

Algoritma Komputasi Generalized Estimating Equation

Bertho Tantular, Zulhanif, Gumgum Darmawan

Departemen Statistika FMIPA Universitas Padjadjaran

bertho@unpad.ac.id

Abstrak—*Generalized estimating equation* merupakan metode yang digunakan dalam penaksiran parameter untuk data yang berkorelasi antar pengamatan didalam kelompok akan tetapi antar kelompok saling bebas. Jenis data yang umumnya dianalisis menggunakan *generalized estimating equation* adalah data longitudinal. *Generalized estimating equation* tidak menggunakan fungsi likelihood yang diperoleh dari distribusi komponen acak, tetapi menggunakan pendekatan *quasi-likelihood* yang selanjutnya dilakukan proses iterasi hingga konvergen. Proses komputasi untuk metode ini menggunakan algoritma *Fisher's scoring*.

Kata kunci: *Generalized estimating equation, Fisher's scoring algorithm*

I. PENDAHULUAN

Generalized Estimating Equations (GEE), pertama kali diperkenalkan oleh Liang dan Zeger pada tahun 1986, metode ini menjadi populer dalam berbagai bidang ilmu. GEE merupakan metode yang digunakan dalam penaksiran parameter untuk data yang berkorelasi antar pengamatan didalam kelompok akan tetapi antar kelompok saling bebas. Jenis data yang umumnya dianalisis menggunakan GEE adalah data *clustered* yaitu data yang di dalam klaster mungkin berkorelasi atau data hasil pengukuran berulang (*repeated measurement*) yang biasa disebut data longitudinal. Menurut Hardin dan Hilbe (2003) meskipun *Generalized Linier Model* (GLM) adalah metode yang kuat dan fleksibel digunakan, namun analisisnya terbatas, asumsi dari GLM adalah setiap individu harus independen dengan individu yang lainnya. Namun pada data longitudinal dan data yang berasal dari pengklasteran asumsi ini mungkin bisa tidak terpenuhi.

Dalam proses penaksirannya GEE tidak menggunakan fungsi likelihood yang diperoleh dari perkalian fungsi densitas atau fungsi massa distribusi variable respon. Penaksiran parameter GEE menggunakan pendekatan *quasi-likelihood* yang tidak membutuhkan asumsi distribusi. Walaupun demikian dibutuhkan varians yang diperoleh dari ekspektasi.

Keuntungan utama dari metode ini adalah prosedur komputasi untuk metode ini lebih sederhana dari *generalized linear model*. Selain itu metode ini juga tidak membutuhkan asumsi distribusi multivariate. Keuntungan lainnya adalah penaksir yang dihasilkan bersifat konsisten meskipun ada kesalahan spesifikasi pada struktur korelasi yang digunakan dan dapat diterapkan pada data yang tidak lengkap (terdapat data hilang), dengan pengamatan yang hilang merupakan MCAR. (Hardin dan Hilbe, 2003). GEE dapat diimplementasikan dengan perangkat lunak statistika yang umum tersedia terutama yang menggunakan *command-base* seperti R, SAS, Stata, atau S-Plus.

Pada paper ini akan diuraikan algoritma komputasi yang digunakan untuk prosedur penaksiran parameter menggunakan GEE. Oleh karena GEE menggunakan *quasi-likelihood* dalam menentukan taksiran parameternya yang pada prosesnya melibatkan matriks korelasi antar pengamatan, maka prosesnya memerlukan iterasi. Algoritma optimasi harus dilibatkan pada proses iterasinya, dalam hal ini metode optimasi yang dipilih adalah algoritma *Fisher's scoring*.

II. LANDASAN TEORI

Metode penaksiran parameter untuk data yang mengandung struktur korelasi seperti data longitudinal tanpa membutuhkan asumsi distribusi yang ketat dapat menggunakan *Generalized Estimating Equation*. Hardin dan Hilbe (2003) menyatakan bahwa meskipun *Generalized Linier Model* (GLM) merupakan metode yang kuat dan fleksibel digunakan, namun analisisnya terbatas, asumsi dari GLM adalah setiap individu harus independen dengan individu yang lainnya. Namun pada data longitudinal asumsi ini mungkin bisa tidak terpenuhi. Keuntungan utama dari penggunaan GEE terletak

pada hasil penaksiran parameter yang konsisten dan tidak bias, bahkan ketika struktur korelasi *misspecified* (Ghisletta dan Spini, 2004). GEE merupakan penyesuaian pada GLM ketika data berkorelasi, maka metode ini dapat diterapkan untuk berbagai variabel respon yang sering dijumpai dalam aplikasi empirik (misalnya, data kontinu, polikotomus, dikotomus, dan pencacahan). Meskipun GEE tidak mengharuskan data mengikuti distribusi tertentu, tetapi mengasumsikan varians dari variabel respon dijelaskan sebagai sebuah fungsi ekspektasi.

Salah satu hal penting yang membedakan GEE dengan metode GLM adalah adanya struktur korelasi yang disertakan dalam penaksiran parameternya. Struktur korelasi ini disebut sebagai *Working Correlation Matrix* (WCM). Penentuan WCM sangat bergantung pada kondisi korelasi yang terjadi pada data yang digunakan. Menurut Liang dan Zeger (1986), GEE tetap *robust* meskipun WCM yang digunakan tidak sesuai. Menurut Swan (2006) beberapa pilihan WCM yang dapat digunakan adalah *Independent Structure* yaitu berupa matriks identitas, *Exchangeable Structure* merupakan matriks dengan nilai korelasi tetap, *Autoregressive Structure* merupakan matriks dengan bentuk sebagai berikut

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{N-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{N-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{N-1} & \alpha^{N-2} & \dots & \alpha & 1 \end{bmatrix} \quad (1)$$

dengan: $\alpha_t = \phi \sum_{i=1}^N e_{it} e'_{i,t+1} / (NT - K)$ dan $\alpha = \sum_{t=1}^{T-1} \hat{\rho}_t / (T - 1)$

dan *Unstructure Correlation* merupakan struktur korelasi yang digunakan ketika tidak ada urutan yang logis pada pengamatan, dan struktur ini sebaiknya digunakan dalam pengamatan kecil dengan sebuah desain yang seimbang (*balance*).

Secara umum model GLM yang digunakan untuk data longitudinal atau *clustered* adalah sebagai berikut.

$$g(Y_{ij}) = X'_{ij} \beta \quad (2)$$

Untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, m_i$. Dengan m_j banyak pengukuran berulang unit i , n adalah banyak unit pengamatan, π_{ij} adalah peluang sukses unit i pada waktu ke j dan β adalah koefisien regresi. Fungsi penghubung (*Link Function*) yang digunakan bergantung pada tipe variable respon.

Untuk menaksir parameter pada Persamaan (2) harus memperhitungkan adanya WCM sehingga penaksiran parameternya tidak dapat menggunakan metode *Maximum Likelihood*. Metode yang digunakan adalah dengan melakukan proses memaksimumkan fungsi *log-likelihood* berbasis distribusi normal meskipun tanpa dipenuhinya asumsi normalitas dari respon. Fungsi likelihood yang dihasilkan disebut *quasi-likelihood*. GEE memodifikasi fungsi *log-likelihood* dengan memasukkan WCM ke dalam persamaan yang memperhitungkan adanya korelasi dalam data (Hardin dan Hilbe, 2003). Persamaan *quasi-score likelihood* dinyatakan sebagai:

$$U(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \omega_{ij} D'_{ij} V_{ij}^{-1} S_{ij} = 0 \quad (3)$$

Dengan : $D_i = \frac{\partial \pi_{ij}}{\partial \beta} = A_{ij} X_{ij}$ dan $V_{ij} = A_{ij}^2 R(\alpha) A_{ij}^2$

$$A_{ij} = \text{diag}\{\pi_{ij}(1 - \pi_{ij})\} \text{ dan } S_{ij} = Y_{ij} - \pi_{ij}$$

Dengan $R(\alpha)$ adalah WCM dari Y_{i1}, \dots, Y_{ij} , struktur korelasi. Apabila Persamaan (3) diturunkan terhadap β maka akan menghasilkan matriks informasi seperti pada persamaan berikut ini:

$$I = \sum_{i=1}^n \sum_{j=1}^{m_i} \omega_{ij} D'_{ij} V_{ij}^{-1} D_{ij} \quad (4)$$

Dengan :

$$\hat{V} = \frac{n}{n-1} \sum_{i=1}^n (B_i - \bar{B})(B_i - \bar{B})'$$

$$B_i = \sum_{j=1}^{m_i} \omega_{ij} X'_{ij} \hat{A}_{ij}^{1/2} \hat{R}(\alpha)^{-1} \hat{A}_{ij}^{1/2} X_{ij} \quad \text{dan} \quad \bar{B} = \frac{1}{n} \sum_{i=1}^n B_i$$

Persamaan (1) tidak linier dalam parameter β , sehingga memerlukan metode optimasi untuk memperoleh solusinya. Metode optimasi yang digunakan untuk menghitung $\hat{\beta}$ adalah Metode *Fisher Scoring*. Berdasarkan Dobson (2002) persamaan metode *Fisher's scoring* untuk *quasi-likelihood* adalah sebagai berikut :

$$\beta^{(t+1)} = \beta^{(t)} - \left(\sum_{i,j} \omega_{ij} D_{ij}^{(t)'} V_{ij}^{(t)-1} D_{ij}^{(t)} \right)^{-1} \left(\sum_{i,j} \omega_{ij} D_{ij}^{(t)'} V_{ij}^{(t)-1} (Y_{ij} - \pi_{ij}^{(t)}) \right) \quad (5)$$

Proses diiterasi hingga tercapai konvergen. Untuk proses iterasi nilai taksiran awal $\hat{\beta}^{(0)}$ diperoleh dari taksiran menggunakan model GLM biasa menggunakan *Maximum Likelihood Estimation* tanpa melibatkan unsur korelasi. Kemudian pada tahap berikutnya nilai taksiran dihitung dengan melibatkan WCM.

Standard error bagi penaksir GEE tidak dapat diperoleh secara langsung dari *quasi-likelihood*. Beberapa pendekatan yang dapat digunakan adalah menggunakan resampling yaitu metode *jackknife* atau metode *bootstrap*, atau menggunakan *sandwich estimator* yang diusulkan oleh Liang dan Zeger (1986) dengan persamaan sebagai berikut

$$\text{cov}(\hat{\beta}) = (X' \hat{W} X)^{-1} \left[\sum_i X_i' (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)' X_i \right] (X' \hat{W} X)^{-1} \quad (6)$$

Standard error diperoleh dari akar diagonal utama matriks tersebut. Penaksir ini akan baik untuk ukuran sampel yang cukup besar.

Algoritma komputasi untuk GEE secara umum dapat dijelaskan dalam uraian berikut ini: Pertama-tama hitung nilai taksiran awal $\hat{\beta}^{(0)}$ dengan menggunakan model regresi logistik dengan penaksiran parameter menggunakan *Maximum Likelihood Estimation* tanpa melibatkan korelasi. Kemudian tentukan WCM menggunakan struktur korelasi yang telah dijelaskan pada bagian sebelumnya. Selanjutnya hitung nilai korelasibentuk matriks $R(\alpha)$. Kemudian hitung nilai statistik *score quasi-likelihood* pada Persamaan (3), setelah itu hitung juga matriks informasi fisher dari *quasi-likelihood* seperti yang dirumuskan pada Persamaan (4) dengan mengambil nilai $\hat{\beta}^{(0)}$ awal. Kemudian taksir parameter model (1) menggunakan algoritma *Fisher Scoring* seperti pada Persamaan (5). proses iterasi dihentikan jika nilainya sudah konvergen. Selanjutnya tentukan *standard error* dari penaksir tersebut menggunakan Persamaan (6).

Metode penaksiran parameter untuk data longitudinal atau *clustered* menggunakan *generalized estimating equation* yang diperkenalkan oleh Liang dan Zeger (1986) diimplementasikan menggunakan perangkat lunak **Stata 11** dengan paket "**xtgee**". Secara umum ada tiga komponen yang dispesifikasi dalam "**xtgee**" yaitu:

- Distribusi komponen acak: dalam hal ini adalah tipe distribusi *family exponential* dari variabel respon yang dispesifikasi menggunakan opsi "**family()**".
- Fungsi penghubung yang digunakan, sesuai dengan distribusi variable responnya. Fungsi penghubung ini dispesifikasikan menggunakan opsi "**link()**".
- Struktur korelasi: dalam hal ini menggunakan WCM yang dispesifikasikan menggunakan opsi "**correlation ()**"

Secara umum opsi "**family()**" dan "**link()**" tidak diperlukan, sebab secara default telah ditetapkan fungsi penghubung yang digunakan adalah *canonical link*.

III. HASIL DAN PEMBAHASAN

Untuk mengilustrasikan algoritma komputasi GEE menggunakan perangkat lunak Stata 11, digunakan data mengenai kasus DBD di Kota Bandung berdasarkan Tantular et. al. (2016). Data yang digunakan adalah data sekunder yang bersumber dari Dinas Kesehatan Kota Bandung pada 30 Kecamatan dari tahun 2009 hingga 2013. Variabel yang diteliti adalah jumlah kasus Demam Berdarah Dengue (DBD) sebagai respon, sedangkan yang menjadi predictor adalah curah hujan, kelembaban,

kepadatan, angka bebas jentik (ABJ) dan rumah sehat di setiap kecamatan. Data jumlah kasus DBD yang diperoleh oleh Dinas Kesehatan Kota Bandung berasal dari rumah sakit yang ada di Kota Bandung. Puskesmas sebagai instansi kesehatan terkecil hanya bersifat pendataan dalam penanganan penderita DBD. Penderita dirujuk ke rumah sakit terdekat untuk penanganan kesehatan secepatnya. Oleh karena itu, jumlah kasus DBD yang digunakan dalam penelitian ini bersumber dari data rumah sakit yang ada di Kota Bandung dengan mengasumsikan dapat merepresentasikan jumlah kasus DBD di setiap kecamatan.

Secara umum karena respon yang digunakan adalah data cacahan (*counting data*) maka opsi “**family(poisson)**” yang dapat digunakan. Begitu pula dengan fungsi penghubung menggunakan opsi “**link(log)**”. Sementara untuk struktur korelasi dalam hal ini digunakan opsi “**correlation(AR(1))**”. Langkah pertama yang harus dilakukan adalah mempersiapkan data agar siap dianalisis. Pada tahap ini, data yang digunakan adalah data longitudinal maka harus dispesifikasi variable mana yang dijadikan sebagai variable waktu dan variable individu kemudian dihitung statistik deskriptif dari respon.

```
. xtset id_kec tahun
panel variable:  id_kec (strongly balanced)
time variable:  tahun, 2009 to 2012
delta: 1 unit

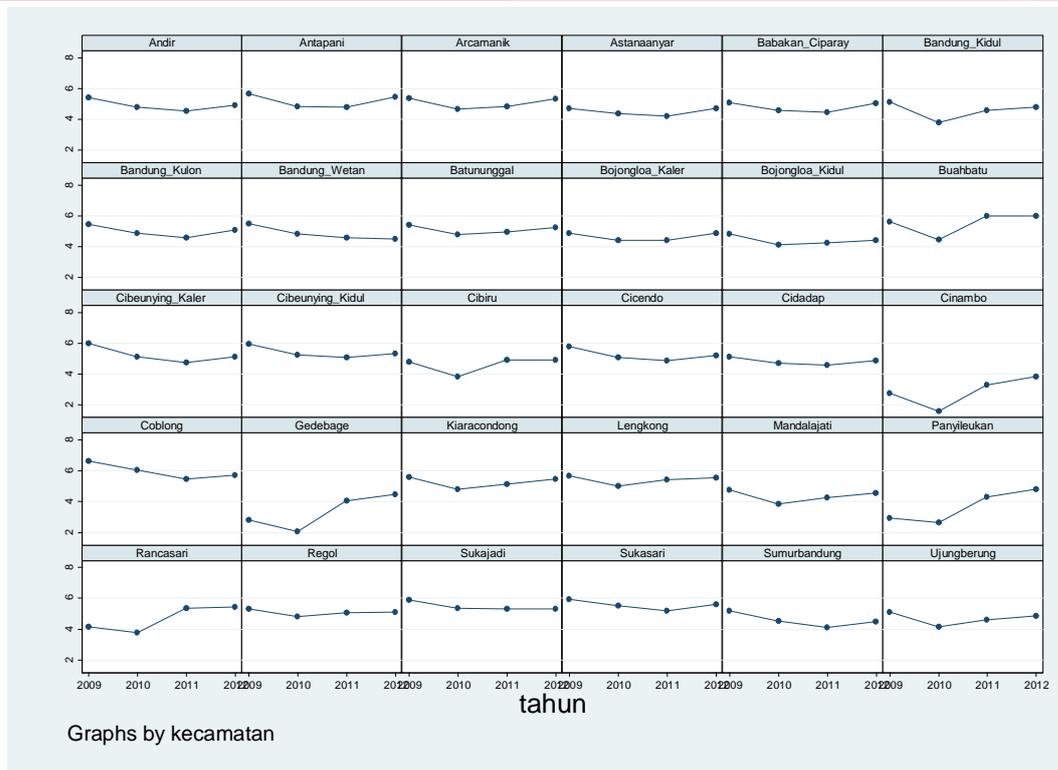
. summarize kasus
```

Variable	Obs	Mean	Std. Dev.	Min	Max
kasus	120	159.2417	107.8154	5	770

Berdasarkan hasil tersebut terlihat bahwa ada kecenderungan bahwa rata-rata respon memiliki nilai yang berbeda dengan varians respon. Hal ini mengindikasikan adanya *overdispersion* pada data tersebut.

Kemudian untuk melihat profil dari setiap unit respon dibuat grafik *individual response profiles*. Untuk respon cacahan gunakan transformasi *log* pada respon, karena fungsi penghubung yang digunakan adalah *log*.

```
. gen log_kasus=log(kasus)
. twoway connect log_kasus tahun, by(kecamatan, style(compact))
yttitle("Jumlah Kasus")
```



Gambar 1. Grafik Profil Respon Kasus DBD Kota Bandung

Dari Gambar 1 terlihat bahwa pola yang terbentuk secara umum relatif mendatar, akan tetapi terdapat perbedaan di beberapa kecamatan yang berfluktuasi dan ada kecenderungan meningkat. Selanjutnya karena ada indikasi overdispersion pada data maka dalam penaksirannya parameter dispersion harus ditaksir. Dalam Stata 11 “**opsi scale(x2)**” dapat digunakan untuk mengatasi masalah tersebut.

```
. xtgee kasus suhu curahhujan kelembaban kepadatan abj rumahsehat, corr(ar1)
family(pois) scale(x2)

Iteration 1: tolerance = .06011716
Iteration 2: tolerance = .00047667
Iteration 3: tolerance = .00013205
Iteration 4: tolerance = .00001084
Iteration 5: tolerance = 6.411e-07

GEE population-averaged model
Group and time vars:      id_kec tahun
Link:                      log
Family:                    Poisson
Correlation:              AR(1)
Scale parameter:          55.24384

Number of obs      =      120
Number of groups   =       30
Obs per group: min =        4
                  avg =       4.0
                  max =        4
Wald chi2(6)      =       64.82
Prob > chi2       =       0.0000
```

kasus	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
suhu	10.40196	1.626916	6.39	0.000	7.213259 13.59065
curahhujan	-.0038729	.0009742	-3.98	0.000	-.0057822 -.0019635
kelembaban	.274426	.0499039	5.50	0.000	.1766161 .3722358
kepadatan	.0000129	9.62e-06	1.34	0.180	-5.96e-06 .0000318
abj	.0042407	.0121585	0.35	0.727	-.0195895 .0280709
rumahsehat	.0010807	.0032255	0.34	0.738	-.0052412 .0074026
_cons	-259.2735	41.58515	-6.23	0.000	-340.7789 -177.7681

(Standard errors scaled using square root of Pearson χ^2 -based dispersion)

Dari hasil diatas terlihat bahwa nilai dispersi sebesar 55.24384 menunjukkan adanya overdispersi yang sangat kuat. Apabila kondisi ini dibiarkan maka *standard error* penaksir akan menjadi sangat besar, hal ini akan mengakibatkan pengujian hipotesis cenderung ditolak. Secara umum meskipun tidak ada dugaan *overdispersion*, penaksiran *standard error* dapat dilakukan dengan menggunakan penaksir *robust*. Opsi “**robust**” bias ditambahkan apabila kita menginginkan standard error ditaksir oleh penaksir *robust*.

```
. xtgee kasus suhu curahhujan kelembaban kepadatan abj rumahsehat, corr(ar1)
family(pois) robust

Iteration 1: tolerance = .06011716
Iteration 2: tolerance = .00047667
Iteration 3: tolerance = .00013205
Iteration 4: tolerance = .00001084
Iteration 5: tolerance = 6.411e-07

GEE population-averaged model
Group and time vars:      id_kec tahun
Link:                     log
Family:                   Poisson
Correlation:              AR(1)
Scale parameter:         1

Number of obs      =      120
Number of groups   =       30
Obs per group: min =       4
                  avg =      4.0
                  max =       4
Wald chi2(6)       =     485.34
Prob > chi2        =      0.0000
```

(Std. Err. adjusted for clustering on id_kec)

kasus	Semirobust		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
suhu	10.40196	1.255709	8.28	0.000	7.940812	12.8631
curahhujan	-.0038729	.0005779	-6.70	0.000	-.0050056	-.0027402
kelembaban	.274426	.0507114	5.41	0.000	.1750334	.3738186
kepadatan	.0000129	.0000117	1.10	0.269	-9.99e-06	.0000358
abj	.0042407	.0111393	0.38	0.703	-.017592	.0260734
rumahsehat	.0010807	.002153	0.50	0.616	-.003139	.0053004
_cons	-259.2735	33.2926	-7.79	0.000	-324.5258	-194.0212

Analisis menggunakan penaksir *standard error robust* meskipun tanpa melibatkan parameter dispersion memiliki hasil yang mirip dengan analisis dengan memperhitungkan adanya parameter *dispersion*. Selanjutnya penaksir struktur korelasi yang digunakan yaitu AR(1) dapat dihitung menggunakan fungsi “**xtcorr**”.

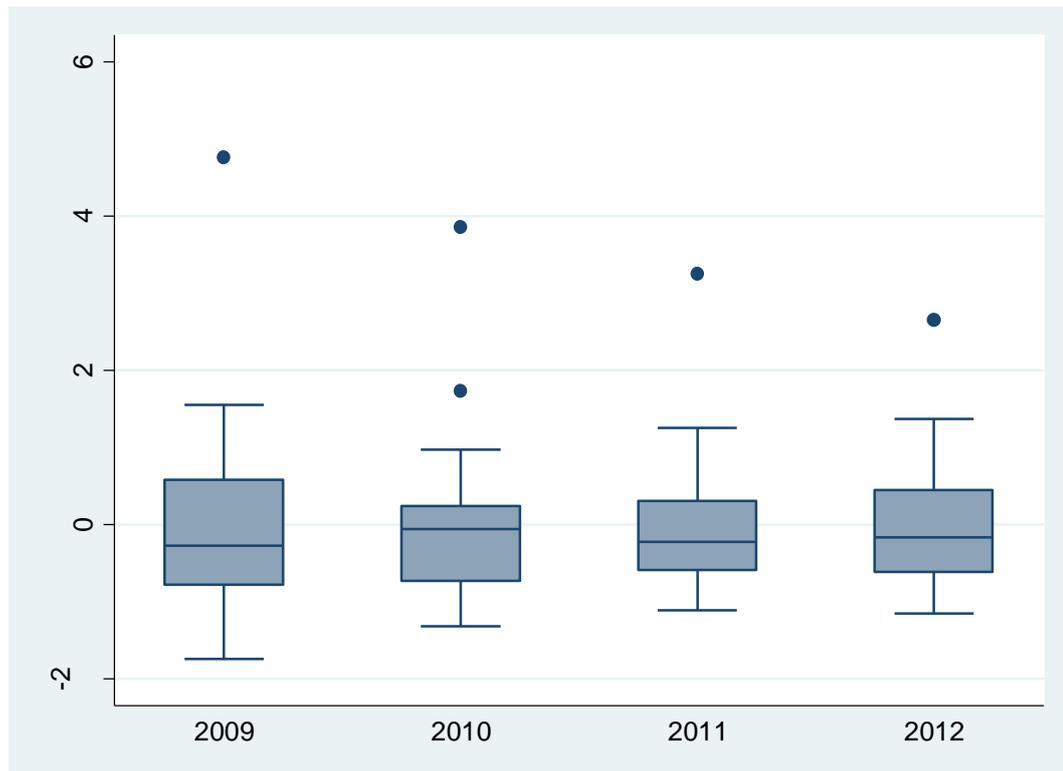
```
. xtcorr

Estimated within-id_kec correlation matrix R:

c1      c2      c3      c4
r1  1.0000
r2  0.7407  1.0000
r3  0.5486  0.7407  1.0000
r4  0.4064  0.5486  0.7407  1.0000
```

Selanjutnya untuk proses *diagnostic* dari residual dapat dilakukan menggunakan *standardized pearson residuals*. Nilai residual yang diperoleh dapat di plot menggunakan boxplot untuk melihat performa dari prediksi ini.

```
. predict pred, mu
. gen pres=(kasus-pred)/sqrt(e(chi2_dis)*pred)
. graph box pres, medtype(line) over(tahun) ytitle("Nilai Prediksi")
```



Gambar 2. *Standardized Pearson Residuals* Kasus DBD Kota Bandung

Berdasarkan Gambar 2 dapat diketahui pengamatan mana saja yang nilainya merupakan *outlier*.

```
. list kasus kecamatan if pres>2
```

```

+-----+
| kasus  kecamatan |
+-----+
47. |    396   Buahbatu |
48. |    407   Buahbatu |
73. |    770   Coblong  |
74. |    425   Coblong  |
+-----+

```

Dari hasil itu terlihat bahwa ada dua kecamatan yang memiliki nilai yang cukup ekstrim yaitu kecamatan Buahbatu dan Kecamatan Coblong.

IV. SIMPULAN DAN SARAN

Penaksir GEE yang diusulkan oleh Liang dan Zeger (1986) merupakan penaksir yang digunakan untuk respon dari data *clustered* atau data longitudinal. Metode GEE memasukkan unsur korelasi antar pengamatan, yang disebut *working correlation matrix*, dalam proses penaksirannya. Metode GEE dapat digunakan pada seluruh model *generalized linear model* yang mengakomodir korelasi antar pengamatan. Algoritma komputasi yang digunakan GEE memiliki proses yang mirip dengan GLM yaitu menggunakan algoritma *Fisher's Scoring*. Akan tetapi *statistic score* dan matriks informasi yang digunakan berasal dari *quasi-likelihood*.

Data kasus DBD Kota Bandung yang memiliki respon cacahan digunakan untuk implementasi algoritma GEE menggunakan Stata 11. Dari hasil analisis dapat disimpulkan bahwa data terindikasi *overdispersion* yang dapat mengakibatkan *standard error* penaksir menjadi sangat besar yang akhirnya

pengujian hipotesis akan cenderung ditolak. Hal ini dapat diatasi dengan memasukkan parameter *dispersion* kedalam analisis. Tetapi apabila hanya ingin menghitung *standard error* yang tidak terpengaruh adanya *dispersion* dapat menggunakan *robust standard error*. Berdasarkan hasil analisis *robust standard error* menghasilkan nilai yang mirip dengan *standard error* dari analisis menggunakan parameter *dispersion*. Untuk diagnostic residual dapat digunakan *Standardized Pearson's Residual*. Hasil analisis *Standardized Pearson's Residual* untuk kasus DBD di Kota Bandung memperlihatkan adanya *outlier* yaitu kecamatan yang memiliki nilai ekstrim yaitu Kecamatan Buahbatu dan Kecamatan Cobleng.

Dalam makalah ini *working correlation matrix* yang digunakan merupakan adalah *autoregressif structure* dengan lag 1, sehingga untuk penelitian selanjutnya disarankan untuk menggunakan *working correlation matrix* yang lainnya seperti *exchangeable* atau *unstructure*. Selain itu dapat pula dibandingkan hasil dari penggunaan berbagai *working correlation matrix* untuk membuktikan sifat *robust* dari penaksir GEE. *Sandwich standard error* merupakan *standard error* alternatif untuk penaksir GEE, sehingga untuk penelitian selanjutnya disarankan untuk melihat performa dari *sandwich standard error* ini.

UCAPAN TERIMAKASIH

Ucapan terima kasih kami sampaikan kepada semua pihak yang membantu dalam penulisan artikel ini.

DAFTAR PUSTAKA

- [1] Alan Agresti. 2002. *Categorical Data Analysis. 2nd edition*. New York: John Wiley & Sons
- [2] Dobson, Annette J. 2002. *An Introduction to Generalized Linear Models 2nd edition*. London. Chapman & Hall.
- [3] Ghisletta, Paolo, and Spini, Dario. 2004. "An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals". *Journal of Educational and Behavioral Statistics Winter*, Vol.29, No.4, pp 421 – 437. University of Geneva and University of Laussane.
- [4] Hardin, J. W., and Hilbe, J. M. 2003. *Generalized Estimating Equations*. United States of America: Chapman & Hall/CRC.
- [5] Liang, K.-Y. and Zeger S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, Vol 73, pp. 45-51.
- [6] McCullagh dan Nelder. 1989. *Generalized Linear Models. 2nd edition*. , London. Chapman & Hall.
- [7] Paul, Sudhir., Zhang, Xuemao., Xu, Jianwen. 2013. Estimation of Regression Parameters for Binary Longitudinal Data Using GEE: Review, Extension and an Application to Environmental Data. *Journal of Environmental Statistics*, Vol 4, Issue 1.
- [8] Rabe-Hesketh, Sophia., Everitt, Brian. 2004. *A Handbook of Statistical Analysis using Stata. 3rd edition*, Washington D.C., Chapman & Hall.
- [9] Swan, Taryn. 2006. "Generalized Estimating Equation when the response variable has a Tweedie distribution: An application for multi-site rainfall modeling". *Thesis* in The University of Southern Queensland, Toowoomba.
- [10] Tantular, Bertho. 2012. Pendekatan Model Multilevel untuk Data Repeated Measures. Paper dipresentasikan pada Seminar Nasional Matematika dan Pendidikan FMIPA UNY, 10 November 2012.
- [11] Tantular, Bertho, Mindra Jaya , I Gede Nyoman, Zulhanif. 2016. Pendekatan Model Multilevel pada Analisis Regresi Poisson untuk Data Longitudinal (Studi Kasus DBD Kota Bandung). *Prosiding Seminar Nasional Matematika dan Pendidikan Jurusan Pendidikan Matematika FMIPA UNY*, ISBN: 978-602-73403-1-2.