975
30	4	2.9993415	2.9921739	0	0.6095810	0.6093975
31	0	0.8082293	0.8067469	0	0.6448165	0.6461143
32	4	3.1183846	3.1126653	0	0.6448165	0.6461143
33	0	0.8082293	0.8067469	0	0.6448165	0.6461143
34	0	0.8385094	0.8375456	0	0.6797270	0.6824709
35	6	4.4335664	4.4284702	8	5.2251989	5.2462918
36	0	0.8385094	0.8375456	0	0.6797270	0.6824709
37	7	5.2104215	5.2085357	5	3.6985682	3.7199324
38	11	7.6917409	7.6889569	0	0.7140780	0.7182028
39	0	0.8681126	0.8677984	0	0.7140780	0.7182028
40	0	0.8969470	0.8973988	0	0.7476504	0.7530637
41	0	0.8969470	0.8973988	0	0.7476504	0.7530637
42	3	2.8197493	2.8211696	8	5.7473400	5.7889535
43	2	2.2467958	2.2499958	7	5.3456961	5.3908244
44	0	0.9249324	0.9262497	0	0.7802453	0.7868321
45	5	4.2295909	4.2356150	8	5.9979034	6.0485376
46	0	0.9520009	0.9542652	0	0.8116883	0.8193157
47	0	0.9520009	0.9542652	6	4.8826316	4.9285134
48	6	5.0336458	5.0456179	0	0.8116883	0.8193157
49	**7**	**5.8705501**	**5.8901969**	0	0.8418328	0.8503555
50	**0**	**0.9780972**	**0.9813706**	0	0.8418328	0.8503555

θ	80%		θ	90%	
	(θ_{ZINB})	(θ_{NB})		(θ_{ZINB})	(θ_{NB})
0	0.3446261	0.3414675	0	0.1188657	0.08450219
0	0.3446261	0.3414675	0	0.1188657	0.08450219
0	0.3446261	0.3414675	0	0.1188657	0.08450219
1	0.6778051	0.6724045	0	0.1306358	0.09654046
0	0.3582417	0.3553873	0	0.1306358	0.09654046
0	0.3582417	0.3553873	0	0.1306358	0.09654046
0	0.3721296	0.3696055	0	0.1433334	0.11009637
0	0.3721296	0.3696055	0	0.1433334	0.11009637
0	0.3721296	0.3696055	0	0.1433334	0.11009637
4	1.7645574	1.7546520	0	0.1569837	0.12530331
0	0.3862755	0.3841071	0	0.1569837	0.12530331
0	0.3862755	0.3841071	0	0.1569837	0.12530331
6	2.5450922	2.5337376	0	0.1716032	0.14228972
0	0.4006632	0.3988757	0	0.1716032	0.14228972
0	0.4006632	0.3988757	0	0.1716032	0.14228972
3	1.5265946	1.5215126	0	0.1871972	0.16117369
0	0.4152757	0.4138932	4	0.9062934	0.78030362
0	0.4152757	0.4138932	0	0.1871972	0.16117369
0	0.4300944	0.4291402	0	0.2037589	0.18205631
0	0.4300944	0.4291402	0	0.2037589	0.18205631
0	0.4300944	0.4291402	0	0.2037589	0.18205631
0	0.4450997	0.4445957	0	0.2212674	0.20501434
0	0.4450997	0.4445957	0	0.2212674	0.20501434
0	0.4450997	0.4445957	0	0.2212674	0.20501434
0	0.4602710	0.4602375	0	0.2396862	0.23009221

0	0.4602710	0.4602375	0	0.2396862	0.23009221
4	2.1025788	2.1024256	5	1.3905943	1.33493270
0	0.4755865	0.4760424	0	0.2589628	0.25729387
0	0.4755865	0.4760424	0	0.2589628	0.25729387
0	0.4755865	0.4760424	0	0.2589628	0.25729387
5	2.6810713	2.6863241	0	0.2790276	0.28657527
0	0.4910238	0.4919859	0	0.2790276	0.28657527
0	0.4910238	0.4919859	0	0.2790276	0.28657527
0	0.5065597	0.5080428	0	0.2997946	0.31783809
4	2.3140316	2.3208065	0	0.2997946	0.31783809
0	0.5065597	0.5080428	0	0.2997946	0.31783809
0	0.5221703	0.5241870	0	0.3211616	0.35092559
0	0.5221703	0.5241870	0	0.3211616	0.35092559
3	1.9195497	1.9269635	0	0.3211616	0.35092559
0	0.5378314	0.5403921	0	0.3430118	0.38562157
0	0.5378314	0.5403921	0	0.3430118	0.38562157
2	1.4973580	1.5044872	0	0.3430118	0.38562157
0	0.5535184	0.5566308	0	0.3652159	0.42165276
0	0.5535184	0.5566308	5	2.1188838	2.44631518
0	0.5535184	0.5566308	3	1.4174166	1.63645021
8	4.6312166	4.6610716	0	0.3876349	0.45869535
0	0.5692068	0.5728761	0	0.3876349	0.45869535
0	0.5692068	0.5728761	0	0.3876349	0.45869535
0	0.5848718	0.5891007	0	0.4101230	0.49638524
0	0.5848718	0.5891007	0	0.4101230	0.49638524

LAMPIRAN C. Skrip Pogram R Model ZINB

```
> 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
> #penetapan-beta
> b0<-2
> b1<-0.3
> b<-as.matrix(c(b0,b1))
> #hitung_miu
> 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,0.6)
> Y
 [1]  0  3  0  0  0  0  2  5  0  2  0  0  0  4  2  0  1  0  0  0  0  3  0  0  0  0  0  3  6  4  0  4  0
0  6  0  7 11  0  0  0  3  2  0  5  0  0  6  7  0
> summary(Y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00    0.00    0.00    1.72    3.00   11.00 
> tabel<-data.frame(Y=Y,X0=X0,X1=X1)
> #ZINB
> library(pscl)
Classes and Methods for R developed in the
Political Science Computational Laboratory
Department of Political Science
Stanford University
Simon Jackman
hurdle and zeroinfl functions by Achim Zeileis
> ZINB<-zeroinfl(Y~X1,data=tabel,dist='negbin',EM=TRUE)
> summary(ZINB)

Call:
zeroinfl(formula = Y ~ X1, data = tabel, dist = "negbin", EM = TRUE)

Pearson residuals:
    Min      1Q  Median      3Q     Max 
```

-0.8686 -0.6981 -0.5762 0.6857 2.9741

Count model coefficients (negbin with log link):

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.82752	0.28437	2.910	0.00361	**
X1	0.06565	0.02556	2.568	0.01022	*
Log(theta)	6.12320	32.82544	0.187	0.85202	

Zero-inflation model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.68907	0.59851	1.151	0.25
X1	-0.04225	0.06266	-0.674	0.50

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 456.3208

Number of iterations in BFGS optimization: 1

Log-likelihood: -73.31 on 5 Df

> y_htZINB<-fitted.values(ZINB)

> print(y_htZINB)

1 2 3 4 5 6 7 8 9 10 11 12

0.7646244 0.7646244 0.7646244 0.8396329 0.8396329 0.8396329 0.9216284 0.9216284 0.9216284 1.0112189 1.0112189 1.0112189

13 14 15 16 17 18 19 20 21 22 23 24

```
1.1090604 1.1090604 1.1090604 1.2158611 1.2158611 1.2158611 1.3323845
1.3323845 1.3323845 1.4594536 1.4594536 1.4594536

       25        26        27        28        29        30        31        32        33        34        35
36

1.5979549 1.5979549 1.5979549 1.7488428 1.7488428 1.7488428 1.9131445
1.9131445 1.9131445 2.0919649 2.0919649 2.0919649

       37        38        39        40        41        42        43        44        45        46        47
48

2.2864916 2.2864916 2.2864916 2.4980011 2.4980011 2.4980011 2.7278645
2.7278645 2.7278645 2.9775537 2.9775537 2.9775537

       49        50

3.2486488 3.2486488

> #pendugaan area kecil

> library(gamlss)

> NB<-glm.nb(Y~X1,data=tabel)

> psi_ZINB<-unlist(NB[24])

> psi_ZINB

   theta

0.3360701

> GAMht_i_ZINB<-y_htZINB/(y_htZINB+(exp(psi_ZINB)))

> direct<-GAMht_i_ZINB*Y

> indirect<-(1-GAMht_i_ZINB)*y_htZINB

> teta_EB<-direct+indirect

> print(teta_EB)

        1         2         3         4         5         6         7         8

0.4944609 1.5544461 0.4944609 0.5247775 0.5247775 0.5247775 1.3498188
2.5410326
```

```
        9        10        11        12        13        14        15        16 
0.5556763 1.4259915 0.5870341 0.5870341 0.6187211 2.3872066 1.5029638 0.6506031 
       17        18        19        20        21        22        23        24 
1.1155065 0.6506031 0.6825440 0.6825440 0.6825440 2.2458971 0.7144077 0.7144077 
       25        26        27        28        29        30        31        32 
0.7460610 0.7460610 0.7460610 2.4438500 4.1103246 2.9993415 0.8082293 3.1183846 
       33        34        35        36        37        38        39        40 
0.8082293 0.8385094 4.4335664 0.8385094 5.2104215 7.6917409 0.8681126 0.8969470 
       41        42        43        44        45        46        47        48 
0.8969470 2.8197493 2.2467958 0.9249324 4.2295909 0.9520009 0.9520009 5.0336458 
       49        50 
5.8705501 0.9780972 
> 
> #MSE-ZINB
> eZINB<-Y-teta_EB
> sse<-as.numeric(t(eZINB)%*%eZINB)
> mseZINB<-sse/n
> mseZINB
[1] 1.076108
```

LAMPIRAN D. Skrip Pogram R Model NB

```
> #glm_NB

> library(gamlss)

> NB<-glm.nb(Y~X1,data=tabel)

> summary(NB)


Call:

glm.nb(formula = Y ~ X1, data = tabel, init.theta = 0.3360701223,

   link = log)


Deviance Residuals:

    Min       1Q   Median       3Q      Max

-1.2640  -1.0830  -0.9298   0.4254   1.2200


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.26388    0.53764  -0.491    0.624

X1           0.09083    0.05655   1.606    0.108


(Dispersion parameter for Negative Binomial(0.3361) family taken to be 1)


    Null deviance: 45.164  on 49  degrees of freedom

Residual deviance: 42.562  on 48  degrees of freedom
```

```
AIC: 170.28

Number of Fisher Scoring iterations: 1


              Theta:  0.336
          Std. Err.:  0.119

 2 x log-likelihood:  -164.283
> y_htNB<-fitted.values(NB)
> print(y_htNB)
        1         2         3         4         5         6         7         8 
0.7680679 0.7680679 0.7680679 0.8410968 0.8410968 0.8410968 0.9210693 0.9210693 
        9        10        11        12        13        14        15        16 
0.9210693 1.0086457 1.0086457 1.0086457 1.1045490 1.1045490 1.1045490 1.2095708 
       17        18        19        20        21        22        23        24 
1.2095708 1.2095708 1.3245783 1.3245783 1.3245783 1.4505208 1.4505208 1.4505208 
       25        26        27        28        29        30        31        32 
1.5884381 1.5884381 1.5884381 1.7394688 1.7394688 1.7394688 1.9048596 1.9048596 
       33        34        35        36        37        38        39        40 
```

```
1.9048596 2.0859760 2.0859760 2.0859760 2.2843132 2.2843132 2.2843132
2.5015085
       41        42        43        44        45        46        47        48
2.5015085 2.5015085 2.7393550 2.7393550 2.7393550 2.9998162 2.9998162
2.9998162
       49        50
3.2850424 3.2850424
> #teta-duga-NB
> psi_NB<-unlist(NB[24])
> psi_NB
    theta
0.3360701
> GAMht_i_NB<-y_htNB/(y_htNB+(exp(psi_NB)))
> direct<-GAMht_i_NB*Y
> indirect<-(1-GAMht_i_NB)*y_htNB
> teta_EB<-direct+indirect
> print(teta_EB)
        1         2         3         4         5         6         7         8
0.4958986 1.5589659 0.4958986 0.5253489 0.5253489 0.5253489 1.3493249
2.5401029
        9        10        11        12        13        14        15        16
0.5554730 1.4238828 0.5861660 0.5861660 0.6173145 2.3817795 1.4995470
0.6487977
       17        18        19        20        21        22        23        24
```

1.1124110 0.6487977 0.6804896 0.6804896 0.6804896 2.2391471 0.7122606 0.7122606

25 26 27 28 29 30 31 32

0.7439800 0.7439800 0.7439800 2.4380099 4.1005020 2.9921739 0.8067469 3.1126653

33 34 35 36 37 38 39 40

0.8067469 0.8375456 4.4284702 0.8375456 5.2085357 7.6889569 0.8677984 0.8973988

41 42 43 44 45 46 47 48

0.8973988 2.8211696 2.2499958 0.9262497 4.2356150 0.9542652 0.9542652 5.0456179

49 50

5.8901969 0.9813706

> #MSE-NB

> eNB<-Y-teta_EB

> sse<-as.numeric(t(eNB)%*%eNB)

> mseNB<-sse/n

> mseNB

[1] 1.076948