



KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI  
UNIVERSITAS LAMPUNG  
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT  
GEDUNG REKTORAT LANTAI 5  
Jalan. Prof. Dr. Sumantri Brojonegoro No.1 Bandar Lampung 35145  
Telp. (0721) 705173, 701609 Ext. 136 Fax. 773798 email: lppm@kpa.Unila.ac.id

**SURAT PENUGASAN  
PENELITIAN HIBAH FUNDAMENTAL BARU  
TAHUN ANGGARAN 2016**

**Nomor : 76 /UN26/8/LPPM/2016**

Pada hari ini Rabu tanggal Tiga Belas bulan April tahun Dua Ribu Enam Belas, saya yang bertandatangan dibawah ini :

Nama : Warsono, Ph.D.  
NIP : 196302161987031003  
Jabatan : Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM)  
Universitas Lampung  
Alamat : Jln. Prof. Soemanteri Brojonegoro No.1 Gedung Meneng Bandar Lampung  
**Yang selanjutnya disebut sebagai Pihak I (Pertama)**

Dengan ini menugaskan kepada peneliti :

Nama : Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP : 196501251990032001  
Jabatan : Ketua Peneliti  
Fakultas : MIPA  
**Yang selanjutnya disebut sebagai Pihak II (Kedua)**

Untuk melakukan tugas Penelitian HIBAH FUNDAMENTAL BARU yang didanai oleh Dana DIKTI Tahun Anggaran 2016, dengan judul: ***Pengembangan Metode Iterated Reweighted Least Trimmed Square untuk Pendugaan Model Generalized Estimating Equation (GEE) pada Data Mengandung Pencilan***

Surat Penugasan Penelitian ini didasari oleh :

1. Peraturan Pemerintah Republik Indonesia No. 19 Tahun 2005 tentang Standar Nasional Pendidikan; Peraturan Pemerintah Nomor 23 Tahun 2005, Tentang Pengelolaan Keuangan badan Layanan Umum;
2. Keputusan Presiden Nomor 73 tahun 1966 tentang pendirian Unila;
3. Keputusan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 335/MIKP/XI/2015 Tentang Pemberhentian dan Pengangkatan Rektor Unila;
4. Peraturan Menteri Keuangan Nomor 129/KMK.05/2009 Tentang Penetapan Unila Pada Departemen Pendidikan Nasional sebagai Instansi Pemerintah yang Menerapkan Pengelolaan Keuangan Badan Layanan Umum;
5. Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 6 Tahun 2015 tentang Statuta Universitas Lampung;
6. Keputusan Menteri Pendidikan dan Kebudayaan RI Nomor 72 Tahun 2014 tentang Organisasi dan Tata Kerja Universitas Lampung;
7. Keputusan Rektor Universitas Lampung Nomor 18/UN26/OT/2015 tentang berdirinya LPPM Unila;
8. Keputusan Rektor Universitas Lampung Nomor 25/UN26/KP/2015 tentang Pengangkatan Ketua dan Sekretaris Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Lampung;

## KEWAJIBAN – KEWAJIBAN YANG HARUS DIPENUHI

### PASAL 1

Pelaksanaan penugasan penelitian sebagaimana dimaksud, nilai penugasan penelitiannya adalah sebesar Rp 50000000,- (Lima Puluh Juta Rupiah) yang dananya bersumber dari dana APBN Universitas Lampung Tahun Anggaran 2016.

### PASAL 2

Pembayaran penugasan penelitian ini dilaksanakan dalam 2 (dua) Termin yaitu :

- (1) Pembayaran Termin Pertama, sebesar 70% dari nilai pekerjaan = Rp. 35000000,- (Tiga Puluh Lima Juta Rupiah) dibayarkan setelah surat perjanjian ini ditanda tangani oleh kedua belah pihak.
- (2) Pembayaran Termin Kedua, sebesar 30% dari nilai pekerjaan = Rp. 15000000,- (Lima Belas Juta Rupiah) dibayarkan setelah Peneliti menyerahkan laporan Akhir dan laporan keuangan Hasil Pelaksanaan Penelitian yang telah dilaksanakan kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung, disertai dengan Berita Acara Serah Terima Laporan Akhir dan Surat Pertanggung Jawaban Mutlak.

### PASAL 3

Hal-hal dan segala sesuatu yang berkenaan dengan kewajiban pajak berupa PPN dan/atau PPh menjadi tanggung jawab PIHAK KEDUA dan harus dibayarkan ke Kas Negara sesuai dengan ketentuan peraturan perundang-undangan yang berlaku.

### PASAL 4

- (1) Peneliti melaksanakan penelitian sesuai dengan proposal yang telah disetujui oleh Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung Tahun 2016.
- (2) Peneliti berkewajiban untuk mengupayakan hasil penelitiannya untuk dapat dipublikasikan baik dalam jurnal Ilmiah di lingkungan Universitas Lampung maupun diluar Universitas Lampung.

### PASAL 5

- (1) Dana penelitian yang diperoleh oleh peneliti dimanfaatkan sebenar-benarnya untuk pembiayaan penelitian yang dilaporkan kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung.
- (2) Perubahan-Perubahan dalam pelaksanaan penelitian dapat dibenarkan apabila telah mendapat persetujuan lebih dahulu dari Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung.

## PASAL 6

- (1) Peneliti harus menyelesaikan Penelitian yang dimaksud dan menyerahkan : Laporan kemajuan dan Laporan penggunaan anggaran 70% yang telah diverifikasi oleh Tim Money Internal LPPM selambat-lambatnya tanggal 18 Juli 2016
- (2) Peneliti harus menyerahkan : Laporan Akhir dan Laporan Keuangan yang telah diverifikasi oleh Tim Money Internal LPPM, selambat-lambatnya tanggal 28 Oktober 2016.
- (3) Laporan sebagaimana dimaksud dalam pasal 6 ayat (1) disampaikan dalam bentuk hard copy (sebanyak 2 eksemplar) dan softcopy (sebanyak 2 keping CD).
- (4) Bentuk/ukuran, format penulisan dan warna cover Sesuai dengan panduan yang telah ditetapkan.

## PASAL 7

- (1) Apabila Peneliti (ketua) sebagaimana dimaksud pada Pasal 1 tidak dapat menyelesaikan penelitian ini, maka peneliti wajib menunjuk pengganti ketua pelaksana sesuai dengan bidang ilmu yang diteliti dan merupakan salah satu anggota tim yang diketahui oleh Dekan Fakultas dan disetujui oleh Lembaga Penelitian dan Pengabdian kepada Masyarakat Unila.
- (2) Apabila batas waktu penelitian habis peneliti belum menyerahkan hasil pekerjaan seluruhnya maka peneliti akan dikenakan denda sebesar 1 o/oo (satu permil) setiap hari keterlambatan sampai dengan setinggi-tingginya 5% (lima persen) dari nilai kontrak.
- (3) Bagi peneliti yang tidak dapat menyerahkan laporan hasil penelitian sampai dengan akhir tahun anggaran yang sedang berjalan maka sisa dana penelitian yang bersangkutan dinyatakan hangus dan peneliti wajib mengembalikan dana penelitian yang sudah dicairkan untuk dikembalikan ke Kas Negara;
- (4) Apabila Peneliti tidak dapat memenuhi pasal-pasal sebagaimana diatur dalam Perjanjian Penugasan Penelitian ini, maka Peneliti wajib mengembalikan seluruh dana penelitian yang telah diterimanya kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung untuk selanjutnya disetorkan ke Kas Negara;
- (5) Apabila dikemudian hari terbukti bahwa judul-judul penelitian sebagaimana dimaksud pada Pasal 1 dijumpai adanya indikasi duplikasi dengan penelitian lain dan/atau diperoleh indikasi ketidak jujuran dan iktikad kurang baik yang tidak sesuai dengan kaidah ilmiah, maka penelitian tersebut dinyatakan batal dan Peneliti wajib mengembalikan dana penelitian yang telah diterimanya kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung untuk selanjutnya disetor ke Kas Negara.

## PASAL 8

Surat Penugasan Penelitian ini dibuat rangkap 2 (dua), dan masing-masing bermeterai sesuai dengan ketentuan yang berlaku, dan biaya meterainya dibebankan kepada PIHAK KEDUA.

## PASAL 9

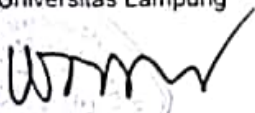
1. Jika terjadi perselisihan antara kedua belah pihak, pada dasarnya akan diselesaikan secara musyawarah.
2. Jika perselisihan itu tidak dapat diselesaikan secara musyawarah, maka akan diselesaikan oleh "Panitia Pendamai" yang berfungsi sebagai juri/wasit, yang dibentuk dan diangkat oleh kedua belah pihak yang terdiri :
  - Seorang wakil dari PIHAK PERTAMA sebagai anggota;
  - Seorang wakil dari PIHAK KEDUA sebagai anggota;
  - Seorang PIHAK KETIGA yang ahli sebagai Ketua, yang telah disetujui oleh PIHAK PERTAMA dan PIHAK KEDUA.
3. Keputusan Panitia Pendamai ini mengikat kedua belah pihak, dan biaya penyelesaian perselisihan yang dikeluarkan akan ditanggung secara bersama.
4. Jika Keputusan ini sebagai mana dimaksud ayat 3 pasal ini tidak dapat diterima oleh salah satu pihak, maka penyelesaian perselisihan akan diteruskan melalui Pengadilan Negeri.
5. Segala akibat yang terjadi dari pelaksanaan perjanjian ini, kedua belah pihak memilih kedudukan (domisili) yang tetap dan sah di Kantor Pengadilan Negeri Bandar Lampung.

## PASAL 10

Segala sesuatu yang belum diatur dalam surat perjanjian ini, atau perubahan-perubahan yang dipandang perlu oleh kedua belah pihak, akan diatur lebih lanjut dalam Surat Perjanjian Tambahan (Adendum) dan merupakan perjanjian yang tidak terpisahkan dari perjanjian ini.

### PIHAK PERTAMA,

Ketua LPPM  
Universitas Lampung

  
Warsono, Ph.D.  
NIP 196302161987031003

### PIHAK KEDUA,

Ketua /Selaku  
Penanggung Jawab Penelitian

  
Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001



## BERITA ACARA PEMBAYARAN

Pada hari ini **Rabu** tanggal **Tiga Belas** bulan **April** tahun **Dua Ribu Enam Belas**, saya yang bertanda tangan dibawah ini :

- I. Nama : Warsono, Ph.D.  
Jabatan : Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung  
Alamat : Jl. Prof. Sumantri Brojonegoro No.1 Gedung Meneng Bandar Lampung  
Disebut Sebagai **PIHAK PERTAMA**.
- II. Nama : Dr., Ir. NETTY HERAWATI, M.Sc.  
Jabatan : Peneliti Utama (penanggung jawab penelitian)  
Fakultas : MIPA  
Alamat : Jl. Prof.Sumantri Brojonegoro No. 1 Bandar Lampung.  
Disebut Sebagai **PIHAK KEDUA**.


Sehubungan dengan pelaksanaan kegiatan Penelitian HIBAH FUNDAMENTAL BARU di Lingkungan Universitas Lampung, sesuai dengan Surat Penugasan Penelitian Nomor. 76/UN26/8/LPPM/2016, tanggal 13 April 2016 dengan judul "**Pengembangan Metode Iterated Reweighted Least Trimmed Square untuk Pendugaan Model Generalized Estimating Equation (GEE) pada Data Mengandung Pencilan**", maka **PIHAK KEDUA** berhak menerima pembayaran dari **PIHAK PERTAMA** sebesar 70% dari nilai kontrak = 70 % x Rp. 50000000,- = Rp. 35000000,- (Tiga Puluh Lima Juta Rupiah) dan disalurkan langsung ke Rekening **PIHAK KEDUA** sebagai Penanggung Jawab Kegiatan Penelitian.

Demikian Berita Acara Pembayaran ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Bandar Lampung, 13 April 2016

### **I. PIHAK PERTAMA.**


Ketua LPPM  
Universitas Lampung.

  
**Warsono, Ph.D.**  
NIP. 196302161987031003

### **II. PIHAK KEDUA.**

Ketua Penelitian/  
nanggung Jawab Kegiatan



  
**Dr., Ir. NETTY HERAWATI, M.Sc.**  
NIP. 196501251990032001



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN REPUBLIK INDONESIA  
UNIVERSITAS LAMPUNG  
LEMBAGA PENELITIAN  
GEDUNG REKTORAT LANTAI 5

Jalan. Prof. Dr. Sumantri Brojonegoro No.1 Bandar Lampung 35145  
Telp. (0721) 705173, 701609 Ext. 136 Fax. 773798 E-Mail: Lemlit@Unila.ac.id

**SURAT PERNYATAAN TANGGUNG JAWAB MUTLAK**

Yang bertanda tangan dibawah ini :

Nama : Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP : 196501251990032001  
Fakultas : MIPA  
Alamat : Jl.Prof.Sumantri Brojonegoro No.1 Gedung Meneng  
Bandar Lampung 35145

Menyatakan dengan sesungguhnya bahwa :

1. Dana penelitian yang saya terima sudah dihitung dengan benar dan akan digunakan sepenuhnya untuk mendanai penelitian yang saya laksanakan yaitu penelitian yang didanai oleh Dana DIKTI TA 2016. Jenis Hibah Penelitian HIBAH FUNDAMENTAL BARU Judul Pengembangan Metode Iterated Reweighted Least Trimmed Square untuk Pendugaan Model Generalized Estimating Equation (GEE) pada Data Mengandung Pencilan dengan jumlah dana sebesar 70% dari nilai pekerjaan Rp. 50000000,- yaitu Rp. 35000000,- (Tiga Puluh Lima Juta Rupiah).
2. Semua penggunaan, pengeluaran keuangan dan pertanggungjawabannya yang terkait dengan *output* kegiatan pelaksanaan penelitian menjadi tanggung jawab saya sepenuhnya.

Demikian pernyataan ini saya buat dengan sebenarnya-benarnya.

Bandar Lampung, 13 April 2016



Peneliti,

Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001



### SURAT PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP : 196501251990032001  
Fakultas : MIPA  
Alamat : Jl.Prof.Sumantri Brojonegoro No.1 Gedung Meneng  
Bandar Lampung 35145

Menyatakan dengan sesungguhnya bahwa :

1. Saya tidak menerima dana penelitian ditahun yang sama, dan dari sumber dana manapun.
2. Apabila terbukti saya menerima dana penelitian dari sumberdana yang lain ditahun yang sama, maka penelitian saya ini dibatalkan dan saya bersedia mengembalikan dana penelitian yang sudah saya terima.

Demikian pernyataan ini saya buat dengan sebenarnya-benarnya.

Bandar lampung, 13 April 2016



Netty Herawati,

Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001

# KWITANSI

No 76

Sudah terima dari : Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung

Banyaknya uang : **Tiga Puluh Lima Juta Rupiah**

Untuk pembayaran : Dana Penelitian HIBAH FUNDAMENTAL BARU yang didanai oleh Dana DIKTI T.A. 2016 Tahap I 70 %  
Dari Nilai Pengawasan sebesar Rp. 50000000,- Berdasarkan Surat Pengawasan Penelitian HIBAH FUNDAMENTAL BARU Nomor. 76/JUN26/8/LPPM/2015 Tanggal 13 April 2016

Rp. 35000000,00

Bandar Lampung, 13 April 2016

Yang Menerima,



terai 6000

Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 195501251990032001

No 76

Sudah terima dari : Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung

Banyaknya uang : **Tiga Puluh Lima Juta Rupiah**

Untuk pembayaran : Dana Penelitian HIBAH FUNDAMENTAL BARU yang didanai oleh Dana DIKTI T.A. 2016 Tahap I 70 %  
Dari Nilai Pengawasan sebesar Rp. 50000000,- Berdasarkan Surat Pengawasan Penelitian HIBAH FUNDAMENTAL BARU Nomor. 76/JUN26/8/LPPM/2015 Tanggal 13 April 2016

Rp. 35000000,00

Bandar Lampung, 13 April 2016

Yang Menerima,

Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 195501251990032001



## BERITA ACARA PEMBAYARAN

Pada hari ini **Senin** tanggal **Sembilan Belas** bulan **Desember** tahun **Dua Ribu Enam Belas**, saya yang bertanda tangan dibawah ini :

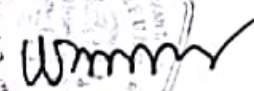
- I. Nama : Warsono, Ph.D.  
Jabatan : Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung  
Alamat : Jl. Prof. Sumantri Brojonegoro No.1 Gedung Meneng Bandar Lampung  
Disebut Sebagai **PIHAK PERTAMA.**
- II. Nama : Dr., Ir. NETTY HERAWATI, M.Sc.  
Jabatan : Peneliti Utama (penanggung jawab penelitian)  
Fakultas : MIPA  
Alamat : Jl. Prof.Sumantri Brojonegoro No. 1 Bandar Lampung.  
Disebut Sebagai **PIHAK KEDUA.**

Sehubungan dengan pelaksanaan kegiatan Penelitian HIBAH FUNDAMENTAL BARU di Lingkungan Universitas Lampung, sesuai dengan Surat Penugasan Penelitian Nomor. 76/UN26/8/LPPM/2016, tanggal 13 April 2016 dengan judul "**Pengembangan Metode Iterated Reweighted Least Trimmed Square untuk Pendugaan Model Generalized Estimating Equation (GEE) pada Data Mengandung Pencilan**", maka **PIHAK KEDUA** berhak menerima pembayaran dari **PIHAK PERTAMA** sebesar 30% dari nilai kontrak = 30 % x Rp. 50000000,- = Rp.15000000,- (Lima Belas Juta Rupiah) dan disalurkan langsung ke Rekening **PIHAK KEDUA** sebagai Penanggung Jawab Kegiatan Penelitian.


Demikian Berita Acara Pembayaran ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Bandar Lampung, 19 Desember 2016

**I. PIHAK PERTAMA.**

Ketua LPPM  
Universitas Lampung,  
  
Warsono, Ph.D.  
NIP. 196302161987031003 #

**II. PIHAK KEDUA.**

Ketua Penelitian/  
Penanggung Jawab Kegiatan  
  
Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001

## SURAT KUASA

Saya yang bertandatangan di bawah ini,

Nama : Dr. Ir. Netti Herawati, M.Sc.

NIP : 196501251990032001

Alamat : Jl. Sukardi Hamdani, Gg Palapa VB No.38 Bandar Lampung

selaku Ketua tim kegiatan penelitian fundamental berjudul "Pengembangan Metode *Iterated Reweighted Least Trimmed Square* untuk Pendugaan Model *Generalized Estimating Equation* (GEE) pada Data Mengandung Pencilan" (No. kontrak 76/ UN26/ LPPM/ 2016), dengan ini memberikan kuasa kepada anggota tim penelitian saya:

Nama : Khoirin Nisa, S.Si., M.Si.

NIP : 197407262000032001

Alamat : Perumahan Griya Kencana Blok C No. 6 Rajabasa Bandar Lampung

untuk mewakili saya dalam pengurusan administrasi pencairan dana penelitian sebesar 30% disebabkan saya akan berada di luar kota untuk suatu keperluan penting yang tidak bisa ditinggalkan.

Demikianlah surat kuasa ini saya buat dengan sungguh-sungguh dan kesadaran untuk digunakan sebagaimana mestinya.

Bandar Lampung, 17 Desember 2016



Dr. Ir. Netti Herawati, M.Sc.



## SURAT PERNYATAAN TANGGUNG JAWAB MUTLAK

Yang bertanda tangan dibawah ini :

Nama : Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP : 196501251990032001  
Fakultas : MIPA  
Alamat : Jl.Prof.Sumantri Brojonegoro No.1 Gedung Meneng  
Bandar Lampung 35145

Menyatakan dengan sesungguhnya bahwa :

1. Dana penelitian yang saya terima sudah dihitung dengan benar dan akan digunakan sepenuhnya untuk mendanai penelitian yang saya laksanakan yaitu penelitian yang didanai oleh Dana DIKTI TA 2016. Jenis Hibah Penelitian HIBAH FUNDAMENTAL BARU Judul Pengembangan Metode Iterated Reweighted Least Trimmed Square untuk Pendugaan Model Generalized Estimating Equation (GEE) pada Data Mengandung Pencilan dengan jumlah dana sebesar 30% dari nilai pekerjaan Rp. 50000000,- yaitu Rp.15000000,- (Lima Belas Juta Rupiah).
2. Semua penggunaan, pengeluaran keuangan dan pertanggungjawabannya yang terkait dengan *output* kegiatan pelaksanaan penelitian menjadi tanggung jawab saya sepenuhnya.

Demikian pernyataan ini saya buat dengan sebenarnya-benarnya.

Bandar Lampung, 19 Desember 2016



Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001

# KWITANSI

Sudah terima dari : Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung

Banyaknya uang : Lima Belas Juta Rupiah

Untuk pembayaran : Dana Penelitian HIBAH FUNDAMENTAL BARU yang didanai oleh Dana DIKTI T.A. 2016 Tahap II 30 %  
: Dari Nilai Penugasan sebesar Rp. 50000000,- Berdasarkan Surat Penugasan Penelitian HIBAH FUNDAMENTAL BARU Nomor: 76/JUN26/8/LPPM/2015 Tanggal 13 April 2016

Rp. 15000000,00

Bandar Lampung, 19 Desember 2016  
Yang Menerima,



Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001

No 76

# KWITANSI

Sudah terima dari : Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Lampung

Banyaknya uang : Lima Belas Juta Rupiah

Untuk pembayaran : Dana Penelitian HIBAH FUNDAMENTAL BARU yang didanai oleh Dana DIKTI T.A. 2016 Tahap II 30 %  
: Dari Nilai Penugasan sebesar Rp. 50000000,- Berdasarkan Surat Penugasan Penelitian HIBAH FUNDAMENTAL BARU Nomor: 76/JUN26/8/LPPM/2015 Tanggal 13 April 2016

Rp. 15000000,00

Bandar Lampung, 19 Desember 2016  
Yang Menerima,

Dr., Ir. NETTY HERAWATI, M.Sc.  
NIP. 196501251990032001

Kode>Nama Rumpun Ilmu : 122 / Statistika

LAPORAN AKHIR  
PENELITIAN FUNDAMENTAL



JUDUL PENELITIAN

**Pengembangan Metode *Iterated Reweighted Least Trimmed Square* untuk  
Pendugaan Model *Generalized Estimating Equation* (GEE) pada Data  
Mengandung Pencilan**

TIM PENELITI

**Dr. Ir. Netti Herawati, M.Sc.                      NIDN 0025016503**

**Khoirin Nisa, S.Si., M.Si                         NIDN 0026077401**

UNIVERSITAS LAMPUNG

2016

**HALAMAN PENGESAHAN  
PENELITIAN FUNDAMENTAL**

**Judul Penelitian** : Pengembangan Metode *Iterated Reweighted Least Trimmed Square* untuk Pendugaan Model *Generalized Estimating Equation (GEE)* pada Data Mengandung Pencilan

**Kode>Nama Rumpun Ilmu** : 122 / Statistika

**Ketua Peneliti**

a. Nama Lengkap : Dr. Ir. Netti Herawati, M.Sc.  
b. NIDN : 0025016503  
c. Jabatan Fungsional : Lektor Kepala  
d. Program Studi : Matematika  
e. Nomor HP : 085769675537  
f. Alamat surel (e-mail) : netti.herawati@fmipa.unila.ac.id

**Anggota Peneliti**

a. Nama Lengkap : Khoirin Nisa, S.Si.,M.Si.  
b. NIDN : 0026077401  
c. Perguruan Tinggi : Universitas Lampung

**Lama Penelitian Keseluruhan** : 2 (dua) tahun

**Penelitian Tahun ke** : 1 (satu)

**Biaya Penelitian Keseluruhan** : Rp. 110.230.000

**Biaya Tahun Berjalan** : - diusulkan ke DIKTI Rp. 58.318.000  
- dana internal PT Rp. 0.00  
- dana institusi lain Rp. 0.00  
- *inkind* sebutkan .....


Bandar Lampung, 22 November 2016

Ketua Peneliti,

Mengetahui,  
Ketua Fakultas MIPA  
Universitas Lampung  
  
Ir. Warsito, S.Si., D.E.A., Ph.D.)  
NIP. 19710212 199512 1 001

  
(Dr. Ir. Netti Herawati, M.Sc.)  
NIP. 19650125 199003 2 001

Menyetujui,  
Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat  
Universitas Lampung

  
(Ir. Warsono, M.S., Ph.D.)  
NIP. 196302161987031003

## DAFTAR ISI

BAB 1. PENDAHULUAN .....	1
BAB 2. TINJAUAN PUSTAKA .....	3
BAB 3. TUJUAN DAN MANFAAT PENELITIAN.....	10
BAB 4. METODE PENELITIAN .....	11
BAB 5. HASIL YANG DICAPAI.....	13
4.1. Algoritma IRLTS.....	13
4.2. Simulasi .....	15
4.3. Pembahasan .....	29
BAB 6. RENCANA TAHAPAN BERIKUTNYA.....	41
BAB 7. KESIMPULAN DAN SARAN.....	42
DAFTAR PUSTAKA .....	43
LAMPIRAN – LAMPIRAN	
Lampiran 1. Paper ilmiah disajikan pada International Conference on Sciences, Technology and Interdisciplinary Research (ICSTAR) 2016.....	48
Lampiran 2. Paper ilmiah disubmit ke International Journal of Pure and Applied Mathematics.....	57
Lampiran 3. R codes untuk simulasi Monte Carlo .....	65
Lampiran 4. Contoh output analisis GEE pada software R.....	73
Lampiran 5. Histogram data Y untuk model GEE sebelum dan sesudah konversi dengan teknik NORTA.....	77

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Keberadaan data pencilan (*outlier*) akan mengganggu dalam proses analisis data dan harus dihindari. Pencilan dalam data menunjukkan adanya pengamatan yang berasal dari kelompok atau distribusi yang berbeda dengan mayoritas data lainnya. Pencilan dapat juga disebabkan oleh kesalahan manusia seperti kesalahan mencatat pengamatan atau kesalahan ketika menyiapkan peralatan pengukuran. Namun, membuang pencilan dalam suatu gugus data bukanlah prosedur yang tepat karena berbagai hal, diantaranya karena pencilan adakalanya dapat memberikan informasi yang tidak bisa diberikan oleh pengamatan lain. Selain itu, pada kasus data berukuran besar, mendeteksi pencilan bukanlah prosedur analisis yang sederhana.

Model linier tergeneralisir (*generalized linear models* : GLM) yang dikembangkan oleh McCullagh dan Nelder (1989) adalah suatu teknik statistika untuk analisis regresi data univariat. Kelebihan GLM dibandingkan dengan analisis regresi linier konvensional adalah bahwa data yang digunakan dalam GLM tidak harus berdistribusi normal, ia dapat mengikuti distribusi-distribusi lain yang termasuk dalam keluarga eksponensial seperti distribusi gamma, beta, Poisson, binomial, log normal, khi-kuadrat dan sebagainya. Hubungan antara variabel respon dan kovariat dalam GLM dijelaskan oleh suatu fungsi penghubung (*link function*). Seperti halnya analisis regresi konvensional, GLM mengasumsikan bahwa masing-masing pengamatan dalam data saling bebas dan sama sekali tidak mempertimbangkan adanya korelasi antara  $n$  respon pengamatan dalam data. Jika terdapat korelasi dalam data maka hal ini dianggap sebagai penyimpangan yang akan mempengaruhi hasil dugaan GLM.

Liang dan Zeger (1986) memperkenalkan suatu pendekatan terhadap masalah data berkorelasi dengan menggunakan *generalized estimating equations* (GEE) untuk mengembangkan GLM menjadi suatu regresi dengan pengamatan-pengamatan yang saling berkorelasi. Selain kelebihan dalam mengatasi data berkorelasi, GEE juga menawarkan suatu framework untuk analisis data kontinu, ordinal, polikotomi, dikotomi dan diskrit. Metode GEE memberikan penduga yang konsisten bagi



parameter-parameter bahkan ketika struktur kovarians dari pengamatan-pengamatan yang berkorelasi salah dipesifikasi (Manel & DeRouen, 2001). Namun demikian, masalah dapat terjadi ketika data yang dianalisa mengandung pencilan (*outlier*), pendugaan GEE yang berbasis pada *quasi likelihood* sangat sensitif terhadap pencilan terutama pencilan ekstrim. Pencilan extreme dalam data akan mengakibatkan pendugaan GEE menjadi bias dan kehilangan efisiensinya (Abebe dkk, 2014). Dalam situasi seperti ini maka diperlukan suatu metode yang ‘*robust*’ agar dihasilkan pendugaan parameter yang tak bias.

Dalam beberapa tahun terakhir ini, beberapa metode GEE *robust* telah diperkenalkan untuk analisis data yang saling berkorelasi. Sebagai contoh, Qaqish dan Preisser (1999) mengajukan sebuah metode GEE *robust* menggunakan pendugaan tipe-M dengan melibatkan penurunan bobot (*downweighting*) titik-titik data yang berpengaruh. Gill (2000) mengajukan metode “*robustified likelihood*” yang didasarkan pada distribusi normal multivariate. Jung dan Ying (2003) mengajukan suatu metode *robust* dengan mengadaptasi metode Wilcoxon-Mann-Whitney untuk digunakan pada analisis data berkorelasi dibawah model kebebasan (*independence model*), metode yang diajukan oleh Jung dan Ying selanjutnya dimodifikasi oleh Wang dan Zhu (2006), dan terakhir Abebe dkk (2014) mengajukan suatu GEE *robust* dengan menggunakan pendugaan berbasis iterasi rank terboboti. Namun pengujian empiris terhadap penduga-penduga GEE *robust* tersebut menunjukkan konsistensinya hanya pada tingkat kontaminasi pencilan sebesar 20%.

Dalam penelitian ini kami akan mengembangkan metode GEE *robust* yang dapat mengatasi tingkat kontaminasi di atas 20%. Metode yang akan kami kembangkan ini berbeda dengan metode-metode yang sudah ada, metode yang kami ajukan didasarkan pada pemangkasan (*trimming*) gugusan kuantil data yang memiliki residual besar dengan mengadopsi algoritma dari metode *least trimmed square* atau LTS (Rousseeuw & Leroy, 1987) yang digunakan untuk analisis regresi konvensional. Metode yang akan kami kembangkan ini kami sebut “*iterated reweighted least trimmed square*” (IRLTS).

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana membangun sebuah algoritma pendugaan model GEE berbasis LTS terboboti beriterasi sehingga menghasilkan pendugaan parameter yang tak bias, efisien dan konsisten walau data mengandung pencilan.
2. Seberapa jauh ketegaran metode IRLTS yang kami kembangkan ini dalam menghadapi banyaknya pencilan pada pendugaan model GEE, untuk itu pengujian empiris terhadap metode IRLTS akan dilakukan dengan mempertimbangkan tingkat kontaminasi pencilan hingga di atas 30%.

## TINJAUAN PUSTAKA

### 2.1 Review GEE

Metode GEE diperkenalkan oleh Liang dan Zeger (1986) sebagai pengembangan metode GLM untuk mengakomodasi korelasi dalam data. Metode ini telah banyak digunakan dalam berbagai bidang khususnya yang berkaitan dengan penelitian longitudinal (Ballinger, 2004; Nakai & Ke, 2009) dan penelitian *cross-section* (Norton dkk, 1996). Berapa contoh penerapan GEE dalam berbagai bidang misalnya dapat dilihat dalam paper-paper berikut: Bishop dkk (2000) dalam bidang teknik kelautan, Manel dan DeRouen (2001) untuk bidang biomedis, Zorn (2001) untuk bidang ilmu politik, Ghisletta & Spini (2004) untuk bidang pendidikan dan Suktitipat dkk (2012) dalam bidang genetika.

GEE awalnya diperkenalkan untuk menduga koefisien regresi dalam GLM pada kondisi di mana struktur kovarian diketahui ataupun tidak diketahui dan harus diduga di awal (yaitu *prior covariance matrix*). Pada pengembangan selanjutnya GEE dapat digunakan untuk menduga parameter-parameter yang menjelaskan struktur kovarian yang tidak diketahui (Prentice & Zhao, 1991). Berikut ini kami uraikan secara singkat teori GEE pada model GLM.

Misalkan  $Y_{ij}$ ,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, K$  menotasikan pengukuran ke- $j$  dari subjek ke- $i$ , sehingga terdapat  $n_i$  kali pengukuran terhadap subjek ke- $i$  dan  $N = n_1 + \dots + n_K$  total pengukuran. Kita notasikan vektor pengukuran subjek ke- $i$  sebagai  $\mathbf{Y}_i = [Y_{i1}, \dots, Y_{in_i}]^T$  dengan vektor mean  $\boldsymbol{\mu}_i = [\mu_{i1}, \dots, \mu_{in_i}]^T$ , kemudian  $\mathbf{X}_i = [X_{i1}, \dots, X_{in_i}]^T$  sebagai matriks kovariat berukuran  $(n_i \times p)$ . Secara umum, Untuk memodelkan hubungan regresi antara respon dan kovariat-kovariat kita dapat menggunakan GLM sebagai berikut:

$$g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta} \quad (1)$$

di mana  $\boldsymbol{\mu}_i = E(\mathbf{Y}_i | \mathbf{X}_i)$ ,  $g$  adalah fungsi penghubung dan  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^T$  merupakan vektor koefisien regresi yang tidak diketahui dan akan diduga. Dalam GLM kita asumsikan bahwa distribusi marjinal dari  $Y_{ij}$  mengikuti distribusi dari keluarga eksponensial berikut:

$$f(y, \theta, \lambda) = \exp\{y\theta - b(\theta) / a(\lambda) + c(y, \lambda)\}$$

di mana  $a(\cdot)$ ,  $b(\cdot)$  dan  $c(\cdot)$  adalah suatu fungsi yang diketahui,  $\theta$  merupakan parameter kanonik dan  $\lambda$  adalah parameter dispersi.

Persamaan GEE untuk menduga vektor parameter regresi  $\beta$  berukuran  $p \times 1$  diberikan oleh:

$$S(\beta) = \sum_{i=1}^K \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} [Y_i - \mu_i(\beta)] = 0, \quad (2)$$

di mana  $V_i$  merupakan matriks kovarian dari  $Y_i$  yang diberikan oleh:

$$V_i = \lambda A_i^{1/2} R(\alpha) A_i^{1/2}, \quad (3)$$

dengan  $A_i$  suatu matriks diagonal dari fungsi varians  $v(\mu_{ij})$ , dan  $R(\alpha)$  adalah matriks korelasi bekerja (*working correlation matrix*) dari  $Y_i$  yang diindeks oleh sebuah vektor parameter  $\alpha$ . Penyelesaian persamaan (2) diperoleh dengan meng-alternasi antara pendugaan terhadap  $\lambda$ ,  $\alpha$  and  $\theta$ . Liang and Zeger memberikan pendugaan yang konsisten terhadap  $\lambda$ ,  $\alpha$  dan  $\theta$  berdasarkan nilai kuadrat dan perkalian silang residual Pearson:

$$r = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}.$$

Dibandingkan dengan persamaan untuk memperoleh penduga pada model regresi linier (normal) dan pada model linier tergeneralisir (GLM), GEE ini mengalami generalisasi atau perbedaan dalam beberapa hal sebagai berikut:

1. Dalam model regresi linier dan GLM respon  $Y_i$ , ekspektasi  $E(Y_i) = \mu_i$  merupakan variabel univariat, sedangkan dalam GEE mereka berupa vektor yang berhubungan dengan subjek ke- $i$ , sebagai konsekuensinya, maka model linier harus digeneralisasi dengan mempertimbangkan jumlah untuk seluruh individu  $Y$ ;
2. Dalam regresi linier, nilai  $(\partial \mu_i / \partial \eta_i)$  adalah 1, pada GLM nilainya bergantung pada fungsi link  $g(\cdot)$ . Dalam GEE, karena ekspektasi dan prediktor liniernya dua-duanya merupakan vektor berukuran  $t$ , maka  $(\partial \mu_i / \partial \eta_i)$  merupakan matrik diagonal berukuran  $t \times t$  dengan unsur diagonalnya adalah  $(\partial \mu_{ij} / \partial \eta_{ij})$  yang nilainya riilnya juga masih bergantung pada fungsi link  $g(\cdot)$  yang digunakan.

3. Dalam regresi linier, varians dari respon,  $\text{var}(Y_i) = \phi v(\mu_i)$  adalah konstan yaitu  $\sigma^2$ . Dalam GLM dia tidak konstan tetapi berupa matriks diagonal, sedangkan dalam GEE dia berupa matriks kovarians yang bersifat umum (simetris) yang tidak saja bergantung pada  $\mu$  atau  $\beta$  tetapi juga pada  $\phi$  dan  $\alpha$ , yang dapat dinyatakan dalam bentuk persamaan (3). Dengan demikian secara keseluruhan  $\text{Var}(\mathbf{Y})$  untuk regresi normal adalah  $2\mathbf{I}$ , untuk GLM adalah matriks diagonal dengan unsur diagonal  $\text{Var}(Y_i)$ , sedangkan pada GEE dia adalah matriks diagonal blok dengan blok ke-i adalah  $\mathbf{V}_i$ .

Ada beberapa metode untuk menentukan matriks korelasi  $\mathbf{R}_i(\alpha)$  yang biasa digunakan untuk memodelkan matriks korelasi dari  $\mathbf{Y}_i$  tersebut. Dimensi dari vector  $\alpha$  (yang diperlakukan sebagai parameter pengganggu) dan bentuk dari penduga  $\alpha$  berbeda untuk setiap metode. Berikut ini beberapa metode yang sering dipilih untuk menentukan matriks  $\mathbf{R}_i(\alpha)$  :

1.  $\mathbf{R}_i(\alpha) = \mathbf{R}_0$ , yaitu menetapkan suatu matriks korelasi dengan nilai tetap. Jika  $\mathbf{R}_0 = \mathbf{I}$  (yaitu matriks identitas), prosedur GEE akan menyederhana menjadi *independence estimating equation*.
2. *Autoregressive-1 (AR1)*, elemen-elemen matriks korelasi diberikan oleh rumus berikut:

$$\text{Cor}(Y_{ij}, Y_{i(j+t)}) = \alpha^t, \quad t = 1, 2, \dots, p-1$$

Struktur matriks *autoregressive-1*:

$$\mathbf{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{p-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{p-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{p-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha^{p-1} & \alpha^{p-2} & \alpha^{p-3} & \dots & 1 \end{pmatrix}$$

3. *Exchangeable*, korelasi antara  $Y_{ij}$  dan  $Y_{ik}$  bernilai sama untuk  $j \neq k$  :

$$\text{Cor}(Y_{ij}, Y_{ik}) = \alpha, \quad j \neq k.$$

Struktur matriks *exchangeable*:

$$\mathbf{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{pmatrix}$$

4. *Unstructured*, korelasi antara  $Y_{ij}$  dan  $Y_{ik}$  berbeda untuk  $j \neq k$ :

$$\text{Cor}(Y_{ij}, Y_{ik}) = \alpha_{jk}$$

Pada metode GEE biasa, penyelesaian untuk  $\beta$  dapat dilakukan dengan menggunakan algoritma berikut (Johnson & Stokes, 1996):

1. Hitung penduga awal dari  $\hat{\beta}_{GEE}$ , misalnya dengan menggunakan GLM biasa dengan asumsi saling bebas.
2. Penduga yang baru  $\hat{\beta}_{GEE}$  diperbaharui dengan meregresikan vektor respon bekerja (*working response vector*)  $\mathbf{Z}^* = \mathbf{X}\hat{\beta} + \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$  terhadap  $\mathbf{X}$ . Sebuah penduga baru  $\hat{\beta}_{new}$  diperoleh dengan rumus:

$$\hat{\boldsymbol{\beta}}_{new} = (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* \mathbf{Z}^* \quad (3)$$

di mana  $\mathbf{W}^*$  adalah matriks diagonal dari blok bobot dengan blok ke- $i$  merupakan

$$\text{matriks berukuran } n_i \times n_i \text{ berikut: } \mathbf{W}_i^* = \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^{-1} \mathbf{A}_i^{-1} \mathbf{R}_i(\hat{\boldsymbol{\alpha}}) \mathbf{A}_i^{-1} \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^{-1}.$$

3. Gunakan  $\hat{\beta}_{new}$  untuk memperbaharui  $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\beta}_{new} = \mathbf{H}\mathbf{Z}^*$ , dengan

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*.$$

4. Ulang langkah-langkah di atas hingga konvergen.

Algoritma pendugaan  $\beta$  di atas memberikan hasil yang baik ketika data tidak mengandung pencilan, namun akan menjadi menyimpang ketika diterapkan pada data yang mengandung pencilan.

Beberapa peneliti telah mengajukan metode *robust* untuk mengatasi masalah pencilan pada GEE, seperti yang dilakukan oleh Qaqish dan Preisser (1999), Gill (2000), Jung dan Ying (2003), Wang dan Zhu (2006) dan Abebe dkk (2014). Dalam penelitian ini

kami mengembangkan suatu metode *robust* GEE dengan pendekatan yang berbeda dari metode-metode di atas.

## 2.2 Review LTS

Metode LTS merupakan sebuah metode statistika yang digunakan untuk menduga koefisien regresi dari data normal yang mengandung pencilan dengan meminimumkan jumlah kuadrat galat terhadap sebaran data yang sudah terpankang (*trimmed*) atau sebaran terwinsorkan (*winsorized distribution*). Dengan kata lain, metode LTS menduga koefisien regresi dengan menggunakan metode *ordinary least square* (OLS) terhadap subhimpunan data terbaik  $H$ , yaitu

$$\hat{\beta}_{LTS} = \arg \min_{\beta} \left( \sum_{i=1}^h e_i^2 \right) = \arg \min_{\beta} \left( \sum_{i=1}^h (|y_i - \hat{y}_i|)^2 \right),$$

dengan  $h$  merupakan banyaknya anggota dalam subhimpunan  $H$  dan  $e_1^2 \leq e_2^2 \leq \dots \leq e_h^2 \leq \dots \leq e_n^2$  adalah kuadrat residual yang diurutkan dari yang terkecil sampai yang terbesar. Koefisien dugaan LTS dihitung dengan meminimalkan  $h$  kuadrat residual terurut di mana  $h$  dipilih pada interval:  $\frac{n}{2} + 1 \leq h \leq \frac{3n + p + 1}{4}$ , dengan  $n$  merupakan ukuran sampel dan  $p$  banyaknya parameter.

Ketika  $n$  kecil, semua kemungkinan subhimpunan berukuran  $H$  dapat disusun untuk menentukan subhimpunan terbaik yang meminimumkan kriteria LTS. Misalnya jika  $n=40$  dan  $H=n/2$ , maka terdapat  $10^{11}$  subhimpunan. Jika  $n$  berukuran besar maka akan semakin banyak kemungkinan subhimpunan yang dapat dibentuk yang mengakibatkan komputasi menjadi memakan waktu (*time consuming*). Untuk mengatasi itu Rousseeuw dan van Driessen (2006) mengajukan sebuah algoritma cepat untuk perhitungan metode LTS. Algoritma cepat ini adalah dengan mengiterasikan beberapa langkah terhadap sejumlah besar pengamatan-pengamatan sebagai nilai awal, kemudian menyimpan beberapa subhimpunan terbaik. Selanjutnya selanjutnya subhimpunan ini digunakan untuk iterasi penuh sampai konvergen dan diperoleh nilai dugaan akhir. Algoritma yang dihasilkan ini membuat pendugaan LTS menjadi lebih cepat.

Rousseeuw dan Leroy (1987) memperlihatkan konsistensi dan kenormalan asimtotik LTS pada tingkat  $n^{1/2}$ , ini sama dengan efisiensi asimtotik dari penduga Huber.

Pembuktian lengkap tentang sifat konsistensi dan kenormalan asimtotik metode LTS ini dapat dilihat dalam dua seri tulisan Visek (2006a, 2006b). Karena keunggulan metode LTS maka beberapa peneliti telah mengadaptasi dan menggunakan metode ini untuk berbagai analisis *robust*, diantaranya dalam analisis pengolahan citra (*image processing*) metode LTS diadaptasi untuk pengukuran jarak Hausdorff pada pengepasan object (*object matching*) (Sim,1999) dan untuk mendeteksi pixel terkontaminasi pada pengenalan wajah (*face recognition*) (Lai & Jiang, 2013), dalam analisis biomedis metode LTS digunakan untuk pengelompokan data genetik (Somorjai dkk, 2003), pada analisis regresi sendiri LTS telah dikembangkan untuk analisis regresi multivariate (Agullo dkk, 2008) dan untuk regresi linier dengan error berkorelasi (Roozbeh & Babaie-Kafaki, 2015). Namun kajian LTS pada analisis GEE *robust* belum pernah dilakukan dan menurut kami merupakan kajian menarik yang cukup menantang dan dapat memberikan kontribusi besar dalam pengembangan ilmu statistika, yaitu khususnya untuk mengatasi masalah analisis GEE pada data mengandung pencilan.

### **BAB 3**

#### **TUJUAN DAN MANFAAT PENELITIAN**

##### **2.1. Tujuan Penelitian**



Tujuan penelitian ini adalah:

1. Mengembangkan suatu metode *robust* untuk pendugaan model GEE yang dapat mengatasi tingkat kontaminasi pencilan pada data hingga di atas 30%.
2. Mengkaji sifat-sifat asimtotik dari metode yang diajukan.
3. Menguji metode yang dikembangkan untuk mengetahui nilai *breakdown point*<sup>1</sup> metode dan membandingkannya dengan metode *robust* GEE lainnya.

## **2.2. Manfaat Penelitian**

Mengingat GEE saat ini telah menjadi populer dan banyak digunakan dalam berbagai bidang penelitian, khususnya pada penelitian longitudinal dan penelitian berkelompok (*cluster*), maka diharapkan hasil dari penelitian ini akan menjadi referensi bagi para peneliti dan menjadi alat bagi para pengguna statistika khususnya pengguna metode GEE. Dengan kemampuan GEE dalam analisis data berkorelasi dan fleksibilitasnya untuk digunakan pada data normal maupun nonnormal, maka GEE robust akan menjadi alat statistika yang penting bagi para ahli statistik dan praktisi untuk mengatasi data yang mengandung pencilan.

## **BAB 4 METODE PENELITIAN**

---

<sup>1</sup> *Breakdown point* adalah jumlah maksimum pencilan yang dapat ditoleransi oleh suatu metode, nilai ini sering digunakan sebagai ukuran ketegaran suatu metode (Croux & Gazen, 2005).

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut

### 1. Studi bibliographi

Pada tahap ini kami melakukan penelusuran dan pengkajian terhadap literatur-literatur tentang GEE, GLM, LTS, pendugaan *robust*, dan teori-teori statistika lainnya yang berkaitan dengan metode yang kami kembangkan. Khusus untuk teknik membangkitkan data non-normal yang saling berkorelasi, kami mengkaji beberapa literatur diantaranya Yahav dan Shmueli (2012), Demirtas (2014) dan Minhajuddin dkk (2004). Teknik yang kami gunakan untuk simulasi model GEE non-normal adalah teknik NORTA (*Normal To Anything*), kajian detail tentang teknik NORTA dapat dilihat pada Niavarani dan Smith (2013), Niaki dan Abbasi (2008) atau Ghosh dan Henderson (2005).

### 2. Pengembangan metode IRLTS

Pada tahap ini penyelidikan dilakukan untuk menemukan solusi tentang bagaimana menduga model GEE dengan metode IRLTS, selanjutnya menyusun algoritma perhitungan dugaan parameter regresi dari model GEE dengan metode IRLTS. Penyelidikan juga dilengkapi dengan mengkaji sifat-sifat asimtotik IRLTS untuk mengetahui keterandalan metode IRLTS sebagai metode *robust*.

### 3. Pengujian Empiris

Pengujian empiris dilakukan dengan melakukan simulasi data menggunakan bantuan perangkat lunak R. Data yang digunakan merupakan data bangkitan (*generated data*). Rancangan simulasi yang kami lakukan sebagai berikut:

Membangkitkan data berukuran  $N=1000$  pengamatan saling berkorelasi dengan desain data terdiri dari 200 subjek dan 5 kali pengukuran berulang. Kami menggunakan model dengan dua variabel bebas  $X$ , maka model yang digunakan dalam simulasi adalah:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij}.$$

Kami membangkitkan variabel respon dari keluarga exponential dari dua jenis respon yaitu kontinu dan diskrit. Untuk respon kontinu kami menggunakan distribusi normal dan gamma, sedangkan untuk respon diskrit kami gunakan distribusi Poisson binomial.

Korelasi pada data kami setting menggunakan dua jenis matriks korelasi yaitu *exchangeable* dan *autoregressive-1* dengan nilai-nilai  $\alpha$  seperti berikut:

1. *Exchangeable*

$$R = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

2. *Autoregressive-1*

$$R = \begin{bmatrix} 1 & 0.5 & 0.25 & 0.125 & 0.0625 \\ 0.5 & 1 & 0.5 & 0.25 & 0.125 \\ 0.25 & 0.5 & 1 & 0.5 & 0.25 \\ 0.125 & 0.25 & 0.5 & 1 & 0.5 \\ 0.0625 & 0.125 & 0.25 & 0.5 & 1 \end{bmatrix}$$

Kami mempertimbangkan kondisi data tanpa pencilan ( $\epsilon = 0\%$ ) dan data terkontaminasi pencilan ( $\epsilon = 1\%, 5\%, 10\%, 20\%, 30\%, 40\%$ ). Selanjutnya menduga koefisien GEE dari data yang sudah dibangkitkan dengan metode IRLTS dan metode GEE standard.

Analisis GEE *robust* dan standard dilakukan terhadap 1000 sampel data bangkitan yang berbeda (1000kali replikasi). Kami mengevaluasi hasilnya berdasarkan nilai kuadrat tengah galat (KTG) atau *mean square error* (MSE) penduga yang merupakan jumlah dari kuadrat bias dan varians penduga.

Dalam simulasi yang kami lakukan, untuk pendugaan dengan metode IRLTS kami menetapkan nilai  $h=n/2$ .

## BAB 5 HASIL YANG DICAPAI

Dalam penelitian ini kami telah berhasil menyusun sebuah algoritma cepat untuk pendugaan metode IRLTS dan telah melakukan pengujian empiris dengan simulasi Monte Carlo untuk beberapa kasus pencilan pada model bedistribusi normal, gamma, Poisson dan binomial untuk dua jenis matriks korelasi yaitu *exchangeable* dan *autoregressive-1*.

### 5.1. Algoritma IRLTS

Pendugaan model GEE dengan metode IRLTS dilakukan dengan algoritma sebagai berikut:

*Concentration-step:*

1. Ambil  $h$  pengamatan dengan  $\frac{n}{2} + 1 \leq h \leq \frac{3n + p + 1}{4}$ , dimana  $p$  adalah jumlah peubah bebas.
2. Hitung penduga awal dari  $\hat{\beta}$  dengan menggunakan metode GEE klasik (IRLS) dengan asumsi saling bebas terhadap  $h$  pengamatan, yaitu menggunakan rumus :

$$\hat{\beta}_{awal} = (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* \mathbf{Z}^*$$

dengan

$$\mathbf{Z}^* = \mathbf{X} \hat{\beta}_{awal} + \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}_{awal}} (\mathbf{y} - \hat{\boldsymbol{\mu}})$$

dan

$$\mathbf{W}_i^* = \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_{awal}} \right)^{-1} \mathbf{A}_i^{-1} \mathbf{R}_i(\hat{\boldsymbol{\alpha}}) \mathbf{A}_i^{-1} \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_{awal}} \right)^{-1}$$

secara iteratif dengan memperbaharui

$$\hat{\boldsymbol{\eta}} = \mathbf{X} \hat{\boldsymbol{\beta}}_{old} = \mathbf{H} \mathbf{Z}^*$$

menggunakan

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*.$$

3. Gunakan dugaan koefisien model diatas untuk mengitung nilai dugaan:

$$\hat{\boldsymbol{\mu}}_i = g^{-1}(\mathbf{X}_i \hat{\boldsymbol{\beta}}_{old})$$

4. Hitung galat dari seluruh  $n$  data:  $e_{ij} = Y_{ij} - \hat{\mu}_{ij}$  berdasarkan persamaan  $\hat{\mu}_i = g^{-1}(\mathbf{X}_i \hat{\boldsymbol{\beta}}_{old})$
5. Urutkan  $|e_{ij}|$  untuk  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, t$  dari yang terkecil hingga yang terbesar :  $|e_{11}| \leq |e_{12}| \leq \dots \leq |e_m|$
6. Pilih  $h$  pengamatan yang memiliki  $h$ -galat terkecil, kita notasikan sebagai subset H.

Pengulangan *concentration-step* akan menghasilkan suatu proses iterasi.

### **IRLTS algorithm:**

1. Tentukan  $h$  pengamatan.
2. Hitung penduga awal dari  $\hat{\boldsymbol{\beta}}$  berdasarkan  $h$  pengamatan dengan menggunakan metode IRLS, kita peroleh  $\hat{\mu}_i = g^{-1}(\mathbf{X}_i \hat{\boldsymbol{\beta}})$ .
3. Hitung residual dari  $n$  pengamatan :  $e_{ij} = Y_{ij} - \hat{\mu}_{ij}$ .
4. Urutkan  $|e_{ij}|$  untuk  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, t$  dari yang terkecil hingga yang terbesar :  $|e_{11}| \leq |e_{12}| \leq \dots \leq |e_m|$
5. Pilih  $h_1$  pengamatan yang memiliki residual terkecil, kita notasikan sebagai subhimpunan  $H_1$ .
6. Jalankan *concentration-step* terhadap  $H_1$  dua kali, dan kita peroleh  $H_1^*$ .
7. Ulangi langkah step 1- step 6 untuk  $\binom{n}{h_1}$  kali
8. Dari  $\binom{n}{h_1}$  hasil, pilih 10 subhimpunan terbaik (yang memiliki jumlah kuadrat galat terkecil)  $H_q, q=1, \dots, 10$ .
9. Jalankan *concentration-step* terhadap 10 subhimpunan terbaik  $H_q$  sampai konvergen.
10. Pilih subhimpunan terbaik H.

## 5.2. Hasil Simulasi

### 5.2.1. Model Regresi Normal

Berikut hasil pengujian empiris untuk model berdistribusi normal.

$$\text{Model : } \mu_i = X_i\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k.$$

1. Dengan matriks korelasi *exchangeable*

Tabel 1. Nilai harapan (E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *exchangeable* untuk model GEE normal

Method	Classical GEE			IRLTS		
	$\hat{\beta}_0$					
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	1.092751	1.084446	6.89E-05	7.22E-05	8.67E-03	0.007134
1%	1.386088	1.047247	0.032525	1.15E-01	1.82E-01	0.002411
5%	6.731881	1.102981	0.002906	3.17E+01	3.29E+01	0.013196
10%	12.65882	0.884682	4.718481	1.39E+02	1.43E+02	0.018111
20%	19.93999	1.081184	6.468163	3.56E+02	3.63E+02	0.007880
30%	23.45578	0.998279	9.034601	5.04E+02	5.13E+02	0.000972
40%	36.60281	1.511892	10.492129	1.23E+03	1.27E+03	0.263016
	$\hat{\beta}_1$					
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	0.963772	0.966477	1.67E-02	1.60E-02	0.001414	1.19E-03
1%	1.161652	0.988342	6.99E-02	1.58E-01	0.04568	2.69E-04
5%	1.113516	0.974106	3.16E+01	3.32E+01	0.059902	8.22E-04
10%	0.247617	1.033552	1.54E+02	1.35E+02	0.672261	1.21E-03
20%	0.127293	0.994076	3.93E+02	3.59E+02	0.815828	8.69E-04
30%	0.985436	0.994466	5.05E+02	5.05E+02	0.014695	9.40E-05
40%	1.388427	1.021476	1.24E+03	1.27E+03	0.150917	5.87E-04
	$\hat{\beta}_2$					
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.001861	1.002416	8.26E-03	8.16E-03	5.18E-06	6.81E-06
1%	1.067707	1.017931	1.44E-01	1.36E-01	4.72E-02	4.03E-04
5%	0.617707	0.989511	3.74E+01	3.30E+01	1.76E-01	1.11E-04
10%	1.105493	0.999549	1.34E+02	1.36E+02	3.85E-01	3.83E-04
20%	1.623563	1.011352	3.36E+02	3.58E+02	5.70E-01	2.82E-04
30%	2.158404	1.048966	4.54E+02	5.02E+02	2.02E+00	2.71E-03
40%	-0.30025	0.871031	1.36E+03	1.28E+03	1.75E+00	1.66E-02

2. Dengan matriks korelasi *autoregressive-1*

Tabel 2. Nilai harapan(E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *autoregressive-1* untuk model GEE normal

Method	Classical GEE			IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	1.030284	1.042708	0.003600	8.86E-03	4.52E-03	0.010527
1%	0.997654	1.150559	0.009454	2.94E-02	9.46E-03	0.028697
5%	6.50429	0.881484	0.259620	3.16E+01	3.06E+01	0.035804
10%	9.994994	0.970323	3.062796	8.15E+01	8.40E+01	0.011416
20%	16.74356	0.918175	4.141901	2.50E+02	2.56E+02	0.008374
30%	30.29975	1.219171	8.995525	8.46E+02	8.63E+02	0.071200
40%	37.75225	1.341671	20.70969	1.33E+03	1.37E+03	0.140363
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	0.980363	0.975022	2.72E-03	3.69E-03	0.000612	1.26E-03
1%	1.234786	0.970659	6.29E-02	8.30E-04	0.061842	9.62E-04
5%	1.068068	1.022869	2.97E+01	3.00E+01	0.127005	2.15E-03
10%	1.679534	1.071079	6.94E+01	7.96E+01	0.725131	6.76E-03
20%	2.021052	0.99121	2.17E+02	2.48E+02	1.136598	9.49E-05
30%	1.082956	0.989612	8.54E+02	8.59E+02	0.335845	2.34E-04
40%	1.297996	0.916068	1.34E+03	1.36E+03	6.931952	1.23E-02
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.010495	1.013388	3.95E-04	2.86E-04	0.000114	0.000179
1%	1.105557	0.986604	1.55E-02	2.06E-03	0.015006	0.002121
5%	0.712987	1.015173	3.38E+01	3.01E+01	0.345357	0.000313
10%	0.55209	0.94338	8.92E+01	8.19E+01	0.223766	0.003229
20%	0.784725	1.03267	2.56E+02	2.47E+02	1.496681	0.001152
30%	0.016743	0.989729	9.19E+02	8.59E+02	2.759813	0.004597
40%	-0.77829	0.983936	1.49E+03	1.35E+03	4.330139	0.000315

### 5.2.2. Model Regresi Gamma

Model untuk distribusi gamma memiliki fungsi link kebalikan (invers), dengan demikian modelnya dapat ditulis sebagai berikut

$$(1/\mu_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k,$$

atau dapat ditulis sebagai

$$\mu_i = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)^{-1}$$

Untuk membangkitkan data berdistribusi gamma yang memenuhi model GEE, yaitu pengamatan saling berkorelasi, diperlukan prosedur tahapan dengan menggunakan teknik khusus untuk membangkitkan data non-normal yang saling berkorelasi. Teknik yang kami gunakan dalam simulasi ini adalah NORTA . Dengan teknik ini maka tahapannya adalah sebagai berikut:

1. Tentukan matriks korelasi **R** dengan struktur elemen sesuai dengan yang diinginkan. Yaitu

*Exchangeable:*

$$R = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

*Autoregressive-1 :*

$$R = \begin{bmatrix} 1 & 0.5 & 0.25 & 0.125 & 0.0625 \\ 0.5 & 1 & 0.5 & 0.25 & 0.125 \\ 0.25 & 0.5 & 1 & 0.5 & 0.25 \\ 0.125 & 0.25 & 0.5 & 1 & 0.5 \\ 0.0625 & 0.125 & 0.25 & 0.5 & 1 \end{bmatrix}$$

2. Bangkitkan data normal multivariate dengan vektor mean  $\mu=(0,0,\dots,0)^T$  dan matriks korelasi  $\Sigma=R$
3. Hitung nilai fungsi densitas kumulatif (CDF=*cumulative density function*) dari setiap pengamatan pada data normal multivariat yang diperoleh di langkah 2



4. Bangkitkan galat (error) berdistribusi gamma berdasarkan nilai kuantil-kuantil gamma yang bersesuaian dengan nilai CDF pada langkah 3 (konversi data normal ke gamma)
5. Bangkitkan  $X_1$  dan  $X_2$ , dalam simulasi ini kami menggunakan distribusi uniform(1,5) dan integer (1-5).
6. Bangkitkan  $Y=1/(1+ X_1 + X_2+ \text{error})$ .
7. Menduga model GEE untuk distribusi gamma terhadap data  $Y$ ,  $X_1$  dan  $X_2$ . Koefisien yang diperoleh dari pembangkitan data awal tanpa pencilan ini merupakan parameter  $\beta_i$  yang ditetapkan.
8. Berikan kontaminasi terhadap data awal dengan pencilan sebesar 1%, 5%, 10%, 20%, 30% dan 40% dari total jumlah data ( $n_{\text{total}}=1000$ ), pencilan dibangkitkan secara acak dari distribusi gamma(2,2).
9. Menduga model GEE untuk distribusi gamma terhadap data terkontaminasi pada langkah 8 menggunakan menggunakan metode GEE standard dan IRLTS, simpan nilai dugaan parameter yang diperoleh dari kedua metode.
10. Ulangi langkah 8-9 sebanyak 1000 kali.
11. Hitung nilai ekspektasi, galat baku ( $se=\text{standard error}$ ) dan kuadrat tengah galat ( $mse=\text{mean square error}$ ) dari nilai dugaan parameter kedua metode.

Berbeda dengan simulasi untuk data normal di mana parameter  $\beta_i$  kami tetapkan secara langsung (yaitu  $\beta_i = 1$ , untuk  $i=0,1,2$ ), pada kasus data non-normal hal tersebut tidak dapat dilakukan karena adanya proses konversi distribusi dari normal ke non-normal (teknik NORTA) yang mengakibatkan adanya perubahan nilai parameter (*downbias* atau *upbias*) untuk setiap segugus sample data yang dibangkitkan. Sehingga untuk mengatasi itu maka data awal tanpa pencilan dianggap sebagai data populasi dan koefisien model dari data awal ditetapkan sebagai nilai parameter. Pada simulasi untuk distribusi gamma kami peroleh nilai parameter  $\beta_i$  untuk kasus matriks *exchangeable* adalah sebagai berikut:  $\beta_0= 1.338535$ ,  $\beta_1= 1.007061$  dan  $\beta_2=1.045502$ . Sedangkan nilai parameter  $\beta_i$  untuk kasus matriks *autoregressive-1* adalah sebagai berikut :  $\beta_0= 1.401825$ ,  $\beta_1= 0.988603$  dan  $\beta_2=1.025601$ .

Berdasarkan simulasi yang kami lakukan membangkitkan data nonnormal dengan konversi distribusi normal menggunakan teknik NORTA memberikan hasil yang baik

dalam mempertahankan nilai-nilai dari elemen matriks korelasi yang diinginkan oleh peneliti. Berikut kami sajikan matriks korelasi dari distribusi normal dan matriks korelasi untuk distribusi gamma hasil konversi dengan teknik NORTA.

a. Kasus *Exchangeable*

Matriks korelasi data normal:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.5638784	0.5521484	0.5781597	0.5205594
[2,]	0.5638784	1.0000000	0.5455666	0.5058149	0.5556930
[3,]	0.5521484	0.5455666	1.0000000	0.5580789	0.5832907
[4,]	0.5781597	0.5058149	0.5580789	1.0000000	0.4855287
[5,]	0.5205594	0.5556930	0.5832907	0.4855287	1.0000000

Matriks korelasi data gamma hasil konversi:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.5487579	0.5749420	0.5700902	0.5304298
[2,]	0.5487579	1.0000000	0.5412140	0.4841419	0.5052032
[3,]	0.5749420	0.5412140	1.0000000	0.4857146	0.6227979
[4,]	0.5700902	0.4841419	0.4857146	1.0000000	0.4483169
[5,]	0.5304298	0.5052032	0.6227979	0.4483169	1.0000000

b. Kasus *Autoregressive-I*

Matriks korelasi data normal:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.000000000	0.51827971	0.2997689	0.04789696	0.001830824
[2,]	0.518279710	1.0000000	0.6165370	0.28675416	0.036428174
[3,]	0.299768861	0.61653698	1.0000000	0.43216752	0.190331985
[4,]	0.047896964	0.28675416	0.4321675	1.00000000	0.567395918
[5,]	0.001830824	0.03642817	0.1903320	0.56739592	1.000000000

Matriks korelasi data gamma hasil konversi:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.00000000	0.49806872	0.2857593	0.03168619	0.01442081
[2,]	0.49806872	1.00000000	0.5478124	0.19118799	0.01586682
[3,]	0.28575929	0.54781238	1.00000000	0.42549574	0.17125361
[4,]	0.03168619	0.19118799	0.4254957	1.00000000	0.55103491
[5,]	0.01442081	0.01586682	0.1712536	0.55103491	1.00000000

Berikut hasil pengujian empiris untuk model berdistribusi gamma.

1. Dengan matriks korelasi *exchangeable*

Tabel 3. Nilai harapan (E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *exchangeable* untuk model GEE gamma

Method	Classical GEE			IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	1.338535	0.000000	0.000000	1.561729	0.000000	0.049816
1%	1.587133	0.387475	0.150137	1.331742	0.277444	0.024127
5%	1.833421	0.417356	0.419098	1.292539	0.280084	0.080563
10%	1.963008	0.551804	0.694453	1.343505	0.345998	0.119739
20%	2.105067	0.522999	0.861099	1.302799	0.245013	0.061308
30%	2.158723	0.712847	1.180859	1.432421	0.581205	0.346613
40%	2.343510	0.758355	1.585077	0.942300	1.080309	1.324070
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	1.007061	0.000000	0.000000	0.923761	0.000000	0.006939
1%	0.905981	0.152981	0.023403	0.954914	0.041541	0.003474
5%	0.664892	0.297197	0.205406	0.957103	0.034595	0.003693
10%	0.587415	0.348410	0.297492	0.968876	0.075239	0.007119
20%	0.314225	0.347369	0.600688	0.964959	0.039372	0.003323
30%	0.297561	0.338291	0.617832	0.937923	0.088894	0.012682
40%	0.566934	0.538557	0.483756	1.124362	0.436399	0.204204
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.045502	0.000000	0.000000	0.952301	0.000000	0.008686
1%	0.930571	0.147962	0.021893	0.964191	0.019529	0.006851
5%	0.743453	0.289166	0.174850	0.974127	0.039936	0.006689
10%	0.549449	0.423348	0.425292	0.958451	0.018566	0.007922
20%	0.355655	0.349199	0.597828	0.975630	0.032734	0.005954
30%	0.332191	0.336312	0.621917	0.969043	0.071512	0.010960
40%	0.548824	0.493729	0.490457	0.995581	0.057913	0.005846

2. Dengan matriks korelasi *autoregressive-1*

Tabel 4. Nilai harapan(E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *autoregressive-1* untuk model GEE gamma.

Method	Classical GEE			IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	1.401825	0.000000	0.000000	1.519920	0.000000	0.013946
1%	1.500068	0.185410	0.034377	1.292374	0.267690	0.031861
5%	1.713422	0.174403	0.127509	1.169724	0.219998	0.102270
10%	2.051958	0.292341	0.476604	1.172668	0.169948	0.081395
20%	2.096749	0.279746	0.561178	1.182726	0.281815	0.127424
30%	2.208355	0.570003	0.975395	1.094088	0.105856	0.105908
40%	2.583383	0.515743	1.662071	1.355114	0.430022	0.187101
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	0.988603	0.000000	0.000000	0.923267	0.000000	0.004269
1%	0.907956	0.097084	0.009425	0.948473	0.035394	0.002228
5%	0.746269	0.169773	0.087549	0.977105	0.032371	0.001180
10%	0.493787	0.257591	0.311197	0.974237	0.026179	0.000892
20%	0.290721	0.257720	0.553459	0.982542	0.035247	0.001279
30%	0.228004	0.231963	0.595925	0.981851	0.023087	0.000579
40%	0.295225	0.306991	0.575017	0.971695	0.052629	0.003056
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.025601	0.000000	0.000000	0.962045	0.000000	0.004039
1%	0.971239	0.086763	0.007528	0.983387	0.028004	0.002111
5%	0.739142	0.224797	0.132593	0.981539	0.021532	0.002405
10%	0.511909	0.302973	0.355672	0.981291	0.028264	0.002762
20%	0.281145	0.270400	0.627330	0.983924	0.028529	0.002551
30%	0.225163	0.377600	0.646722	1.003760	0.035796	0.001758
40%	0.297677	0.352980	0.654468	0.974769	0.068322	0.007252

### 5.2.3. Model Regresi Poisson

Model regresi Poisson pada GLM memiliki fungsi link logaritma sebagai berikut:

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

atau dapat ditulis

$$\mu_i = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k).$$

Data berdistribusi Poisson yang memenuhi model GEE dibangkitkan menggunakan prosedur yang sama dengan model gamma yaitu dengan menggantikan distribusi gamma pada tahapan ke 5 dengan distribusi Poisson yang diperoleh menggunakan kuantil-kuantil Poisson yang bersesuaian dengan nilai CDF dari data normal. Data Y dibangkitkan menggunakan persamaan  $Y = \exp(1 + X_1 + X_2 + \text{error})$ . Sedangkan kontaminasi pencilan dibangkitkan secara acak dari distribusi Poisson ( $\lambda=5000$ ).

Nilai parameter  $\beta_i$  untuk distribusi Poisson kami peroleh untuk kasus matriks *exchangeable* adalah:  $\beta_0 = 0.874938$ ,  $\beta_1 = 1.153085$  dan  $\beta_2 = 1.096566$ . Sedangkan nilai parameter  $\beta_i$  untuk kasus matriks *autoregressive-1*:  $\beta_0 = 2.202333$ ,  $\beta_1 = 0.947705$  dan  $\beta_2 = 0.970118$ .

Berikut ini matriks korelasi dari distribusi normal dan matriks korelasi untuk distribusi Poisson hasil konversi serta hasil uji empiris pendugaan model GEE.

#### 1. Kasus *Exchangeable*

Matriks korelasi data normal:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.5395777	0.5457184	0.5434695	0.5488964
[2,]	0.5395777	1.0000000	0.5129178	0.4928067	0.5026590
[3,]	0.5457184	0.5129178	1.0000000	0.5835206	0.5335449
[4,]	0.5434695	0.4928067	0.5835206	1.0000000	0.4614220
[5,]	0.5488964	0.5026590	0.5335449	0.4614220	1.0000000

Matriks korelasi data gamma hasil konversi:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.4889263	0.5086858	0.4883726	0.4602878
[2,]	0.4889263	1.0000000	0.4946574	0.3798874	0.4331820
[3,]	0.5086858	0.4946574	1.0000000	0.5300413	0.4698371
[4,]	0.4883726	0.3798874	0.5300413	1.0000000	0.3899928
[5,]	0.4602878	0.4331820	0.4698371	0.3899928	1.0000000

## 2. Kasus Autoregressive-I

Matriks korelasi data normal:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.00000000	0.57266668	0.24785668	0.1243428	0.09897073
[2,]	0.57266677	1.00000000	0.5351493	0.2671806	0.10015096
[3,]	0.24785683	0.5351493	1.00000000	0.5173734	0.19540986
[4,]	0.12434275	0.2671806	0.5173734	1.00000000	0.43768665
[5,]	0.09897073	0.1001510	0.1954099	0.4376866	1.00000000

Matriks korelasi data gamma hasil konversi:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.00000000	0.51415361	0.1926366	0.1762554	0.07541080
[2,]	0.5141536	1.00000000	0.4869834	0.2370013	0.06037798
[3,]	0.1926366	0.48698337	1.00000000	0.4682331	0.16894839
[4,]	0.1762554	0.23700127	0.4682331	1.00000000	0.39812270
[5,]	0.0754108	0.06037798	0.1689484	0.3981227	1.00000000

Hasil pengujian empiris pendugaan model GEE Poisson disajikan dalam Tabel 5 dan Tabel 6 sebagai berikut.

Tabel 5. Nilai harapan (E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *exchangeable* untuk model GEE Poisson

Method	Classical GEE			IRLTS		
	$\hat{\beta}_0$					
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	0.874938	0.000000	0.000000	1.382454	0.000000	0.257572
1%	5.135087	0.432026	18.664606	1.906352	0.966794	0.724036
5%	7.851928	0.578389	49.012922	2.242959	0.411831	1.041084
10%	9.299508	0.776268	71.575965	2.084935	0.454712	1.670854
20%	10.710302	0.425014	96.915019	2.351178	0.546663	1.478125
30%	11.413164	0.621123	111.439999	1.988002	0.486656	1.475743
40%	11.616069	1.109335	116.602513	2.449406	0.478231	2.707652

Method	Classical GEE			IRLTS		
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	1.153085	0.000000	0.000000	1.037534	0.000000	0.013352
1%	0.676815	0.486207	0.236397	0.929211	0.234359	0.093310
5%	0.358318	0.060367	0.635299	0.901860	0.025510	0.063765
10%	0.193887	0.078982	0.926299	0.911826	0.029340	0.059067
20%	0.099502	0.062118	1.113897	0.909217	0.030959	0.060430
30%	0.069531	0.038471	1.175571	0.953702	0.037684	0.041174
40%	0.076081	0.042942	1.161782	0.972846	0.053091	0.035305
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.096566	0.000000	0.000000	0.994616	0.000000	0.010394
1%	0.627813	0.480668	0.231042	1.015382	0.132211	0.023639
5%	0.320968	0.047317	0.603791	0.914446	0.028837	0.033999
10%	0.210708	0.044968	0.786766	0.959878	0.040024	0.020285
20%	0.098703	0.054113	0.998659	0.925985	0.018188	0.029429
30%	0.041398	0.040679	1.115033	0.981431	0.042721	0.015081
40%	0.025875	0.029411	1.147245	0.851497	0.034816	0.061271

Tabel 6. Nilai harapan(E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *autoregressive-1* untuk model GEE Poisson.

Method	Classical GEE			IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	2.202333	0.000000	0.000000	2.474529	0.000000	0.074091
1%	5.110614	0.777663	9.062855	3.027439	0.824689	0.360911
5%	8.493574	0.689463	40.055071	1.614313	0.331758	0.455831
10%	9.788645	0.962589	58.478701	1.765065	0.204619	0.523307
20%	10.525858	0.877176	70.050502	2.356166	0.470975	0.724548
30%	10.635832	0.570803	71.449706	3.071576	0.974599	1.054251
40%	11.321868	0.519095	83.435379	2.725903	1.113441	1.513876

Method	Classical GEE			IRLTS		
	$\hat{\beta}_1$					
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	0.947705	0.000000	0.000000	0.875558	0.000000	0.005205
1%	0.667902	0.098130	0.087919	0.836358	0.054700	0.015390
5%	0.309635	0.108437	0.418891	1.052717	0.059291	0.014543
10%	0.194983	0.044670	0.568586	1.011790	0.074132	0.009602
20%	0.097441	0.077894	0.729016	0.918343	0.038557	0.002349
30%	0.092040	0.031461	0.733153	0.809874	0.101063	0.029211
40%	0.057724	0.029418	0.792931	0.948095	0.173304	0.030034
	$\hat{\beta}_2$					
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	0.970118	0.000000	0.000000	0.885872	0.000000	0.007097
1%	0.630164	0.089934	0.123657	0.817922	0.048233	0.025490
5%	0.281547	0.064601	0.478304	0.947537	0.021808	0.000986
10%	0.159566	0.060507	0.660655	0.967633	0.065401	0.004284
20%	0.087528	0.031481	0.779957	0.900333	0.028748	0.005696
30%	0.067109	0.035320	0.816673	0.825752	0.057643	0.024164
40%	0.052893	0.041477	0.843022	0.839545	0.083307	0.023989

#### 5.2.4. Model Regresi Binomial (Model Logit)

Model GLM untuk distribusi binomial memiliki fungsi link logit seperti berikut:

$$\log \left[ \frac{\pi}{1-\pi} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k,$$

atau

$$\pi = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))} \quad (4)$$

dengan  $\pi$  merupakan peluang sukses suatu kejadian variabel  $Y$  yang berdistribusi binomial.

Dalam model binomial ini, pengamatan regresi  $y \in \{0,1\}$  diasumsikan mengikuti distribusi Bernoulli dengan parameter mean (suatu peluang bernilai antara 0 dan 1) tergantung pada nilai-nilai dugaan. Jadi untuk setiap nilai dugaan yang diberikan,



hanya ada dua kemungkinan nilai galat/ error, yaitu:  $(1-\pi)$  dengan peluang terjadi sebesar  $\pi$ , dan  $(0-\pi)$  dengan peluang terjadi sebesar  $(1-\pi)$ . Untuk dugaan lain nilai galat akan sama dengan  $(1-\pi')$  dengan peluang terjadi sebesar  $\pi'$ , dan  $(0-\pi')$  dengan peluang terjadi sebesar  $(1-\pi')$ . Dengan demikian tidak ada distribusi galat bersama (*common error*) yang bebas dari nilai dugaan.

Berbeda dengan model normal dimana galat dari nilai dugaan mengikuti distribusi normal, dalam model binomial distribusi galat tidak mengikuti binomial, karena nilai galat berupa probabilitas yang bernilai kontinu antara 0 dan 1 sebagaimana yang dijelaskan di atas.

Untuk mengaplikasikan algoritma IRLTS dalam model GEE binomial, dimana pemilihan subsample terbaik untuk pendugaan parameter dilakukan berdasarkan pengamatan yang memiliki  $h$  galat terkecil (setelah diurutkan), maka kami mencoba menggunakan galat model yang dihitung sebagai selisih antara nilai dugaan  $\hat{\pi} = 1/\exp(1 + \exp(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k))$  dengan  $\pi = \exp(Y)/(1+\exp(Y))$ .

Data yang memenuhi model GEE binomial dibangkitkan menggunakan prosedur yang sama dengan model gamma dan Poisson yaitu dengan mengkonversi data normal ke kuantil-kuantil Poisson yang bersesuaian dengan nilai CDF dari data normal. Data  $Y$  dibangkitkan dengan membangkitkan secara random data berdistribusi Bernoulli sebanyak  $n=1000$  dengan parameter peluang menggunakan rumus (4). Sedangkan kontaminasi pencilan dibangkitkan secara acak dari distribusi Bernoulli (0.3).

Nilai parameter  $\beta_i$  untuk distribusi binomial kami peroleh untuk kasus matriks *exchangeable* adalah:  $\beta_0= 0.664901$ ,  $\beta_1=0.971221$  dan  $\beta_2=1.014409$ . Sedangkan nilai parameter  $\beta_i$  untuk kasus matriks *autoregressive-I*:  $\beta_0= 1.864317$ ,  $\beta_1= 0.86504$  dan  $\beta_2=1.240741$ .

Berikut ini matriks korelasi dari distribusi normal dan matriks korelasi untuk distribusi binomial hasil konversi serta hasil uji empiris pendugaan model GEE.

1. Kasus *exchangeable*

Matriks korelasi data normal:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.4215716	0.5052906	0.4329946	0.5297598
[2,]	0.4215716	1.0000000	0.5131713	0.4344994	0.4630810
[3,]	0.5052906	0.5131713	1.0000000	0.5108226	0.4449479
[4,]	0.4329946	0.4344994	0.5108226	1.0000000	0.4836009
[5,]	0.5297598	0.4630810	0.4449479	0.4836009	1.0000000

Matriks korelasi data binomial hasil konversi:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.4196585	0.4723432	0.3960880	0.5111212
[2,]	0.4196585	1.0000000	0.4847993	0.3703963	0.4165814
[3,]	0.4723432	0.4847993	1.0000000	0.4921482	0.4276644
[4,]	0.3960880	0.3703963	0.4921482	1.0000000	0.4630630
[5,]	0.5111212	0.4165814	0.4276644	0.4630630	1.0000000

## 2. Kasus *autoregressive-1*

Matriks korelasi data normal:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.5733886	0.2490881	0.1028065	0.05530921
[2,]	0.5733886	1.0000000	0.5324482	0.3225246	0.15249245
[3,]	0.2490880	0.5324482	1.0000000	0.5853229	0.30738284
[4,]	0.1028065	0.3225246	0.5853229	1.0000000	0.54004235
[5,]	0.05530921	0.1524924	0.3073828	0.5400423	1.0000000

Matriks korelasi data binomial hasil konversi:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.5184128	0.2076530	0.07753672	0.03107838
[2,]	0.51841277	1.0000000	0.4942312	0.30237469	0.14677034
[3,]	0.20765301	0.4942312	1.0000000	0.55932585	0.29498380

[4,] 0.07753672 0.3023747 0.5593258 1.00000000 0.49065763  
 [5,] 0.03107838 0.1467703 0.2949838 0.49065763 1.00000000

Hasil uji empiris metode IRLTS untuk model binomial disajikan dalam Table 7 dan Tabel 8 berikut.

Tabel 7. Nilai harapan (E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *exchangeable* untuk model GEE binomial

Method	Classical GEE			IRLTS		
	$\hat{\beta}_0$					
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	0.798286	0.000000	0.000000	-3.743076	0.000000	8.671788
1%	0.826341	0.314811	0.124382	0.129459	3.936774	0.788359
5%	0.755716	0.342582	0.125610	-0.068959	0.358534	0.667097
10%	0.800070	0.267484	0.089818	0.199516	0.495592	0.462195
20%	1.080128	0.298918	0.261765	0.448911	0.324755	0.152118
30%	0.976034	0.282935	0.176856	0.530003	0.462669	0.232260
40%	1.148409	0.416390	0.407160	0.698038	0.199287	0.040813
	$\hat{\beta}_1$					
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	0.887979	0.000000	0.000000	5.317442	0.000000	19.620140
1%	0.900073	0.463693	0.219927	9.488864	0.463693	356.529651
5%	1.007786	0.337339	0.115135	1.342950	0.337339	0.351971
10%	0.824783	0.326249	0.127883	1.403068	0.326249	0.553849
20%	1.108860	0.501206	0.270152	9.667001	0.501206	359.618618
30%	1.158490	0.512883	0.298119	13.748455	0.512883	524.377095
40%	1.227528	0.471894	0.288377	9.673974	0.471894	350.300052
	$\hat{\beta}_2$					
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.008863	0.000000	0.000000	5.448792	0.000000	19.712970
1%	0.921239	0.175044	0.031643	1.547949	3.941664	0.604805
5%	1.002532	0.196164	0.038621	1.701675	0.263213	0.541616
10%	0.995142	0.075742	0.006108	1.481478	0.302178	0.309465
20%	0.936824	0.192072	0.042911	1.510147	0.333506	0.356983
30%	1.008159	0.267506	0.071598	1.388333	0.381452	0.285325
40%	0.976736	0.255019	0.066454	1.349414	0.235920	0.167886

Tabel 8. Nilai harapan (E), galat baku (SE) dan kuadrat tengah galat (KTG) dengan matriks korelasi *autoregressive-1* untuk model GEE binomial

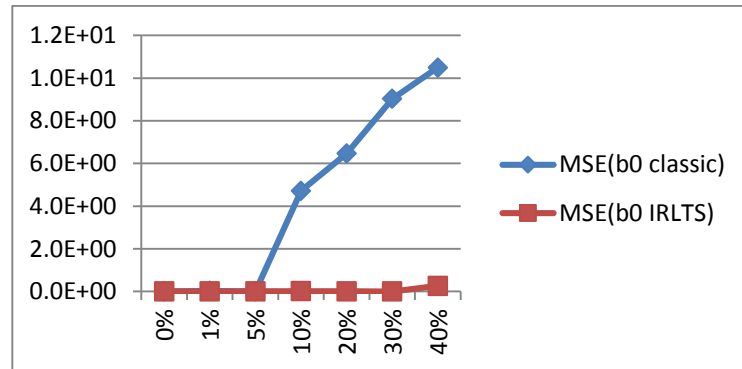
Classical GEE				IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
0%	1.74467	0.00000	0.00000	-5.42067	0.00000	51.34212
1%	2.11300	0.80956	0.58157	-6.65737	21.05245	443.20573
5%	2.44271	0.85109	1.05889	-6.97260	13.84667	269.82120
10%	2.16186	0.92322	0.94087	-7.95193	15.21610	327.88853
20%	0.62640	9.57975	93.30400	-10.10238	28.30811	944.55112
30%	2.76878	1.27944	2.45502	-7.13136	34.31551	1258.47654
40%	-0.82634	9.65979	100.55122	-11.76560	16.96862	473.70852
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
0%	1.63576	0.00000	0.00000	3.10670	0.00000	2.16366
1%	1.44579	1.32431	2.05496	-14.90069	45.82162	209.96211
5%	0.79594	1.72154	2.96848	2.17247	3.45786	13.66617
10%	1.33547	1.53931	2.59078	1.48661	2.45174	6.39737
20%	-0.63493	27.99170	100.85259	0.13308	5.20460	27.62360
30%	0.79108	1.32836	1.77000	-1.84077	8.11408	73.15970
40%	2.30069	2.60340	8.83876	3.66669	3.61055	20.88529
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
0%	1.08464	0.00000	0.00000	7.96662	0.00000	47.36170
1%	1.00526	0.37029	0.18626	232.74806	736.01398	54171.65802
5%	1.10350	0.53972	0.31014	10.31821	13.79839	272.79599
10%	1.09107	0.49412	0.26656	11.39512	15.54519	344.76425
20%	4.14569	10.03358	109.11142	19.74334	19.51259	723.08727
30%	1.01415	0.50207	0.30342	17.85189	29.36697	1138.34921
40%	4.21358	9.90539	106.95459	15.00391	17.33806	490.03300

### 5.3. Pembahasan

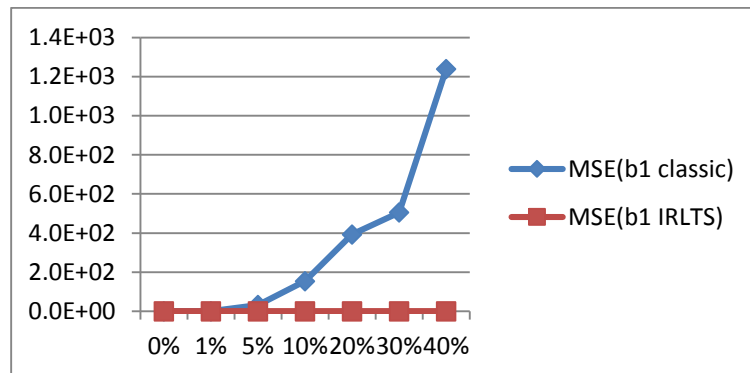
#### 5.3.1. Perilaku Metode IRLTS pada Model Normal

Untuk model GEE normal, berdasarkan hasil pada Tabel 1 di atas dapat disimpulkan bahwa metode IRLTS yang kami kembangkan ini memiliki kemampuan yang jauh lebih baik daripada metode klasik dalam mengatasi masalah pencilan pada data pada

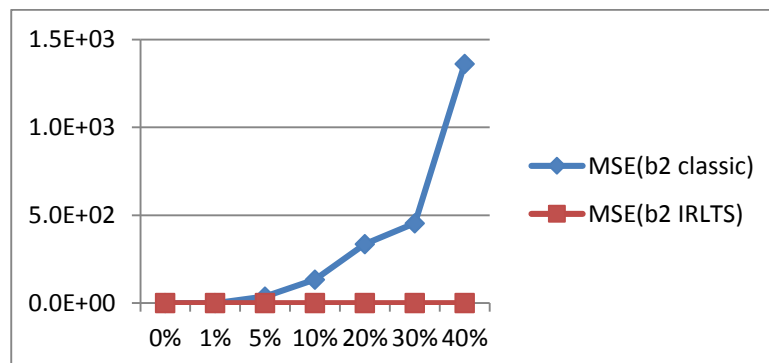
kasus model berdistribusi normal. Untuk lebih jelas perbandingan MSE kedua metode dapat dilihat pada Gambar 1- Gambar 6 di bawah. Pada ketiga gambar tersebut terlihat bahwa MSE dugaan IRLTS cenderung stabil dan tidak terpengaruh dengan jumlah pencilan. Sementara MSE dugaan klasik terpengaruh pencilan, yaitu semakin banyak jumlah pencilan maka nilai MSE metode klasik semakin meningkat.



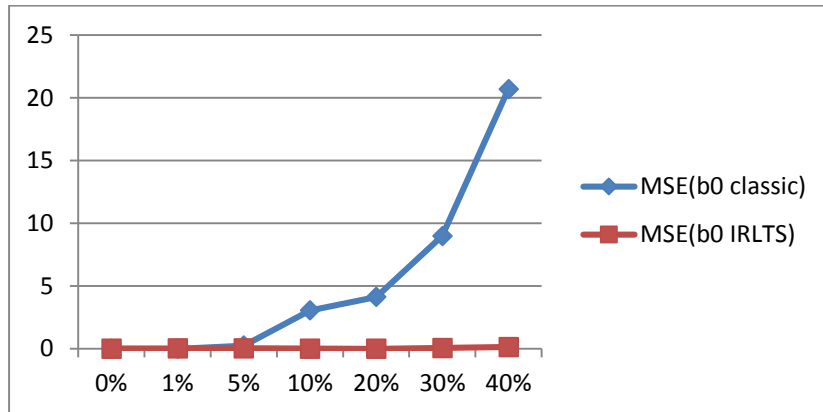
Gambar 1. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *exchangeable* model normal



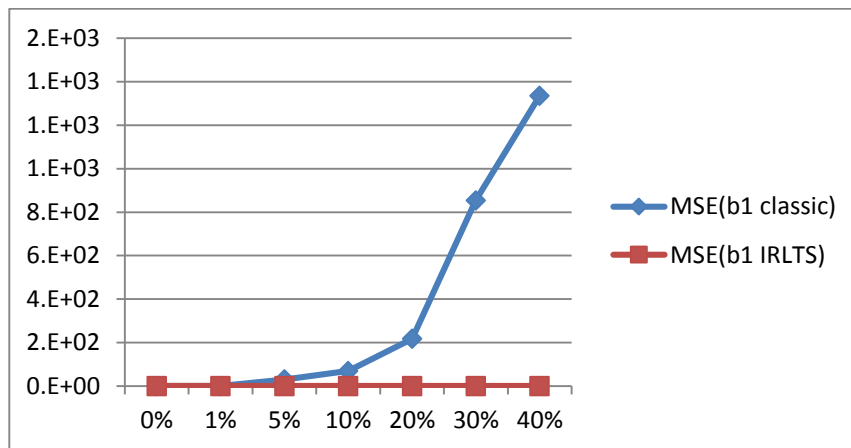
Gambar 2. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *exchangeable* model normal



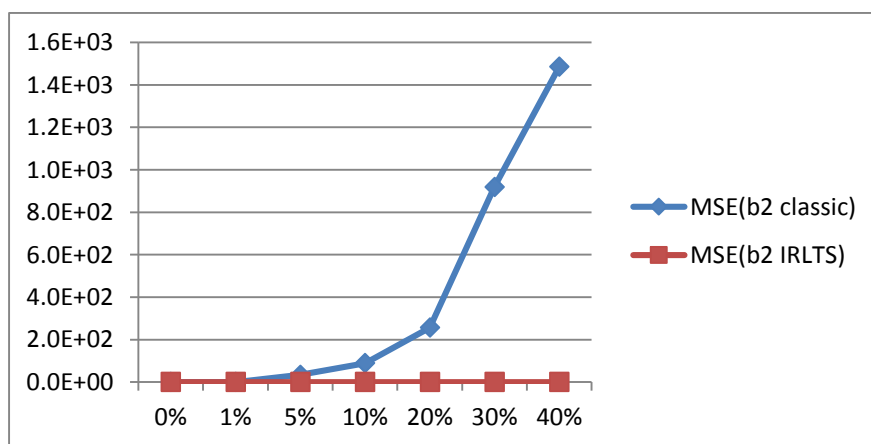
Gambar 3. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *exchangeable* model normal



Gambar 4. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *autoregressive-1* model normal



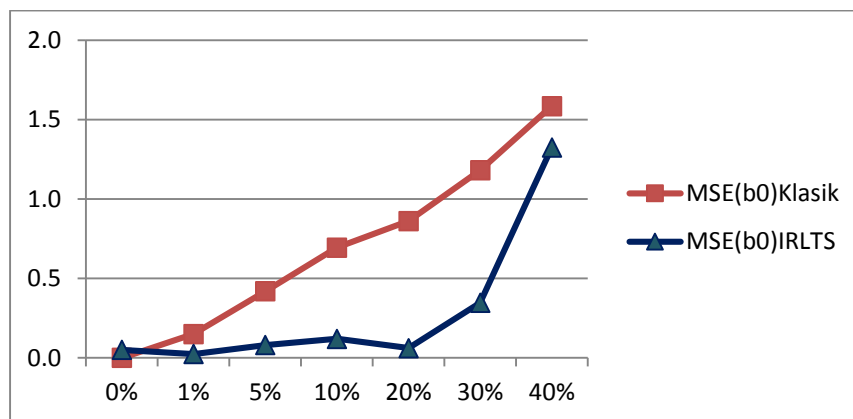
Gambar 5. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *autoregressive-1* model normal



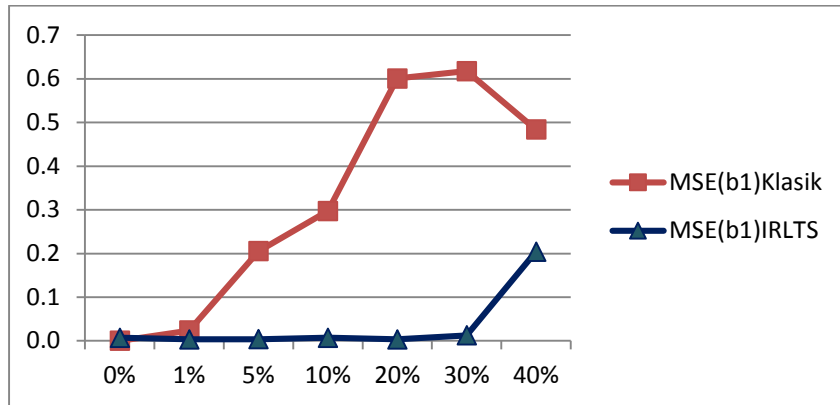
Gambar 6. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *autoregressive-1* model normal

### 5.3.2. Perilaku Metode IRLTS pada Model Gamma

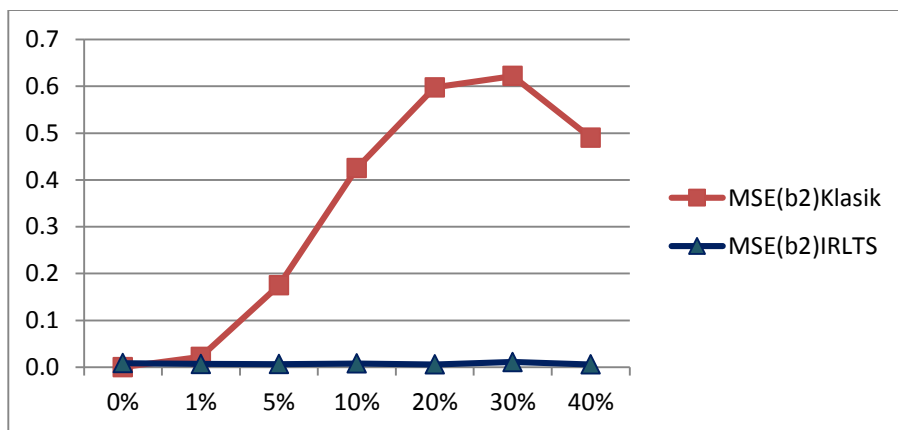
Pada uji empiris metode IRLTS untuk model GEE distribusi gamma data awal tanpa pencilan merupakan data populasi dengan dugaan koefisien model merupakan parameter model GEE. Dengan penetapan ini maka nilai pemberian kontaminasi sebesar 0% tidak akan merubah data awal, sehingga nilai dugaan koefisiennya selalu sama untuk setiap replikasi yang mengakibatkan nilai galat baku menjadi nol. Hal ini juga mengakibatkan nilai KTG metode GEE klasik (IRLS) sama dengan nol, sedangkan KTG metode IRLTS merupakan kuadrat dari selisih antara nilai parameter dengan dugaan IRLTS. Berdasarkan hasil pada Tabel 3 dan 4 dapat disimpulkan bahwa metode IRLTS juga memiliki kemampuan yang jauh lebih baik daripada metode klasik dalam mengatasi masalah pencilan pada data pada kasus model berdistribusi gamma. Perbandingan MSE kedua metode dapat dilihat pada Gambar 7- Gambar 12 di bawah. Pada ketiga gambar tersebut terlihat bahwa MSE dugaan IRLTS tidak terpengaruh dengan jumlah pencilan. Sementara MSE dugaan klasik terpengaruh oleh pencilan, yaitu semakin banyak jumlah pencilan maka nilai MSE metode klasik semakin meningkat. Dalam hal ini maka kami simpulkan bahwa metode IRLTS cukup *robust* dalam menghadapi pencilan pada model gamma.



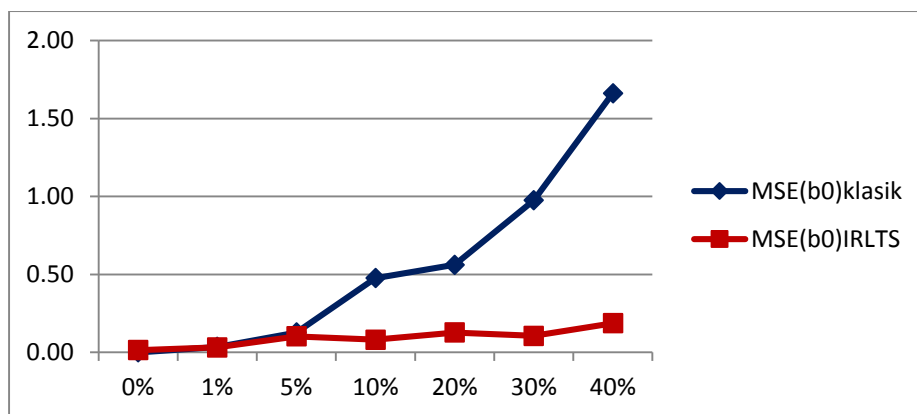
Gambar 7. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *exchangeable* model gamma



Gambar 8. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *exchangeable* model gamma

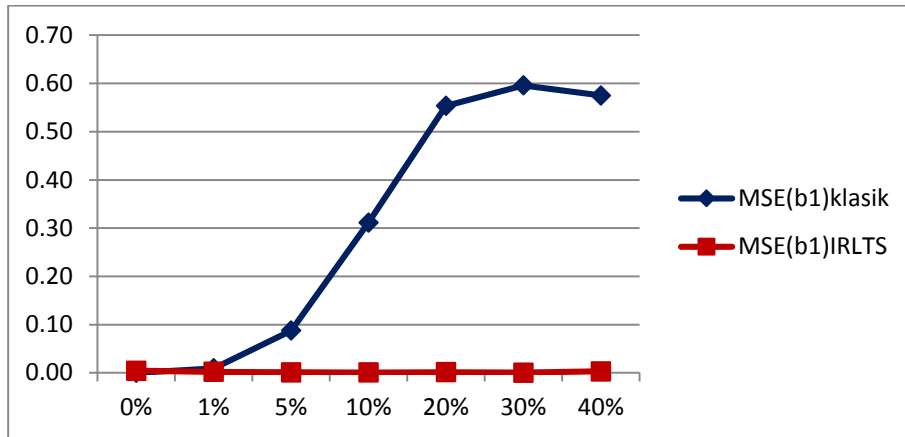


Gambar 9. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *exchangeable* model gamma

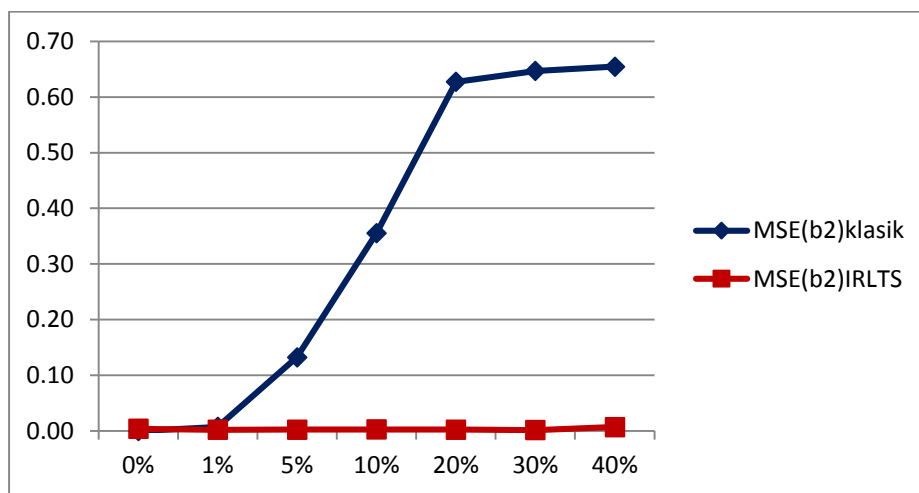


Gambar 10. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *autoregressive-1* model gamma





Gambar 11. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *autoregressive-1* model gamma

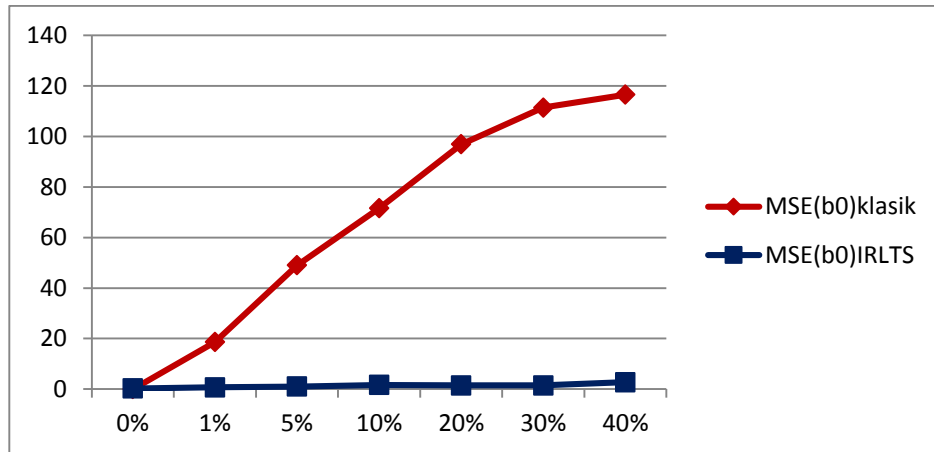


Gambar 12. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *autoregressive-1* model gamma

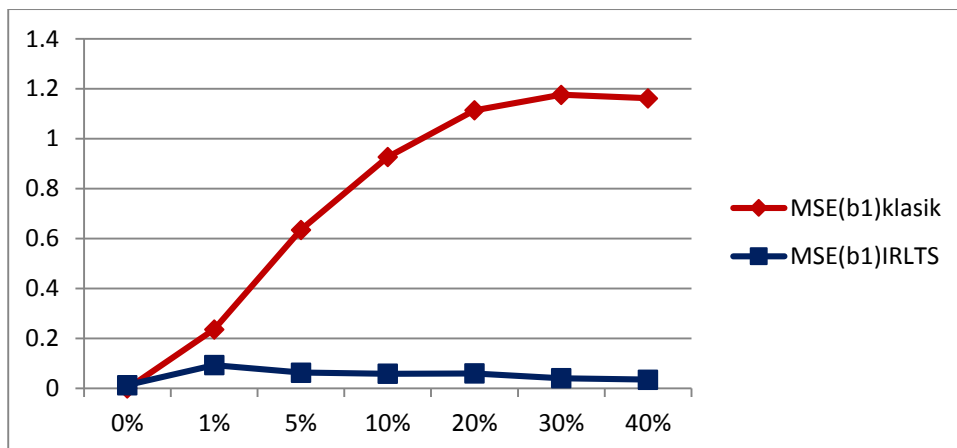
### 5.3.3. Perilaku Metode IRLTS pada Model Poisson

Serupa dengan hasil uji empiris metode IRLTS untuk model gamma, untuk model Poisson ini data awal tanpa pencilan merupakan data populasi dengan dugaan koefisien model merupakan parameter model GEE. Dengan penetapan ini maka nilai KTG metode GEE klasik (IRLS) sama dengan nol, sedangkan KTG metode IRLTS merupakan kuadrat dari selisih antara nilai parameter dengan dugaan IRLTS. Berdasarkan hasil pada Tabel 5 dan 6 dapat disimpulkan bahwa metode IRLTS juga memiliki kemampuan yang jauh lebih baik daripada metode klasik dalam mengatasi

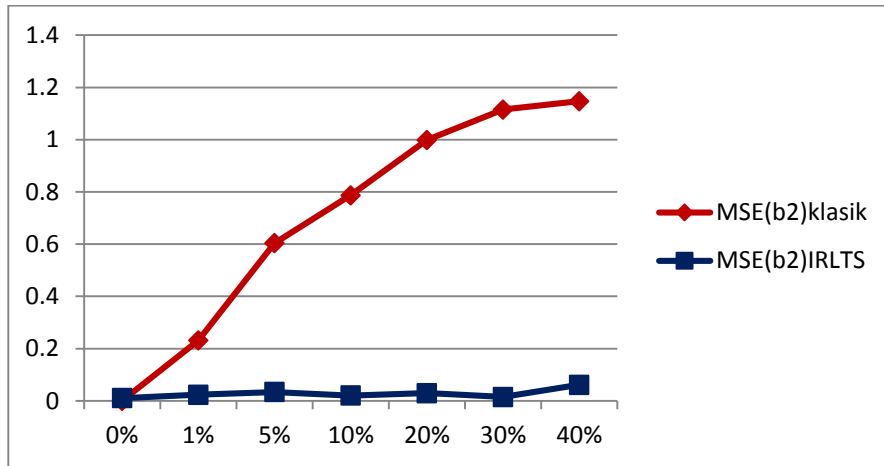
masalah pencilan pada data. Perbandingan MSE kedua metode disajikan dalam bentuk grafik pada Gambar 13 - Gambar 18 di bawah. Perbandingan perilaku MSE dari kedua metode untuk model Poisson ini mirip dengan model gamma. Dalam hal ini maka kami simpulkan bahwa metode IRLTS cukup *robust* dalam menghadapi pencilan pada model Poisson.



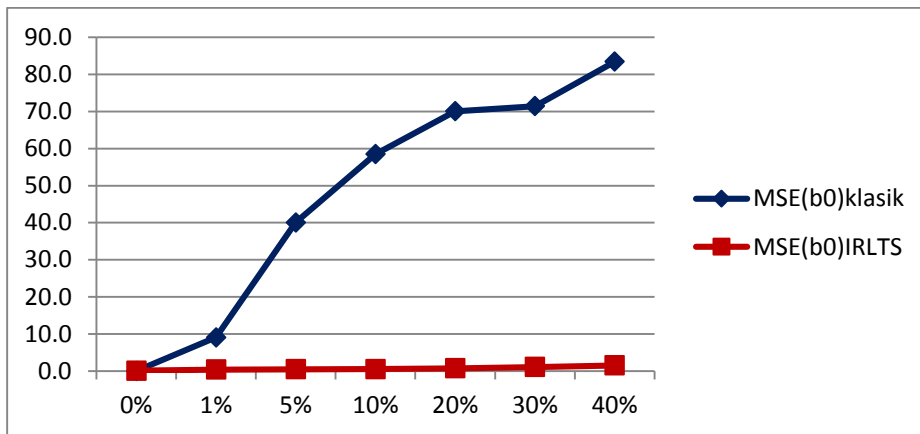
Gambar 13. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *exchangeable* model Poisson



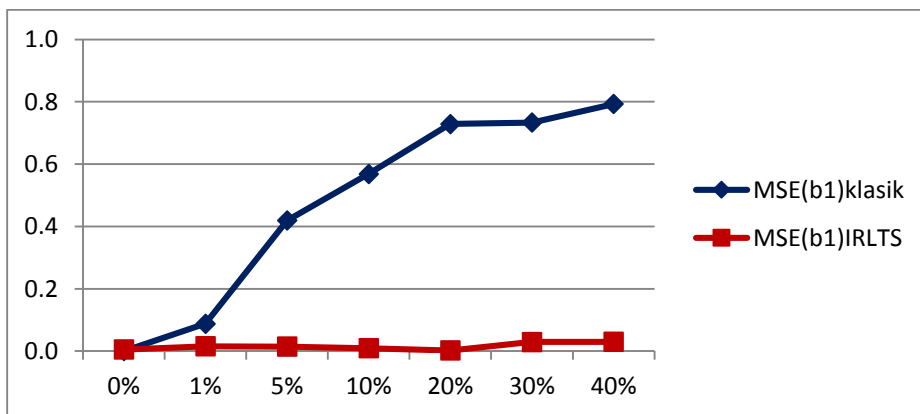
Gambar 14. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *exchangeable* model Poisson



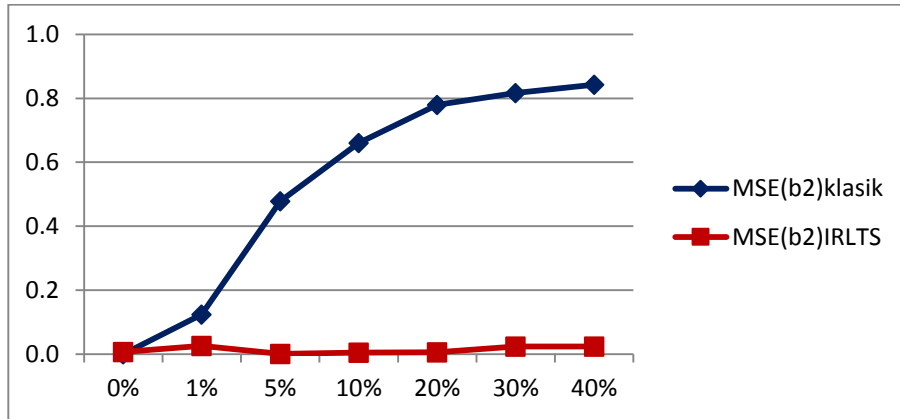
Gambar 15. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *exchangeable* model Poisson



Gambar 16. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *autoregressive-1* model Poisson



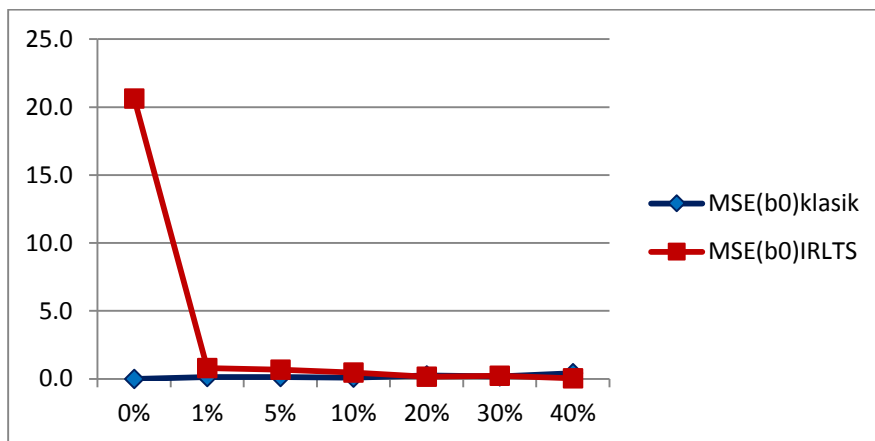
Gambar 17. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *autoregressive-1* model Poisson



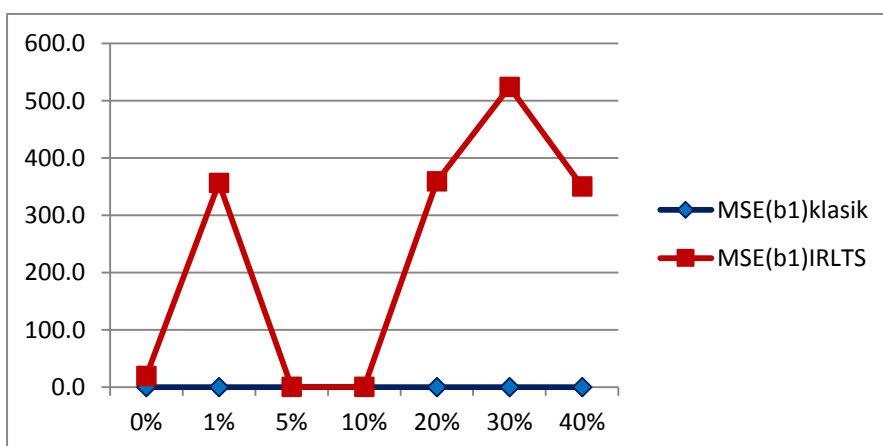
Gambar 18. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *autoregressive-I* model Poisson

#### 5.3.4. Perilaku Metode IRLTS pada Model Binomial

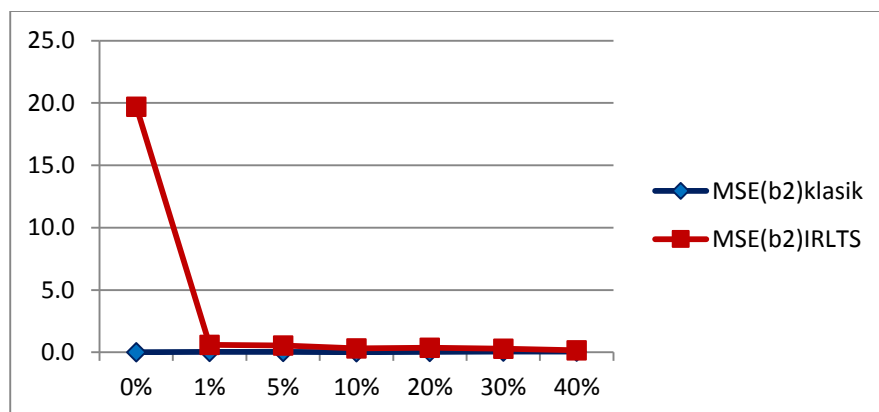
Berdasarkan hasil pada Tabel 7 dan 8, kami menyimpulkan bahwa metode IRLTS memberikan hasil yang buruk untuk pendugaan model GEE binomial. Hal ini dapat dilihat dari nilai penduganya dimana nilai MSE untuk setiap kasus pencilan tidak stabil dan bahkan memiliki nilai yang sangat besar untuk beberapa dugaan koefisien model. Perilaku dari metode IRLTS tidak dapat diprediksi, nilainya sangat berfluktuasi tanpa pola seragam. Menurut peneliti hal ini disebabkan oleh penggunaan galat model sebagai alat seleksi data pada algoritma IRLTS. Setiap replikasi menghasilkan peluang galat yang berbeda, sehingga memberikan nilai dugaan yang tidak konsisten. Nilai MSE kedua metode disajikan dalam bentuk grafik pada Gambar 19 - Gambar 24. Dari gambar-gambar tersebut terlihat bahwa metode IRLTS memberikan hasil yang buruk pada pendugaan model GEE binomial.



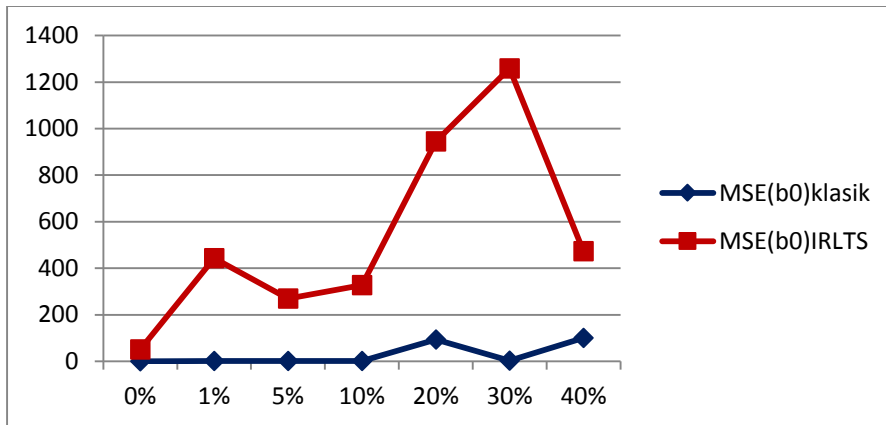
Gambar 13. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *exchangeable* model binomial



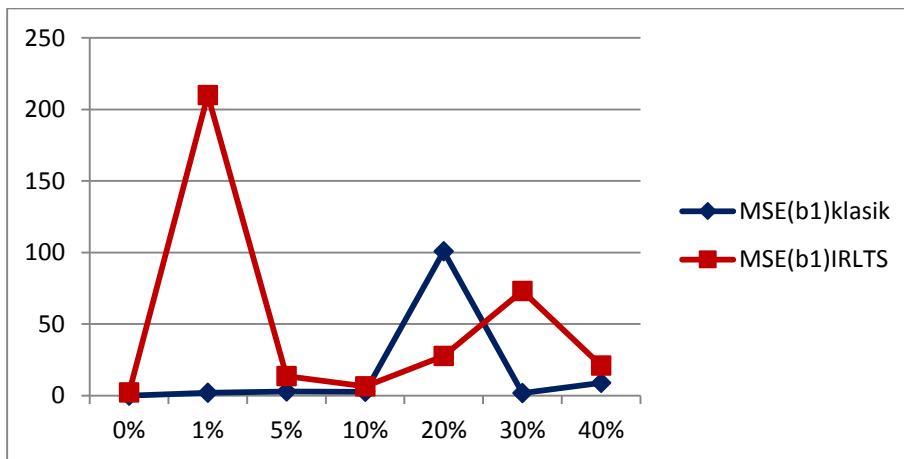
Gambar 14. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *exchangeable* model binomial



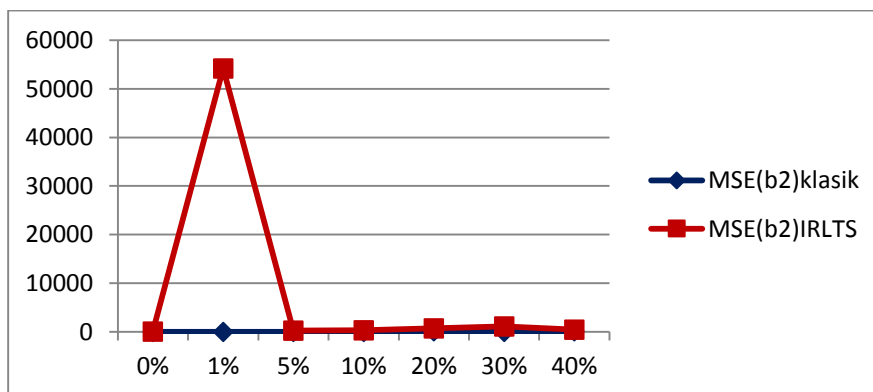
Gambar 15. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *exchangeable* model binomial



Gambar 16. Grafik nilai MSE untuk  $b_0 = \hat{\beta}_0$  pada kasus *autoregressive-1* model binomial



Gambar 17. Grafik nilai MSE untuk  $b_1 = \hat{\beta}_1$  pada kasus *autoregressive-1* model binomial



Gambar 18. Grafik nilai MSE untuk  $b_2 = \hat{\beta}_2$  pada kasus *autoregressive-1* model binomial

## 5.4. Sifat Asimtotik Penduga IRLTS

### 5.4.1. Normal Asimtotik

Jika  $n \rightarrow \infty$  dan  $h$  sangat besar, maka  $\hat{\beta}^{(IRLTS)}$  adalah normal asimtotik dengan mean sama dengan  $\beta$  dan matriks kovarians:

$$S = \left[ \sum_{i=1}^h D_i^T V_i^{-1} D_i \right]^{-1} \left\{ \sum_{i=1}^h D_i^T V_i^{-1} (Y - \hat{\mu}_i) (Y - \hat{\mu}_i)^T V_i^{-1} D_i \right\}_i \times \left[ \sum_{i=1}^h D_i^T V_i^{-1} D_i \right]^{-1}$$

dimana  $D_i = \frac{\partial \mu_i}{\partial \beta}$ . Atau dapat ditulis :

$$\sqrt{n}(\hat{\beta}^{IRLTS} - \beta) \sim N(\mathbf{0}, S)$$

Sifat normal asimtotik ini diperoleh dengan menggunakan sifat asimtotik IRLS (Liang and Zeger, 1986) untuk seluruh pengamatan  $n$  dari sampel yang digunakan dalam pendugaan.

### 5.4.2. Konsistensi

Kekonsistenan satu penduga diukur berdasarkan kuadrat tengah galatnya. Untuk metode IRLTS KTGnya sebagai berikut:

$$\begin{aligned} \text{KTG}(\hat{\beta}^{IRLTS}) &= E(\hat{\beta}^{IRLTS} - \beta)^2 + \text{Var}(\hat{\beta}^{IRLTS}) \\ &= \mathbf{0}^2 + \left[ \sum_{i=1}^n D_i^T V_i^{-1} D_i \right]^{-1} \left\{ \sum_{i=1}^n D_i^T V_i^{-1} (Y - \hat{\mu}_i) (Y - \hat{\mu}_i)^T V_i^{-1} D_i \right\}_i \times \left[ \sum_{i=1}^n D_i^T V_i^{-1} D_i \right]^{-1} \\ &= \left[ \sum_{i=1}^n D_i^T V_i^{-1} D_i \right]^{-1} \left\{ \sum_{i=1}^n D_i^T V_i^{-1} (Y - \hat{\mu}_i) (Y - \hat{\mu}_i)^T V_i^{-1} D_i \right\}_i \times \left[ \sum_{i=1}^n D_i^T V_i^{-1} D_i \right]^{-1} \end{aligned}$$

Sehingga  $\text{KTG}(\hat{\beta}^{IRLTS}) = S$ .

Untuk  $h \rightarrow \infty$ , perlu ditunjukkan bahwa  $\text{KTG}(\hat{\beta}^{IRLTS})$  akan mendekati nol. Yaitu :

$\lim_{h \rightarrow \infty} \text{KTG}(\hat{\beta}^{IRLTS}) = \mathbf{0}$ . Pembuktian sifat ini akan dikaji lebih lanjut.

## BAB 6

### RENCANA TAHAPAN BERIKUTNYA

Rencana tahapan berikutnya (tahun kedua) dari penelitian ini adalah membandingkan metode IRLTS dengan dua metode robust GEE lain yaitu

1. Iterated Reweighted Rank-Based Estimates (Abebe dkk, 2014)
2. Resistant estimator of GEE (Qaqish dan Preisser, 1999)

Metode *Iterated Reweighted Rank-Based Estimates* (IRRE) dari Abebe dkk merupakan suatu GEE *robust* dengan menggunakan pendugaan berbasis iterasi rank terboboti. Sedangkan penduga Resistant estimator of GEE (REGEE) dari Qaqish dan Preisser merupakan sebuah metode GEE *robust* menggunakan pendugaan tipe-M dengan melibatkan penurunan bobot (*downweighting*) titik-titik data yang berpengaruh.

Pembandingan akan dilakukan secara empiris dengan simulasi data menggunakan software R. Rancangan simulasi yang kami lakukan sebagai berikut:

Membangkitkan data berukuran  $N=1000$  pengamatan saling berkorelasi dengan desain data terdiri dari 200 subjek dan 5 kali pengukuran berulang. Kami menggunakan model dengan dua variabel bebas  $X$ , maka model yang digunakan dalam simulasi adalah:

Model normal:  $\mu_i = \eta_i = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij}$

Model gamma :  $1/\mu_i = \eta_i = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij}$

Model Poisson :  $\log(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij}$ .

Kami akan membangkitkan variabel respon dari distribusi normal, gamma, dan Poisson. Untuk membangkitkan data berdistribusi gamma dan Poisson yang saling berkorelasi kami akan menggunakan teknik NORTA seperti pada penelitian yang sudah kami lakukan. Berikut tahapan pembangkitan data gamma dan Poisson yang saling berkorelasi menggunakan NORTA:



1. Tentukan matriks korelasi  $\mathbf{R}$  dengan struktur elemen sesuai dengan yang diinginkan.
2. Bangkitkan data normal multivariate dengan vektor mean  $\boldsymbol{\mu}=(0,0,\dots,0)^T$  dan matriks korelasi  $\boldsymbol{\Sigma}=\mathbf{R}$
3. Hitung nilai fungsi densitas kumulatif (CDF=*cumulative density function*) dari setiap pengamatan pada data normal multivariat yang diperoleh di langkah 2
4. Bangkitkan galat (error) berdistribusi gamma berdasarkan nilai kuantil-kuantil gamma dan Poisson yang bersesuaian dengan nilai CDF pada langkah 3 (konversi data normal ke gamma dan Poisson)
5. Bangkitkan  $X_1$  dan  $X_2$ , dalam simulasi ini kami menggunakan distribusi uniform(1,5) dan integer (1-5).
6. Bangkitkan  $Y=1/(1+ X_1 + X_2+ \text{error})$ .
7. Menduga model GEE untuk distribusi gamma dan Poisson terhadap data  $Y$ ,  $X_1$  dan  $X_2$ . Koefisien yang diperoleh dari pembangkitan data awal tanpa pencilan ini merupakan parameter  $\beta_i$  yang ditetapkan.
8. Berikan kontaminasi terhadap data awal dengan pencilan sebesar 5%, 10%, 20%, 30% dan 40% dari total jumlah data ( $n_{\text{total}}=1000$ ), pencilan dibangkitkan secara acak dari distribusi gamma(2,2) dan Poisson(5000) untuk masing-masing model.
9. Menduga model GEE terhadap data terkontaminasi pada langkah 8 menggunakan menggunakan metode IRLTS, IRRE, dan REGEE, simpan nilai dugaan parameter yang diperoleh dari kedua metode.
10. Ulangi langkah 8-9 sebanyak 1000 kali.
11. Hitung nilai ekspektasi, galat baku ( $se=\text{standard error}$ ) dan kuadrat tengah galat ( $mse=\text{mean square error}$ ) dari nilai dugaan parameter ketiga metode.

## BAB 7

## KESIMPULAN DAN SARAN

### 7.1. Kesimpulan

Algoritma IRLTS yang kami kembangkan ini membutuhkan dua iterasi pada perhitungannya, yaitu iterasi dalam pendugaan koefisien regresi menggunakan metode IRLS dan iterasi dalam menentukan subsampel terbaik berukuran  $h$  pada *Concentration-step*.

Berdasarkan hasil uji empiris dapat disimpulkan bahwa metode IRLTS yang kami kembangkan dapat mengatasi pencilan dengan baik pada model GEE berdistribusi normal, gamma, dan Poisson. Nilai galat baku yang rendah menunjukkan bahwa metode IRLTS adalah penduga yang efisien. Kekonsistenan penduga IRLTS juga cukup baik berdasarkan nilai KTG penduga IRLTS yang kecil dan jauh lebih rendah dibandingkan metode GEE klasik. Namun pada simulasi model gamma untuk kasus matriks *exchangeable*, dugaan metode IRLTS terlihat memburuk pada jumlah pencilan 40%.

Untuk model binomial, metode IRLTS memberikan hasil yang buruk dan tidak *robust* terhadap pencilan.

### 7.2. Saran

Berdasarkan hasil yang kami peroleh maka saran kami kepada para pengguna statistika untuk menggunakan metode IRLTS ini dalam pendugaan model GEE normal, gamma, dan Poisson jika terindakasikan data mengandung pencilan.

Algoritma IRLTS yang kami kembangkan ini membutuhkan waktu komputasi yang lebih lama dibandingkan dengan metode GEE klasik karena menggunakan dua iterasi. Maka studi mengenai algoritma cepat untuk pendugaan dengan metode IRLTS merupakan hal yang menarik untuk dikaji lebih lanjut.

## DAFTAR PUSTAKA

- Abebe, A., McKean, J. W. & Kloke, J. D., Bilgic, Y. (2014). Iterated Reweighted Rank-Based Estimates for GEE Models. Technical Report.
- Agullo, J., Croux, C., & Van Aelst, S. (2008), The Multivariate Least Trimmed Squares Estimator. *Journal of Multivariate Analysis*, 99(3), 311-318.
- Ballinger, G. A. (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods*, 7, 127-150.
- Bishop, J. , Die, D. & Wang, Y. (2000). A Generalized Estimating Equations Approach For Analysis of The Impact Of New Technology On A Trawl Fishery. *Australian & New Zealand Journal of Statistics*, 42, 159-177.
- Croux C. & Ruiz-Gazen A. (2005). High Breakdown Estimators for Principal Components: the Projection-Pursuit Approach Revisited. *Journal of Multivariate Analysis*, 95, 206-226.
- Demirtas, H. (2014). Joint Generation of Binary and Nonnormal Continuous Data. *Biometrics and Biostatistics*, S12 - 001.
- Ghisletta, P. & Spini, D. (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*, 29, 421–437.
- Ghosh, S. and Henderson, S. G. 2005. Corrigendum: Behaviour of the NORTA method for correlated random vector generation as the dimension increases. *ACM TOMACS*, 16, 93-94
- Gill, P.S. (2000). A Robust Mixed Linear Model Analysis for Longitudinal Data. *Statistics in Medicine*, 19, 975-987.
- Johnston, G. & Stokes, M. (1996). Repeated Measures Analysis With Discrete Data Using The SAS System. *SUGI Proceeding*. SAS Institute Inc., Cary, NC.
- Jung, S. H. & Ying, Z. (2003). Rank-Based Regression With Repeated Measurements Data, *Biometrika*, 90, 732-740.
- Kalina, J. (2012). Highly Robust Statistical Methods in Medical Image Analysis. *Biocybernetics and Biomedical Engineering*, 32 (2), 3-16.

- Lai, J. & Jiang X. (2013). Robust Face Recognition Using Trimmed Linear Regression. *IEEE International Conference In Acoustics, Speech, and Signal Processing (ICASSP)*. Vancouver, Canada.
- Liang, K.Y. & Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73, 13-22.
- Manel, L. A & DeRouen, T. A. (2001). A Covariance Estimator for GEE with Improved Small-Sample Properties. *Biometrics*, 57, 126-134.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- Minhajuddin, A.T.M., Harris, I. R., dan Schucany, W.R. (2004). Simulating Multivariate Distributions with Specific Correlations. *Journal of Statistical Computation and Simulation*, 74(8), 599-607.
- Nakai, M. & Ke, W. (2009). Statistical Models for Longitudinal Data Analysis. *Applied Mathematical Sciences*, 3(40), 1979 – 1989.
- Niaki, S.T.A. dan Abbasi, B. (2008) Generating Correlation Matrices for Normal Random Vectors in NORTA Algorithm Using Artificial Neural Networks. *Journal of Uncertain Systems*, 2 (3), 192-201.
- Niavarani, M.R. & Smith, A.J.R. (2013). Modeling and Generating Multi-Variate-Attribute Random Vectors Using a New Simulation Method Combined with NORTA Algorithm. *Journal of Uncertain Systems*, 7 (2), 83-91.
- Norton, E. C., Bieler, G. S., Ennett, S. T., Zarkin, G. A. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations. *Journal of Consulting and Clinical Psychology*, 64, 919-926.
- Prentice, R. L. & Zhao, L. P. (1991). Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses. *Biometric*, 47, 825-839.
- Qaqish, B. F. & Preisser, J. S. (1999). Resistant fits for regression with correlated outcomes an estimating equations approach. *Journal of Statistical Planning and Inference*, 75, 15-431.

- R Development Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna. <http://www.R-project.org>.
- Roosbeh, M. & Babaie-Kafaki, S. (2015). Extended Least Trimmed Square Estimator in Semiparametric Regression Models with Correlated Errors. *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655.2015.1014371.
- Rousseeuw, R. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
- Rousseeuw, P. J., dan van Driessen, K. (2006). Computing LTS Regression for Large Data Sets. *Data mining and Knowledge Discovery*, 12, 29-45.
- Sim, D. G., Kwon O.K., Park R. H. (1999) Object Matching Algorithms Using Robust Hausdorff Distance Measures. *IEEE Transactions On Image Processing*, 8 (3), 425-2429.
- Somorjai , R. L., Dolenko, B. & Baumgartner, B. (2003). Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics*, 19 (12), 1484-1491.
- Visek, J. A. (2006a). The Least Trimmed Squares. Part II:  $\sqrt{n}$ -consistency. *Kybernetika*, 42, 181-202.
- Visek, J. A. (2006b). The Least Trimmed Squares. Part III: Asymptotic normality. *Kybernetika*, 42, 203-224.
- Wang, Y.-G. and Zhu, M. (2006). Rank-based regression for analysis of repeated measures. *Biometrika*, 93, 459-464.
- Yahav, I dan Shmueli, G. (2012). On generating multivariate poisson data in management science applications. *Applied Stochastic Models in Business and Industry*, 28(1), 91-102
- Zorn, C. J. W. (2001). Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *American Journal of Political Science*, 45(2), 470-490.

**LAMPIRAN**

# Robust Estimation of Generalized Estimating Equation when Data Contain Outliers

Khoirin Nisa and Netti Herawati

**Abstract**—In this paper, a robust procedure for estimating parameters of regression model when generalized estimating equation (GEE) applied to longitudinal data that contains outliers is proposed. The method is called ‘iteratively reweighted least trimmed square’ (IRLTS) which is a combination of the iteratively reweighted least square (IRLS) and least trimmed square (LTS) methods. To assess the proposed method a simulation study was conducted and the result shows that the method is robust against outliers.

**Keywords**—GEE, IRLS, LTS, longitudinal data, regression model.

## I. Introduction

LONGITUDINAL studies are increasingly common in many scientific research areas, for example in the social, biomedical, and economical fields. In longitudinal studies, multiple measurements are taken on the same subject at different points in time. Thus, observations for the same subject are correlated. The analysis of data resulting from such studies often becomes complicated due to the within-subject correlation. This correlation must be considered for any appropriate analysis method.

Generalized linear models (GLM) as described by McCullagh and Nelder [1] is a standard method used to fit regression models for univariate data that are presumed to follow an exponential family distribution. The association between the response variable and the covariates is given by the link function. GLM assume that the observations are independent and do not consider any correlation between the outcome of the  $n$  observations. Liang and Zeger [2] introduced an approach to this correlation problem using GEE to extend GLM into a regression setting with correlated observations within subjects.

The GEE method of Liang and Zeger gives consistent estimators of the regression parameter. The parameter estimates are consistent even when the variance structure is misspecified under mild regularity conditions. However, problems can occur when data contain outliers. The method is not robust against outliers since it is based on score equations from the quasi likelihood method of estimation. The working correlation matrix would be affected by the outliers and also the parameter estimates. In this situation, we need a robust method that can minimize the effect of outliers.

In recent years, a few authors have considered robust methods for longitudinal data analysis. For example, Qaqish and Preisser [3] proposed a resistant version of the GEE using M-type estimation by involving down-weighting influential data points. Gill [4] proposed a

robustified likelihood based on multivariate normal distribution. Jung and Ying [5] proposed an adaptation of the Wilcoxon-Mann-Whitney method of estimating linear regression parameters for use in longitudinal data analysis under the working independence model. And recently, Abebe *et al.* [6] proposed a robust GEE using iterated reweighted rank-based estimation.

In this paper, we adopt the LTS [7] method for robust linear regression in the sense of trimming the data for estimating the regression coefficients so that the observations with high residuals are not included in the parameter estimation. In Section 2 we present a brief review of GEE. In Section 3 we describe our proposed method IRLTS. In Section 4 we discuss some results from our simulation study.

## II. Generalized Estimating Equation and IRLS method

Let  $Y_{ij}, j = 1, \dots, m_i, i = 1, \dots, n$  represent the  $j$ th measurement on the  $i$ th subject. There are  $m_i$  measurements on subject  $i$  and  $N = \sum_{i=1}^n m_i$  total measurements. Assume that the marginal distribution of  $y_{ij}$  is of the exponential class of distributions and is given by:

$$f(y, \theta, \phi) = \exp\{y\theta - b(\theta)/a(\phi) + c(y, \phi)\}$$

where  $a(\cdot)$ ,  $b(\cdot)$ , and  $c(\cdot)$  are given,  $\theta$  is the canonical parameter, and  $\phi$  is the dispersion parameter.

Let the vector of measurements on the  $i$ th subject be  $\mathbf{Y}_i = [Y_{i1}, \dots, Y_{im_i}]^T$  with corresponding vector of means  $\boldsymbol{\mu}_i = [\mu_{i1}, \dots, \mu_{im_i}]^T$  and  $\mathbf{X}_i = [X_{i1}, \dots, X_{im_i}]^T$  be the  $m_i \times p$  matrix of covariates. In general, the components of  $\mathbf{Y}_i$  are correlated but  $\mathbf{Y}_i$  and  $\mathbf{Y}_k$  are independent for any  $i \neq k$ . To model the relation between the response and covariates, we can use a regression model similar to the generalized linear models:

$$g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta}$$

where  $\boldsymbol{\mu}_i = E(\mathbf{Y}_i | \mathbf{X}_i)$ ,  $g$  is a specified link function, and  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^T$  is a vector of unknown regression coefficients to be estimated. The GEE for estimating the  $p \times 1$  vector of regression parameters  $\boldsymbol{\beta}$  is given by :

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} [\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] = 0 \quad (1)$$

where  $\mathbf{V}_i$  be the covariance matrix of  $\mathbf{Y}_i$  modeled as  $\mathbf{V}_i = \lambda \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$ ,  $\mathbf{A}_i$  is a diagonal matrix of variance function  $V(\mu_{ij})$ , and  $\mathbf{R}(\boldsymbol{\alpha})$  is the working correlation matrix of  $\mathbf{Y}_i$  indexed by a vector of parameters  $\boldsymbol{\alpha}$ . Solutions to (2) are obtained by alternating between estimation of  $\lambda$ ,  $\boldsymbol{\alpha}$  and  $\theta$ .

There are several specific choices of the form of working correlation matrix  $\mathbf{R}_i(\boldsymbol{\alpha})$  commonly used to model the correlation matrix of  $\mathbf{Y}_i$ . A few of the choices are shown below, one can refer to [1] for additional choices. The dimension of the vector  $\boldsymbol{\alpha}$ , which is treated as



a nuisance parameter, and the form of the estimator of  $\boldsymbol{\alpha}$  are different for each choice. Some typical choices are:

1.  $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{R}_0$ , a fixed correlation matrix. For  $\mathbf{R}_0 = \mathbf{I}$ , the identity matrix, the GEE reduces to the independence estimating equation.
2. Exchangeable:  $Cor(Y_{ij}, Y_{ik}) = \alpha$ ,  $j \neq k$ .
3. Autoregressive-1:  $Cor(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}$ .
4. Unstructured:  $Cor(Y_{ij}, Y_{ik}) = \alpha_{jk}$ .

Solving for  $\boldsymbol{\beta}$  is done with iteratively reweighted least squares (IRLS). The following is the algorithm for fitting the specified model using GEEs [3] :

5. Compute an initial estimate of  $\hat{\boldsymbol{\beta}}_{GEE}$ , for example with an ordinary generalized linear model assuming independence.
6. A current estimate  $\hat{\boldsymbol{\beta}}_{GEE}$  is updated by regressing the working response vector

$$\mathbf{Z}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$$

on  $\mathbf{X}$ . A new estimate  $\hat{\boldsymbol{\beta}}_{new}$  is obtained by :

$$\hat{\boldsymbol{\beta}}_{new} = (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* \mathbf{Z}^* \quad (2)$$

where  $\mathbf{W}^*$  is a block diagonal weight matrix whose  $i$ th block is the  $m_i \times m_i$  matrix

$$\mathbf{W}_i^* = \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^{-1} \mathbf{A}_i^{-1} \mathbf{R}_i(\hat{\boldsymbol{\alpha}}) \mathbf{A}_i^{-1} \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^{-1}.$$

7. Use  $\hat{\boldsymbol{\beta}}_{new}$  to update  $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}_{new} = \mathbf{H}\mathbf{Z}^*$ , where  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*$ .
8. Iterate until convergence.

### III. Iteratively Reweighted Least Trimmed Square Algorithm

First let us briefly recall that the robust estimation of regression parameters using LTS method is given by:

$$\hat{\boldsymbol{\beta}}_{LTS} = \arg \min \sum_{i=1}^h e_i^2 \quad (3)$$

which is based on the ordered absolute residuals  $|e_1| \leq |e_2| \leq \dots \leq |e_n|$ . LTS estimation is calculated by minimizing the  $h$  ordered squares residuals, where  $h$  can be chosen within  $\frac{n}{2} + 1 \leq h \leq \frac{3n}{4} + \frac{p+1}{4}$ , with  $n$  and  $p$  being sample size and number of parameters respectively. When  $h = [n/2]$ , LTS locates that half of the observations which has the smallest estimated variance. In that case, the breakdown point is 50%. When  $h$  is set to the sample size, LTS and ordinary least square (OLS) coincide.

In [7] Rousseeuw and Leroy shows  $n^{1/2}$  consistency and asymptotic normality of LTS in the location–scale model. Vížek [8] extends this to the regression model with random regressors, the proof for fixed regressors is in later series of his papers: [9][10].

When  $n$  is very small, it is possible to generate all subsets of size  $h$  to determine which one minimizes the LTS criterion.

Rousseeuw and Leroy computation of LTS based on subsets of size  $k$  requires  $q = \binom{n}{k}$  subsets which is usually still too large for realistic applications. When  $n$  is small enough:

1. Select  $h$ .
2. Generate all possible subsets with  $k$  observations, and compute the regression coefficients, say  $\hat{\beta}(1), \dots, \hat{\beta}(l)$ .
3. Compute the residuals using all  $n$  observations, and from this the LTS criterion.
4. The LTS estimate corresponds to the  $\hat{\beta}(l)$  that minimizes the objective function (3).

Rousseeuw and van Driessen [11] propose a fast algorithm for computing LTS. The trick is to iterate a few steps on a large number of starting values, and keep the 10 (say) most promising ones. These are then used for full iteration, yielding the final estimate. The resulting algorithm makes LTS estimation faster.

Our proposed procedure is a combination of IRLS and LTS methods. IRLTS estimator involves computing the hyperplane that minimizes the sum of the smallest  $h$  squared residuals and use the weighted least square estimation for  $\beta$  in each iteration. To motivate our estimator and following the fast LTS algorithm [11], it is convenient to write IRLTS algorithm with involving the residuals as follow.

Concentration-step:

7. Choose  $h$  observations.
8. Compute  $\hat{\beta}$  based on  $h$  observations using IRLS method .
9. Use the estimate  $\hat{\beta}$  to calculate residuals:  $e_{ij} = Y_{ij} - \hat{\mu}_{ij}$  based on equation  $\hat{\mu}_i = g^{-1}(\mathbf{X}_i \hat{\beta})$  of  $n$  observations.
10. Sort  $|e_{ij}|$  for  $j = 1, \dots, m_i, i = 1, \dots, n$  in ascending order:  $|e_{11}| \leq |e_{12}| \leq \dots \leq |e_{ij}|$ .
11. Choose  $h$  observations which have the lowest  $h$  residuals, we denote the  $h$  observations as subset  $H$ .

The repetitions of concentration-step will produce an iteration process.

IRLTS algorithm:

11. Choose  $h$  observations.
12. Compute  $\hat{\beta}$  based on  $h$  observations using IRLS by (2), we obtain  $\hat{\mu}_i = g^{-1}(\mathbf{X}_i \hat{\beta})$ .
13. Calculate residuals:  $e_{ij} = Y_{ij} - \hat{\mu}_{ij}$  of  $n$  observations.
14. Sort  $|e_{ij}|$  in ascending order:  $|e_{11}| \leq |e_{12}| \leq \dots \leq |e_{ij}|$
15. Choose  $h_1$  observations which have the lowest  $h_1$  residuals, we denote as subset  $H_1$ .
16. Run concentration-step on  $H_1$  twice, and we obtain  $H_1^*$ .
17. Repeat step 1- step 6 for  $\binom{n}{h}$  times
18. From the  $\binom{n}{h}$  results, choose the best 10 subsets  $H_q, q=1, \dots, 10$ .

19. Run concentration-step on the best 10 subsets  $H_q$  until convergence.
20. Choose the best subset  $H$ .

#### IV. Simulation Study

To look at the performance of the proposed method, we have done a simulation study by generating  $N=1000$  observations from 200 subjects with 5 repeated measures. The model for data generation is as follows:

$$u_{ij} = \beta_0 + \beta_1 x_{ij}$$

where  $\beta_0 = \beta_1 = 1$ ,  $i=1,2,\dots, 200$  and  $j=1,2,\dots,5$ . The covariates  $x_{ij}$  are i.i.d. from a uniform distribution  $\text{Unif}(1,5)$ . For this longitudinal data the normal distributed model is used. We generated data based on the underlying true correlation structures as exchangeable (EXCH) and autoregressive-1 (AR1) with  $\alpha=0.3$  and  $0.7$ . We considered data without outliers ( $\varepsilon = 0\%$ ) as well as contaminated data ( $\varepsilon = 10\%$ ,  $20\%$  and  $30\%$ ). The contamination is generated from normal distribution  $N(100,1)$ , we set two cases for the contamination, i.e. randomly spread over the sample (case A) and randomly spread over the half upper  $x_{ij}$  values of the sample (case B). For each scenario 1000 Monte Carlo data sets were generated. We evaluated the results using relative efficiency (RE) of IRLTS to IRLS and the mean square error (MSE) of  $\hat{\beta}$  which we define as

$$RE_{IRLTS/IRLS} = \text{Var}(\hat{\beta}_i^{IRLTS}) \{ \text{Var}(\hat{\beta}_i^{IRLS}) \}^{-1}$$

and

$$MSE = \frac{1}{1000} \sum_{s=1}^{1000} \left( \hat{\beta}_i^{(s)} - \beta_i \right)^2, \text{ with } i = 0,1,$$

where  $\text{Var}(\cdot)$  is the variance. We provide the expected values (E), and the relative efficiency resulted from our simulation in Table I - Table IV and the MSEs in Table V- Table VI.

The efficiency of IRLTS and IRLS for clean data (i.e. when  $\varepsilon = 0\%$ ) is almost equal since  $RE \sim 1$  for each case, but IRLTS is more efficient than IRLS when data contain outliers. The parameter estimates of IRLS are much more influenced by the outliers than the parameter estimates of IRLTS. From the expected values we can see that the the more outliers contained in the data the larger the deviation of IRLS estimates from the parameter (i.e.  $\beta_0 = \beta_1 = 1$ ), while the parameter estimates of IRLTS are almost stable and close to the parameter.

TABLE I  
SIMULATION RESULT FOR LONGITUDINAL DATA WITH EXCHANGEABLE CORRELATION MATRIX WITH  $\alpha = 0.3$

Case	Coeff.	$\varepsilon$	$E(\hat{\beta}_{IRLS})$	$E(\hat{\beta}_{IRLTS})$	$RE_{(IRLTS/IRLS)}$
Case A		0%	1.00500	1.00499	1.00178

		10%	10.08781	1.02474	0.00294
	$\hat{\beta}_0$	20%	17.55050	1.08504	0.00401
		30%	23.73423	1.15940	0.00657
		0%	0.99846	0.99848	1.00278
	$\hat{\beta}_1$	10%	0.99604	0.99948	0.00267
		20%	0.98073	0.99775	0.00393
		30%	0.94544	0.99577	0.00649
Case B		0%	1.02179	1.01050	0.99466
	$\hat{\beta}_0$	10%	-5.20456	1.00195	0.00577
		20%	-10.30572	0.95625	0.01606
		30%	-14.27400	0.84177	0.03562
		0%	0.99318	0.99693	0.99369
	$\hat{\beta}_1$	10%	5.95079	0.98675	0.00483
		20%	9.99122	0.99425	0.05797
		30%	13.21288	1.04650	0.11281

TABLE II  
SIMULATION RESULT FOR LONGITUDINAL DATA WITH EXCHANGEABLE CORRELATION MATRIX WITH  $\alpha = 0.7$

Case	Coeff.	$\varepsilon$	$E(\hat{\beta}_{IRLS})$	$E(\hat{\beta}_{IRLTS})$	$RE_{(IRLTS/IRLS)}$
Case A		0%	1.01266	1.01187	1.01763
	$\hat{\beta}_0$	10%	9.93676	1.01672	0.00433
		20%	17.31940	1.05446	0.00505
		30%	23.68107	1.16184	0.00769
		0%	0.99600	0.99613	1.01882
	$\hat{\beta}_1$	10%	1.04154	1.00103	0.00403
		20%	1.05223	1.00661	0.00478
		30%	0.95609	0.99443	0.00770
Case B		0%	1.05174	1.01922	0.99052
	$\hat{\beta}_0$	10%	-5.20125	1.00911	0.00837
		20%	-10.28303	0.98818	0.00528
		30%	-14.25209	0.83684	0.04304
		0%	0.98159	0.99245	0.99948
	$\hat{\beta}_1$	10%	5.93523	0.98438	0.00722
		20%	9.96316	0.97828	0.00489
		30%	13.21954	1.05368	0.12544

TABLE III  
SIMULATION RESULT FOR LONGITUDINAL DATA WITH AUTOREGRESSIVE-1 CORRELATION MATRIX WITH  $\alpha = 0.3$

Case	Coeff.	$\varepsilon$	$E(\hat{\beta}_{IRLS})$	$E(\hat{\beta}_{IRLTS})$	$RE_{(IRLTS/IRLS)}$
Case A		0%	0.99848	0.99744	1.03077

		10%	10.04370	1.02103	0.00247
	$\hat{\beta}_0$	20%	17.60751	1.07994	0.00359
		30%	23.63332	1.15166	0.00621
		0%	1.00043	1.00079	1.04004
	$\hat{\beta}_1$	10%	1.00954	1.00069	0.00217
		20%	0.96174	0.99879	0.00356
		30%	0.97699	0.99897	0.00625
Case B		0%	1.01800	1.00963	1.00481
	$\hat{\beta}_0$	10%	-5.40516	0.99210	0.00113
		20%	-10.07597	0.94809	0.00393
		30%	-14.18714	0.85628	0.05029
		0%	0.99341	0.99621	1.00108
	$\hat{\beta}_1$	10%	5.98712	0.99180	0.00541
		20%	9.89834	0.99916	0.01079
		30%	13.19889	1.04016	0.16313

TABLE IV  
SIMULATION RESULT FOR LONGITUDINAL DATA WITH AUTOREGRESSIVE-I CORRELATION MATRIX WITH  $\alpha = 0.7$

Case	Coeff.	$\varepsilon$	$E(\hat{\beta}_{IRLS})$	$E(\hat{\beta}_{IRLTS})$	$RE_{(IRLTS/IRLS)}$
Case A	$\hat{\beta}_0$	0%	1.00024	1.00079	1.07947
		10%	9.93814	1.00872	0.00412
		20%	17.52048	1.07392	0.00486
		30%	23.77262	1.16751	0.00790
	$\hat{\beta}_1$	0%	0.99960	0.99939	1.06915
		10%	1.04489	1.00404	0.00379
		20%	0.98595	1.00065	0.00481
		30%	0.94520	0.99331	0.00785
Case B	$\hat{\beta}_0$	0%	1.03666	1.01330	0.99589
		10%	-5.18276	1.01138	0.00618
		20%	-10.37403	0.97097	0.00484
		30%	-14.34357	0.80114	0.04723
	$\hat{\beta}_1$	0%	0.98765	0.99545	0.99343
		10%	5.92778	0.98511	0.00464
		20%	10.00611	0.98523	0.00461
		30%	13.25732	1.07041	0.14581

The consistency of the estimators is assessed through their MSEs (see Table V and Table VI). When data contain outliers, the MSEs of IRLTS are relatively small compared to the MSEs of the classical GEE (IRLS). From the result we conclude that IRLTS is robust against outliers.

TABLE V  
MEAN SQUARE ERROR OF PARAMETER ESTIMATES FOR DATA WITH EXCHANGEABLE CORRELATION MATRIX

Case	Coeff.	$\varepsilon$	$\alpha = 0.3$		$\alpha = 0.7$	
			IRLS	IRLTS	IRLS	IRLTS
Case A	$\hat{\beta}_0$	0%	0.01664	0.01666	0.03141	0.03194
		10%	90.44780	0.02372	88.48682	0.03765
		20%	291.41893	0.07749	283.90429	0.09182
		30%	544.84787	0.20944	540.15839	0.22414
	$\hat{\beta}_1$	0%	0.00165	0.00165	0.00299	0.00305
		10%	0.87157	0.00233	0.94041	0.00378
		20%	1.91855	0.00754	1.92692	0.00924
		30%	3.08263	0.02001	2.82052	0.02174
Case B	$\hat{\beta}_0$	0%	0.01726	0.01680	0.03467	0.03206
		10%	41.61809	0.01801	41.89402	0.02886
		20%	136.26362	0.13754	133.06421	0.03051
		30%	240.63699	0.28655	240.56976	0.36852
	$\hat{\beta}_1$	0%	0.00171	0.00167	0.00346	0.00318
		10%	24.88252	0.00197	24.75140	0.00309
		20%	82.04616	0.06984	80.97956	0.00361
		30%	149.95039	0.09196	150.19042	0.11242

TABLE VI  
MEAN SQUARE ERROR OF PARAMETER ESTIMATES FOR DATA WITH AUTOREGRESSIVE-1 CORRELATION MATRIX

Case	Coeff.	$\varepsilon$	$\alpha = 0.3$		$\alpha = 0.7$	
			IRLS	IRLTS	IRLS	IRLTS
Case A	$\hat{\beta}_0$	0%	0.01309	0.01349	0.02329	0.02515
		10%	89.85427	0.02033	87.75463	0.03249
		20%	292.90715	0.06774	289.43715	0.08572
		30%	539.61490	0.19286	543.38625	0.22402
	$\hat{\beta}_1$	0%	0.00131	0.00136	0.00231	0.00247
		10%	0.88483	0.00192	0.86341	0.00328
		20%	1.85020	0.00659	1.77790	0.00856
		30%	2.98370	0.01864	2.71468	0.02134

Case	Coeff.	$\varepsilon$	$\alpha = 0.3$		$\alpha = 0.7$	
			IRLS	IRLTS	IRLS	IRLTS
Case B		0%	0.01354	0.01337	0.02704	0.02577

$\hat{\beta}_0$	10%	87.20910	0.05218	42.39676	0.02589
	20%	164.20519	0.16609	135.05622	0.02836
	30%	237.06421	0.34325	243.38623	0.41559
$\hat{\beta}_1$	0%	0.00135	0.00133	0.00273	0.00258
	10%	27.51679	0.01437	24.80085	0.00262
	20%	86.50801	0.07905	81.76371	0.00323
	30%	149.53554	0.11950	151.10469	0.13076

## V. Conclusion

Our proposed method have two different iterations in its procedure, one is the iteration for the estimation of regression parameter using IRLS method, and the other iteration is for selecting the best subset  $H$  for calculating the parameter estimate. We have shown that this procedure can minimize the effect of outliers on parameter estimation; IRLTS can produce a relatively efficient and consistent estimator compared to the classical GEE (IRLS). Base on the MSE, IRLTS performs much better than the classical GEE. Hence, robust GEE using IRLTS is a good choice for longitudinal data analysis when data contains outliers.

## REFERENCES

- [1] P. McCullagh and J.A. Nelder, *Generalized Linear Models*. London : Chapman and Hall, 1989.
- [2] K.Y. Liang and S.L. Zeger, "Longitudinal data analysis using generalized linear models", *Biometrika*, vol 73, pp. 13-22, 1986.
- [3] B.F. Qaqish and J.S. Preisser, "Resistant fits for regression with correlated outcomes an estimating equations approach", *Journal of Statistical Planning and Inference*, vol. 75, pp. 15-431, 1999.
- [4] P.S. Gill, "A Robust Mixed Linear Model Analysis for Longitudinal Data", *Statistics in Medicine*, vol. 19, pp. 975-987, 2000.
- [5] S.H. Jung and Z. Ying, "Rank-Based Regression With Repeated Measurements Data", *Biometrika*, vol. 90, pp. 732-740, 2003.
- [6] Abebe, A., McKean, J. W. & Kloke, J. D., Bilgic, Y. "Iterated Reweighted Rank-Based Estimates for GEE Models", submitted.
- [7] R.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
- [8] J.A. Vížek, "The Least Trimmed Squares – random carriers". *Bulletin of the Czech Econometric Society*, vol. 6, pp. 1-30, 1999.
- [9] J.A. Vížek, The Least Trimmed Squares. Part II:  $\sqrt{n}$ -consistency. *Kybernetika*, vol. 42, pp. 181-202, 2006a.
- [10] J.A. Vížek, "The Least Trimmed Squares. Part III: Asymptotic normality". *Kybernetika*, vol. 42, pp. 203-224, (2006b).
- [11] P.J. Rousseeuw, and K. van Driessen, " Computing LTS Regression for Large Data Sets". *Data mining and Knowledge Discovery*, vol. 12, pp. 29-45, 2006.

**Lampiran 2. Paper ilmiah disubmit ke International Journal of Pure and Applied Mathematics**

# Robust Generalized Estimating Equation Using Iterated Reweighted Least Trimmed Square

Netti Herawati and Khoirin Nisa

*Department of Mathematics, Lampung University  
Jl. Soemantri Borojonegoro No. 1, Bandar Lampung 35145, Indonesia  
Email: netti.herawati@fmipa.unila.ac.id*

## Abstract

A robust procedure for estimating parameters of regression model when generalized estimating equation (GEE) is applied to longitudinal data that contains outliers is proposed. The procedure is a combination of the iterated reweighted least square (IRLS) and least trimmed square (LTS) methods and is called *iterated reweighted least trimmed square* (IRLTS). We conducted a simulation study for gamma model and Poisson model using the proposed method, the result shows that our approach can provide a better result than the classical GEE.

**Keywords:** longitudinal data, outlier, regression model.

## 1. Introduction

In statistics, generalized estimating equation (GEE) is used to estimate the parameters of a generalized linear model (GLM) with a possible unknown correlation between outcomes. It is a general statistical approach to fit a marginal model for longitudinal or clustered data analysis, and it has been popularly applied into clinical trials and biomedical studies. GEEs belong to a class of regression techniques that are referred to as semiparametric because they rely on specification of only the first two moments. Under correct model specification and mild regularity conditions, parameter estimates from GEEs are consistent. The generalised estimating equation approach requires correct specification of the first two moments of a model. However, these moment assumptions can be distorted by contaminated or irregular measurements namely outliers. As a result, the generalised estimating equation method fails to give consistent estimators, and more seriously this will lead to incorrect conclusions (Qaqish & Preisser, 1999; Abebe et al, 2014). In this situation, we need a robust method that can minimize the effect of outliers.

In recent years, a few authors have considered robust methods for longitudinal data analysis. For example, Qaqish and Preisser (1999) proposed a resistant version of the GEE using M-type estimation by involving down-weighting influential data points. Gill (2000) proposed a robustified likelihood based on multivariate normal distribution. Jung and Ying (2003) proposed an adaptation of the Wilcoxon-Mann-Whitney method of estimating linear regression parameters to be used in longitudinal data analysis under the working independence model. The method proposed by Jung



and Ying was then modified by Wang and Zhu (2006). The last and recent robust method was proposed by Abebe et al (2014), he used iterated reweighted rank-based estimation for GEE model.

In this paper, we adopt the least trimmed square (LTS) method proposed by Rousseeuw and Leroy (1987) for robust linear regression. We adopt the LTS in the sense of trimming the data for estimating the regression coefficients. In Section 2 we present a brief review of GEE. In Section 3 we describe our proposed method and discuss the results from our simulation study. An example from real data is given in Section 4.

## 2. Generalized Estimating Equation

Let the vector of measurements on the  $i$ th subject be  $\mathbf{Y}_i = [Y_{i1}, \dots, Y_{i n_i}]^T$  with corresponding vector of means  $\boldsymbol{\mu}_i = [\mu_{i1}, \dots, \mu_{i n_i}]^T$  and  $\mathbf{X}_i = [X_{i1}, \dots, X_{i n_i}]^T$  be the  $n_i \times p$  matrix of covariates. In general, the components of  $\mathbf{Y}_i$ , are correlated but  $\mathbf{Y}_i$  and  $\mathbf{Y}_k$  are independent for any  $i \neq k$ . To model the relation between the response and covariates, we can use a regression model similar to the generalized linear models:

$$g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta}$$

where  $\boldsymbol{\mu}_i = E(\mathbf{Y}_i | \mathbf{X}_i)$ ,  $g$  is a specified link function, and  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^T$  is a vector of unknown regression coefficients to be estimated. The GEE of Liang and Zeger (1986) for estimating the  $p \times 1$  vector of regression parameters  $\boldsymbol{\beta}$  is given by :

$$S(\boldsymbol{\beta}) = \sum_{i=1}^K \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} [\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] = 0,$$

where  $\mathbf{V}_i$  be the covariance matrix of  $\mathbf{Y}_i$  modeled as  $\mathbf{V}_i = \lambda \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$ ,  $\mathbf{A}_i$  is a diagonal matrix of variance functions  $v(\mu_{ij})$ , and  $\mathbf{R}(\boldsymbol{\alpha})$  is the working correlation matrix of  $\mathbf{Y}_i$  indexed by a vector of parameters  $\boldsymbol{\alpha}$ . Solutions to Eq. (2) are obtained by alternating between estimation of  $\lambda$ ,  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ . There are several specific choices of the form of working correlation matrix  $\mathbf{R}_i(\boldsymbol{\alpha})$  commonly used to model the correlation matrix of  $\mathbf{Y}_i$ , among them are *exchangeable* and *autoregressive* correlation matrices.

Solving for  $\boldsymbol{\beta}$  is done with iteratively reweighted least squares (IRLS). The following is an algorithm for fitting the specified model using GEEs as described by Johnson and Stokes (1996) and Qaqish and Preisser (1999) :

1. Compute an initial estimate of  $\hat{\boldsymbol{\beta}}_{GEE}$ , for example with an ordinary generalized linear model assuming independence.
2. A current estimate  $\hat{\boldsymbol{\beta}}_{GEE}$  is updated by regressing the working response vector

$$\mathbf{Z}^* = \mathbf{X} \hat{\boldsymbol{\beta}} + \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} (\mathbf{y} - \hat{\boldsymbol{\mu}})$$

on  $\mathbf{X}$ . A new estimate  $\hat{\boldsymbol{\beta}}_{new}$  is obtained by :

$$\hat{\boldsymbol{\beta}}_{new} = (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* \mathbf{Z}^* \quad (1)$$

where  $\mathbf{W}^*$  is a block diagonal weight matrix whose  $i$ th block is the  $n_i \times n_i$  matrix

$$\mathbf{W}_i^* = \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^{-1} \mathbf{A}_i^{-1} \mathbf{R}_i(\hat{\boldsymbol{\alpha}}) \mathbf{A}_i^{-1} \left( \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^{-1}.$$

3. Use  $\hat{\boldsymbol{\beta}}_{new}$  to update  $\hat{\boldsymbol{\eta}} = \mathbf{X} \hat{\boldsymbol{\beta}}_{new} = \mathbf{H} \mathbf{Z}^*$ , where  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*$
4. Iterate until convergence.

### 3. Iterated Reweighted Least Trimmed Square

Let us briefly recall that the robust estimation of regression parameters using LTS method is given by:

$$\hat{\boldsymbol{\beta}}_{LTS} = \arg \min \sum_{i=1}^h e_i^2$$

where  $e_1^2 \leq e_2^2 \leq \dots \leq e_h^2 \leq \dots \leq e_n^2$  are the ordered squared residuals, from smallest to largest. LTS is calculated by minimizing the  $h$  ordered squares residuals, where  $h$  can be chosen between the range  $\frac{n}{2} + 1 \leq h \leq \frac{3n}{4} + \frac{n+1}{4}$ , with  $n$  being sample size and number of parameters, respectively. One can refer to e.g. Rousseeuw and van Driessen (2006), Roozbeh and Babaie-Kafaki (2015) for some details on LTS method.

We state our proposed method in the following short algorithm, combining IRLTS and LTS to solve  $\boldsymbol{\beta}$  of GEE model. To motivate our estimator, it is convenient to write the algorithm with involving the residuals.

1. Compute an initial estimate of  $\hat{\boldsymbol{\beta}}_{GEE}$  using IRLS
2. Use the estimate to calculate fitted value :  $\hat{\boldsymbol{\mu}}_i = g^{-1}(\mathbf{X}_i \hat{\boldsymbol{\beta}})$
3. Calculate residuals :  $e_{ij} = Y_{ij} - \hat{\boldsymbol{\mu}}_{ij}$ . Sort  $|e_{ij}|$  for  $i = 1, 2, \dots, t$  and  $j = 1, 2, \dots, n$  in ascending order :  $|e_{11}| \leq |e_{12}| \leq \dots \leq |e_m|$
4. Choose  $h$  observations which have the lowest  $h$ -residuals, we denote as subset H
5. Improve estimates of  $\boldsymbol{\beta}$  by solving  $\hat{\boldsymbol{\beta}}_{new}$  based on subset H using IRLS
6. Iterate until convergence.

### 3.2 Simulation

In this section, we compare the performances of IRLTS and IRLS through simulation study. Two types of outcomes are considered, continuous and count responses. The models for data generation are as follows:

$$\begin{aligned} 1/\mu_{ij} &= \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} \\ \text{Log}(\mu_{ij}) &= \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} \end{aligned}$$

where  $\beta_{ks}$  for  $k=0,1,2$  are randomly generated,  $i=1,2,\dots, 200$  and  $j=1,2,\dots,5$ . The covariates  $x_{1ij}$  are i.i.d. from a uniform distribution  $\text{Unif}(1,5)$ , and  $x_2$  is the measurement time variable, i.e.  $x_{2i} = 1,2,3,4,5$ . For each scenario, we generate the data based on the underlying true correlation structures as exchangeable (EXCH) with  $\alpha=0.5$ . For the first model (inverse link) the gamma distributed model is used, and for the second model (log link) the Poisson distributed model is applied. In this simulation, 1000 Monte Carlo data sets are generated for each scenario. We considered contamination proportion in data  $\varepsilon = 5\%, 10\%, 20\%$  and  $30\%$ . We evaluate the results using the mean square error (MSE) of the parameter estimates.

We provide the expected values and MSEs of parameter estimates resulted from our simulation on Table 1- Table 4. Table 1 and Table 2 show the expected values and MSEs of parameter estimates for the first model, while Table 3 and Table 4 for the second model.

Table 1. The expected values, standard errors and MSEs of  $\hat{\beta}_0$  of gamma distributed model with exchangeable correlation matrix

Method	Classical GEE			IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
5%	1.833421	0.417356	0.419098	1.292539	0.280084	0.080563
10%	1.963008	0.551804	0.694453	1.343505	0.345998	0.119739
20%	2.105067	0.522999	0.861099	1.302799	0.245013	0.061308
30%	2.158723	0.712847	1.180859	1.432421	0.581205	0.346613
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
5%	0.664892	0.297197	0.205406	0.957103	0.034595	0.003693
10%	0.587415	0.348410	0.297492	0.968876	0.075239	0.007119
20%	0.314225	0.347369	0.600688	0.964959	0.039372	0.003323
30%	0.297561	0.338291	0.617832	0.937923	0.088894	0.012682
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
5%	0.743453	0.289166	0.174850	0.974127	0.039936	0.006689
10%	0.549449	0.423348	0.425292	0.958451	0.018566	0.007922
20%	0.355655	0.349199	0.597828	0.975630	0.032734	0.005954
30%	0.332191	0.336312	0.621917	0.969043	0.071512	0.010960

Table 2. The expected values, standard errors and MSEs of  $\hat{\beta}_1$  of gamma distributed model with autoregressive correlation matrix

Method	Classical GEE			IRLTS		
	$\hat{\beta}_0$					
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
5%	1.713422	0.174403	0.127509	1.169724	0.219998	0.102270
10%	2.051958	0.292341	0.476604	1.172668	0.169948	0.081395
20%	2.096749	0.279746	0.561178	1.182726	0.281815	0.127424
30%	2.208355	0.570003	0.975395	1.094088	0.105856	0.105908
	$\hat{\beta}_1$					
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
5%	0.746269	0.169773	0.087549	0.977105	0.032371	0.001180
10%	0.493787	0.257591	0.311197	0.974237	0.026179	0.000892
20%	0.290721	0.257720	0.553459	0.982542	0.035247	0.001279
30%	0.228004	0.231963	0.595925	0.981851	0.023087	0.000579
	$\hat{\beta}_2$					
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
5%	0.739142	0.224797	0.132593	0.981539	0.021532	0.002405
10%	0.511909	0.302973	0.355672	0.981291	0.028264	0.002762
20%	0.281145	0.270400	0.627330	0.983924	0.028529	0.002551
30%	0.225163	0.377600	0.646722	1.003760	0.035796	0.001758

As shown in Table 1 and Table 2, our approach (IRLTS) performs better than the classical GEE. The MSEs of IRLTS are smaller than the MSEs of classical GEE, the outliers influence the estimation of  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . The parameter estimates of classical GEE are much more influenced than the parameter estimates of IRLTS. The more outliers contained in the data the larger the deviation of classical GEE estimates from the parameter value. In Table 3 and Table 4, the behavior of MSEs of both methods is the same as the first case, here we can see that IRLTS performs better than the classical GEE because the MSEs of IRLTS are smaller than the MSEs of classical GEE.

Table 3. The expected values, standard errors and MSEs of  $\hat{\beta}_2$  of Poisson distributed model with exchangeable correlation matrix

Method	Classical GEE			IRLTS		
$\hat{\beta}_0$						
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
5%	7.851928	0.578389	49.012922	2.242959	0.411831	1.041084
10%	9.299508	0.776268	71.575965	2.084935	0.454712	1.670854
20%	10.710302	0.425014	96.915019	2.351178	0.546663	1.478125
30%	11.413164	0.621123	111.439999	1.988002	0.486656	1.475743
$\hat{\beta}_1$						
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
5%	0.309635	0.108437	0.418891	1.052717	0.059291	0.014543
10%	0.194983	0.044670	0.568586	1.011790	0.074132	0.009602
20%	0.097441	0.077894	0.729016	0.918343	0.038557	0.002349
30%	0.092040	0.031461	0.733153	0.809874	0.101063	0.029211
$\hat{\beta}_2$						
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
5%	0.281547	0.064601	0.478304	0.947537	0.021808	0.000986
10%	0.159566	0.060507	0.660655	0.967633	0.065401	0.004284
20%	0.087528	0.031481	0.779957	0.900333	0.028748	0.005696
30%	0.067109	0.035320	0.816673	0.825752	0.057643	0.024164

The result for Poisson model shows similar behaviour to the result for gamma model. For the result of Poisson model in Table 3 and Table 4, IRLTS also performs better than the classical GEE. The MSEs of IRLTS are smaller than the MSEs of classical GEE, the outliers influence the estimation of  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . The parameter estimates of classical GEE are much more influenced than the parameter estimates of IRLTS. The more outliers contained in the data the larger the deviation of classical GEE estimates from the parameter value. In Table 3 and Table 4, the behavior of MSEs of both methods is the same as the first case, here we can see that IRLTS performs better than the classical GEE.

The estimation of IRLTS yields better results than classical GEE for both case we considered here. The MSEs of IRLTS is smaller than classical GEE, this means that IRLTS can reduce the influence of the high leverage points better than the classical GEE.

Table 4. The expected values, standard errors and MSEs of  $\hat{\beta}_0$  of Poisson distributed model with autoregressive correlation matrix

Method	Classical GEE			IRLTS		
	$\hat{\beta}_0$					
Jumlah pencilan	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG	$E(\hat{\beta}_0)$	$SE(\hat{\beta}_0)$	KTG
5%	8.493574	0.689463	40.055071	1.614313	0.331758	0.455831
10%	9.788645	0.962589	58.478701	1.765065	0.204619	0.523307
20%	10.525858	0.877176	70.050502	2.356166	0.470975	0.724548
30%	10.635832	0.570803	71.449706	3.071576	0.974599	1.054251
	$\hat{\beta}_1$					
Jumlah pencilan	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG	$E(\hat{\beta}_1)$	$SE(\hat{\beta}_1)$	KTG
5%	0.309635	0.108437	0.418891	1.052717	0.059291	0.014543
10%	0.194983	0.044670	0.568586	1.011790	0.074132	0.009602
20%	0.097441	0.077894	0.729016	0.918343	0.038557	0.002349
30%	0.092040	0.031461	0.733153	0.809874	0.101063	0.029211
	$\hat{\beta}_2$					
Jumlah pencilan	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG	$E(\hat{\beta}_2)$	$SE(\hat{\beta}_2)$	KTG
5%	0.281547	0.064601	0.478304	0.947537	0.021808	0.000986
10%	0.159566	0.060507	0.660655	0.967633	0.065401	0.004284
20%	0.087528	0.031481	0.779957	0.900333	0.028748	0.005696
30%	0.067109	0.035320	0.816673	0.825752	0.057643	0.024164

## 5 Concluding Remark

In this paper we have shown the effect of outliers on parameter estimates resulted by classical GEE and our proposed procedur (IRLTS). IRLTS performs better than the classical GEE for the two models we considered; gamma model and Poisson model; when data contain outliers.

## References

- [1] A. Abebe, J. W. McKean, J. D. Kloke and Y. Bilgic, Iterated Reweighted Rank-Based Estimates for GEE Models, Technical Report. (2014).
- [2] P. S. Gill, A Robust Mixed Linear Model Analysis for Longitudinal Data. *Statistics in Medicine*, **19** (2000), 975-987.

- [3] G. Johnston and M. Stokes, Repeated Measures Analysis With Discrete Data Using The SAS System. *SUGI Proceeding*. SAS Institute Inc., Cary, NC, (1996).
- [4] S. H. Jung and Z. Ying, Rank-Based Regression With Repeated Measurements Data, *Biometrika*, **90** (2003), 732-740.
- [5] K. Y. Liang and S. L. Zeger, Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73** (1986), 13-22.
- [6] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, London, 1989.
- [7] B. F. Qaqish and J. S. Preisser, Resistant fits for regression with correlated outcomes an estimating equations approach. *Journal of Statistical Planning and Inference*, **75** (1999), 15-431.
- [8] R. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.
- [9] P. J. Rousseeuw and K. van Driessen, Computing LTS Regression for Large Data Sets. *Data mining and Knowledge Discovery*, **12** (2006), 29-45.
- [10] Y. G. Wang and M. Zhu, Rank-based regression for analysis of repeated measures. *Biometrika*, **93** (2006), 459-464.

**Received: Month XX, 2016**

### Lampiran 3. R codes untuk simulasi data

```
library(MASS)
library(geepack)
library(xlsxjars)

#NORMAL MODEL

SIMULATE=function(n,m,p,R,Corr){
### description n=number of individuals, ###
###           m= number of observations for each individual, ###
###           p= percent of outliers ###
###           Corr= Working correlation matrix: 1="Exchangeable"
and 2="AR1" ###

Result<- matrix(NA,R,6)
beta0 <- 1.0
beta1 <- 1.0
beta2 <- 1.0

##Generate ID
ID <- rep(1:n, m)
ID<- sort(ID)

## A Toeplitz 1 correlation matrix has the following structure
if (Corr==1){
Work<-matrix(c(1,    0.5,  0.5,  0.5,  0.5,
              0.5,  1,    0.5,  0.5,  0.5,
              0.5,  0.5,  1,    0.5,  0.5,
              0.5,  0.5,  0.5,  1,    0.5,
              0.5,  0.5,  0.5,  0.5,  1),5,5)}else

{Work <- matrix(c(1,  0.5,  0.25,  0.125,  0.0625,
                0.5,  1,  0.5,  0.25,  0.125,
                0.25,  0.5,  1,  0.5,  0.125,
                0.125,  0.25,  0.5,  1,  0.5,
                0.0625,  0.125,  0.25,  0.5,  1),5,5)}

for (i in (1:R))
{
##person-level random effect (X) having a uniform(1,5) distribution
X<-runif(n,1,5)
X<-rep(X,each=m)

##20 people with a person-level fixed effect (X2)
X2<-rep(c(1:5),n)

## create five error variables that have a Toeplitz 1 structure
a<-rep(0,m)
errormean<-t(t(a))
e <- mvrnorm(n, mu=errormean, Sigma=Work)

##reshape long
error<-c(t(e))

##simulate contaminations
Outlier<-(p*n*m)/100
```



```

location <- round(runif(Outlier,1,n*m-1))
out<-t(location)
error[out]<-rnorm(1,100,1)

## generate Y
Y<-beta0+beta1*X+beta2*X2+error

### set up data frame
Data <- data.frame(ID, X, X2, Y)

##GEE estimation
if (Corr==1){
geeEst <- gee(Y~X+X2, id=ID, data=Data,
family=gaussian,corstr="exchangeable")
}else
{geeEst <- gee(Y~X+X2, id=ID, data=Data, family=gaussian,
corstr="ar1")}

##store the coefficients of classical estimation
est<-summary(geeEst)$coeff
b0<-est[1,1]
b1<-est[2,1]
b2<-est[3,1]
Result[i,1]=b0
Result[i,2]=b1
Result[i,3]=b2

##IRLTS Algorithm
repeat
{
h<-[n/2]+1
combination<-choose(n*m,h)
for (i in (1:combination))
{
Yhat<-b0+b1*X+b2*X2
DataIRLTS<-Data
DataIRLTS$res<-abs(Data$Y-Yhat)
DataIRLTS<-DataIRLTS[order(DataIRLTS$res),]
DataIRLTS<-DataIRLTS[1:h,]
if (Corr==1){
geeIRLTS <- gee(Y~X+X2, id=ID, data=DataIRLTS, family=gaussian,
corstr="exchangeable")
}else
{geeIRLTS <- gee(Y~X+X2, id=ID, data=DataIRLTS, family=gaussian,
corstr="ar1")}

##store the coefficients of robust estimation
est<-summary(geeIRLTS)$coeff
b0new<-est[1,1]
b1new<-est[2,1]
b2new<-est[3,1]
Result[i,4]=b0new
Result[i,5]=b1new
Result[i,6]=b2new
def<-((b0-b0new)+(b1-b1new)+(b2-b2new))/3
b0<-b0new
b1<-b1new
b2<-b2new
if (def < 0.001) break

```

```

}}
return(Result)
}
## Simulate GEE Calculating the average estimates and MSE
REPLICATE=function(R)
{
## Calculating the expected values, SE and MSE
GEE0<-SIMULATE(200,5,0,R)
EstMean0<-colMeans(GEE0)
SSE<- (GEE0-rep(1, EstMean0,1))^2
Var0<- colSums(SSE)/R
EstSSE0<- (GEE0-rep(1,R,1))^2
MSE0<- colSums(EstSSE0)/(R)

GEE1<-SIMULATE(200,5,1,R)
EstMean1<-colMeans(GEE1)
SSE<- (GEE1-rep(1, EstMean1,1))^2
Var1<- colSums(SSE)/R
EstSSE1<- (GEE1-rep(1,R,1))^2
MSE1<- colSums(EstSSE1)/(R)

GEE5<-SIMULATE(200,5,5,R)
EstMean5<-colMeans(GEE5)
SSE<- (GEE5-rep(1, EstMean5,1))^2
Var5<- colSums(SSE)/R
EstSSE5<- (GEE5-rep(1,R,1))^2
MSE5<- colSums(EstSSE5)/(R)

GEE10<-SIMULATE(200,5,10,R)
EstMean10<-colMeans(GEE10)
SSE<- (GEE10-rep(1, EstMean10,1))^2
Var10<- colSums(SSE)/R
EstSSE10<- (GEE10-rep(1,R,1))^2
MSE10<- colSums(EstSSE10)/(R)

GEE20<-SIMULATE(200,5,20,R)
EstMean20<-colMeans(GEE20)
SSE<- (GEE20-rep(1, EstMean20,1))^2
Var20<- colSums(SSE)/R
EstSSE20<- (GEE20-rep(1,R,1))^2
MSE20<- colSums(EstSSE20)/(R)

GEE30<-SIMULATE(200,5,30,R)
EstMean30<-colMeans(GEE30)
SSE<- (GEE30-rep(1, EstMean30,1))^2
Var30<- colSums(SSE)/R
EstSSE30<- (GEE30-rep(1,R,1))^2
MSE30<- colSums(EstSSE30)/(R)

GEE40<-SIMULATE(200,5,40,R)
EstMean40<-colMeans(GEE40)
SSE<- (GEE40-rep(1, EstMean40,1))^2
Var40<- colSums(SSE)/R
EstSSE40<- (GEE40-rep(1,R,1))^2
MSE40<- colSums(EstSSE40)/(R)

Ebo<-rbind(EstMean0,EstMean1, EstMean5, EstMean10, EstMean20,
EstMean30, EstMean40)
Varb<-rbind(Var0,Var1,Var5,Var10,Var20,Var30,Var40)

```

```

MSEb<-rbind(MSE0,MSE1,MSE5,MSE10,MSE20,MSE30,MSE40)

outlier<-c(0,1,5,10,20,30,40)
FinalResult<-cbind(outlier,Ebo, Varb, MSEb)
colnames(FinalResult)<-c("Outlier","E(bo)", "E(b1)",
"E(b2)", "SE(bo)", "SE(b1)",
"SE(b2)", "MSE(bo)", "MSE(b1)", "MSE(b2)", "E(boIRLTS)", "E(b1IRLTS)",
"E(b2IRLTS)", "SE(boIRLTS)", "SE(b1IRLTS)",
"SE(b2IRLTS)", "MSE(boIRLTS)", "MSE(b1IRLTS)", "MSE(b2IRLTS)")
return(FinalResult)
}
Simulation<-REPLICATE(1000)
write.xlsx(Simulation, "/Simulation/Fundamental/NORMALRESULT.xlsx")

```

### #GAMMA MODEL

```

library(MASS)
library(geepack)
library(xlsx)

##Generate ID
ID <- rep(1:n, m)
ID<- sort(ID)

## A Toeplitz 1 correlation matrix has the following structure
Work <- matrix(c(1, 0.5, 0.5, 0.5, 0.5,
0.5, 1, 0.5, 0.5, 0.5,
0.5, 0.5, 1, 0.5, 0.5,
0.5, 0.5, 0.5, 1, 0.5,
0.5, 0.5, 0.5, 0.5, 1),5,5)

## A Toeplitz 1 correlation matrix has the following structure
Work <- matrix(c(1, 0.5, 0.25, 0.125, 0.0625,
0.5, 1, 0.5, 0.25, 0.125,
0.25, 0.5, 1, 0.5, 0.25,
0.125, 0.25, 0.5, 1, 0.5,
0.0625, 0.125, 0.25, 0.5, 1),5,5)

##person-level random effect (X) having a uniform(0,1) distribution
X<-runif(n,1,5)
X<-rep(X,each=m)

##200 people with a person-level fixed effect (X2)
X2<-rep(c(1:5),n)

## create five error variables that have a Toeplitz 1 structure
a<-rep(0,m)
errormean<-t(t(a))
e <- mvrnorm(n, mu=errormean, Sigma=Work)
cdf<-pnorm(e)
er<-qgamma(cdf,1,1)

##reshape long
error<-c(t(er))

```

```

## generate Y
#mu<-1/(1+ X+ X2+error)
mu<-1/(beta0+beta1*X+beta2*X2+error)
Y<-mu

### set up data frame
Data <- data.frame(ID, X, X2, Y)

##Model parameter
gee0 <- geeglm(Y~X+X2, id=ID, data=Data,
family=Gamma(link=inverse),corstr="exchangeable")

#gee0 <- geeglm(Y~X+X2, id=ID, data=Data,
#family=Gamma(link=inverse),corstr="ar1")

gee<-summary(gee0)$coeff
beta0<-gee[1,1]
beta1<-gee[2,1]
beta2<-gee[3,1]
Result[i,1]=beta0
Result[i,2]=beta1
Result[i,3]=beta2

SIMULATE=function(n,m,p,R)
### description n=number of individuals,
###          m= number of observations for each individual,
###          p= percent of outliers
###          Corr= Working correlation matrix: 1="Exchangeable"
and 2="AR-1"
{
Result<- matrix(NA,R,6)

for (i in (1:R))
{
##simulate contaminations
Outlier<-(p*n*m)/100
location <- round(runif(Outlier,1,n*m-1))
#my<-max(Y)
Y[location]=rgamma(1,1,1)

##GEE estimation
DataGEE<- data.frame(ID, X, X2, Y)
geeEst <- geeglm(Y~X+X2, id=ID, data=DataGEE, family=Gamma(link =
inverse),corstr="exchangeable")
#geeEst <- geeglm(Y~X+X2, id=ID, data=DataGEE, family=Gamma(link =
#inverse),corstr="ar1")

##store the coefficients of classical estimation
est<-summary(geeEst)$coeff
b0<-est[1,1]
b1<-est[2,1]
b2<-est[3,1]
Result[i,1]=b0
Result[i,2]=b1
Result[i,3]=b2
}

```

```

##IRLTS Algorithm
repeat
{
h<-n/2
combination<-choose(n*m,h)
for (i in (1:combination))
{
Yhat<-1/(b0+b1*X+b2*X2)
DataIRLTS<-DataGEE
DataIRLTS$res<-abs(DataGEE$Y-Yhat)
DataIRLTS<-DataIRLTS[order(DataIRLTS$res),]
DataIRLTS<-DataIRLTS[1:h,]
if (Corr==1){
geeIRLTS <- geeglm(Y~X+X2, id=ID, data=DataIRLTS,
family=Gamma(link=inverse), corstr="exchangeable")
}else
{ geeIRLTS <- geeglm(Y~X+X2, id=ID, data=DataIRLTS,
family=Gamma(link=inverse), corstr="ar1")
}

##store the coefficients of robust estimation
est<-summary(geeIRLTS)$coeff
b0new<-est[1,1]
b1new<-est[2,1]
b2new<-est[3,1]
Result[i,4]=b0new
Result[i,5]=b1new
Result[i,6]=b2new
def<-((b0-b0new)+(b1-b1new)+(b2-b2new))/3
b0<-b0new
b1<-b1new
b2<-b2new
if (def < 0.001) break
}}
return(Result)
}

## Simulate GEE Calculating the average estimates and MSE
REPLICATE=function(R)
{

GEE0<-SIMULATE(200,5,0,R)
EstMean0<-colMeans(GEE0)
SSE0<- (GEE0- t(replicate(R,EstMean0)))^2
Var0<- colSums(SSE0)/R
SE0<-sqrt(Var0)
EstMean<-t(EstMean0)
Beta<-EstMean[,1:3]
mBeta<-rep(Beta,3)
EstSSE0<- (GEE0-t(replicate(R,mBeta)))^2
MSE0<- colSums(EstSSE0)/(R)

GEE1<-SIMULATE(200,5,1,R)
EstMean1<-colMeans(GEE1)
SSE1<- (GEE1- t(replicate(R,EstMean0)))^2
Var1<- colSums(SSE1)/R
SE1<-sqrt(Var1)
EstSSE1<- (GEE1- t(replicate(R, mBeta)))^2

```

```

MSE1<- colSums (EstSSE1) / (R)

GEE5<-SIMULATE (200, 5, 5, R)
EstMean5<-colMeans (GEE5)
SSE5<- (GEE5- t(replicate (R, EstMean5))) ^2
Var5<- colSums (SSE5) /R
SE5<-sqrt (Var5)
EstSSE5<- (GEE5- t(replicate (R, mBeta))) ^2
MSE5<- colSums (EstSSE5) / (R)

GEE10<-SIMULATE (200, 5, 10, R)
EstMean10<-colMeans (GEE10)
SSE10<- (GEE10- t(replicate (R, EstMean10))) ^2
Var10<- colSums (SSE10) /R
SE10<-sqrt (Var10)
EstSSE10<- (GEE10- t(replicate (R, mBeta))) ^2
MSE10<- colSums (EstSSE10) / (R)

GEE20<-SIMULATE (200, 5, 20, R)
EstMean20<-colMeans (GEE20)
SSE20<- (GEE20- t(replicate (R, EstMean20))) ^2
Var20<- colSums (SSE20) /R
SE20<-sqrt (Var20)
EstSSE20<- (GEE20- t(replicate (R, mBeta))) ^2
MSE20<- colSums (EstSSE20) / (R)

GEE30<-SIMULATE (200, 5, 30, R)
EstMean30<-colMeans (GEE30)
SSE30<- (GEE30- t(replicate (R, EstMean30))) ^2
Var30<- colSums (SSE30) /R
SE30<-sqrt (Var30)
EstSSE30<- (GEE30- t(replicate (R, mBeta))) ^2
MSE30<- colSums (EstSSE30) / (R)

GEE40<-SIMULATE (200, 5, 40, R)
EstMean40<-colMeans (GEE40)
SSE40<- (GEE40- t(replicate (R, EstMean40))) ^2
Var40<- colSums (SSE40) /R
SE40<-sqrt (Var40)
EstSSE40<- (GEE40- t(replicate (R, mBeta))) ^2
MSE40<- colSums (EstSSE40) / (R)

Ebo<-rbind (EstMean0, EstMean1, EstMean5, EstMean10, EstMean20,
EstMean30, EstMean40)
Varb<-rbind (SE0, SE1, SE5, SE10, SE20, SE30, SE40)
MSEb<-rbind (MSE0, MSE1, MSE5, MSE10, MSE20, MSE30, MSE40)

outlier<-c (0, 1, 5, 10, 20, 30, 40)
FinalResult<-cbind (outlier, Ebo, Varb, MSEb)
colnames (FinalResult) <-c ("Outlier", "Beta0", "Beta1", "Beta2", "E (bo) ",
"E (b1) ", "E (b2) ", "E (boIRLTS) ", "E (b1IRLTS) ", "E (b2IRLTS) ",
"SE (Beta0) ", "SE (Beta1) ", "SE (Beta2) ", "SE (bo) ", "SE (b1) ", "SE (b2) ", "SE (b
oIRLTS) ", "SE (b1IRLTS) ", "SE (b2IRLTS) ",
"MSE (Beta0) ", "MSE (Beta1) ", "MSE (Beta2) ", "MSE (bo) ", "MSE (b1) ", "MSE (b2) "
, "MSE (boIRLTS) ", "MSE (b1IRLTS) ", "MSE (b2IRLTS) ")
return (FinalResult)
}
Simulation<-REPLICATE (10)
write.xlsx (Simulation, "/Simulation/Fundamental/GAMMAERESULT.xlsx")

```

```
#POISSON
```

Serupa dengan R-code untuk simulasi model normal dan gamma, bagian yang berbeda pada beberapa tahapan berikut :

```
## create five error variables that have a Toeplitz 1 structure
a<-rep(0,m)
errormean<-t(t(a))
e <- mvrnorm(n, mu=errormean, Sigma=Work)
cdf<-pnorm(e)
er<-qpois(cdf,1)
error<-c(t(er))

## generate Y
mu<-exp(beta0+beta1*X+beta2*X2+error)
Y<-rpois(1000,mu)

##Model estimation
gee0 <- geeglm(Y~X+X2, id=ID, data=Data, family=poisson(link =
log),corstr="exchangeable")

##simulate contaminations
Outlier<-(p*n*m)/100
location <- round(runif(Outlier,1,n*m-1))
Y[location]=rpois(1,5000)
```

```
#BINOMIAL
```

Beberapa tahapan yang berbeda sebagai berikut :

```
## create five error variables that have a Toeplitz 1 structure
a<-rep(0,m)
errormean<-t(t(a))
e <- mvrnorm(n, mu=errormean, Sigma=Work)
cdf<-pnorm(e)
er<-qbinom(cdf,5,0.5)

## generate Y
z<- 1+X+X2+error
pr = 1/(1+exp(-z))
Y<-rbern(1000,pr)

##Model estimation
gee0 <- geeglm(Y~X+X2, id=ID, data=Data,
family=binomial,corstr="exchangeable")

##simulate contaminations
Outlier<-(p*n*m)/100
location <- round(runif(Outlier,1,n*m-1))
Y[location]=rbern(1,0.3)
```

## Lampiran 4. Contoh output analisis GEE pada software R

### Output model normal

```
Call:
geeglm(formula = Y ~ X + X2, family = gaussian, data =
Data,
       id = ID, corstr = "exchangeable")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W )	
(Intercept)	0.65249	0.17180	14.42	0.000146	***
X	1.11271	0.04936	508.14	< 2e-16	***
X2	0.97691	0.01478	4366.79	< 2e-16	***

---

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

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	1.055	0.07301

Correlation: Structure = exchangeable Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.5325	0.0345

Number of clusters: 200 Maximum cluster size: 5

Analysis of 'Wald statistic' Table

Model: gaussian, link: identity

Response: Y

Terms added sequentially (first to last)

	Df	X2	P(> Chi )	
X	1	508	<2e-16	***
X2	1	4367	<2e-16	***

---

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

### Output model gamma

```
Call:
geeglm(formula = Y ~ X + X2, family = Gamma(link =
inverse),
       data = Data, id = ID, corstr = "ar1")
```



```

Coefficients:
              Estimate Std.err   Wald Pr(>|W|)
(Intercept)  1.2379    0.1557   63.2  1.9e-15 ***
X             1.0971    0.0462  563.5 < 2e-16 ***
X2           1.0193    0.0333  936.0 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

```

```

Estimated Scale Parameters:
              Estimate Std.err
(Intercept)  0.0337    0.0075

```

Correlation: Structure = ar1 Link = identity

```

Estimated Correlation Parameters:
              Estimate Std.err
alpha        0.425    0.0596
Number of clusters: 200 Maximum cluster size: 5

```

Analysis of 'Wald statistic' Table  
Model: Gamma, link: inverse  
Response: Y  
Terms added sequentially (first to last)

```

      Df  X2 P(>|Chi|)
X      1 584 <2e-16 ***
X2     1 936 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

```

## Output model Poisson

```

Call:
geeglm(formula = Y ~ X + X2, family = poisson(link = log),
data = Data,
       id = ID, corstr = "exchangeable")

```

```

Coefficients:
              Estimate Std.err   Wald Pr(>|W|)
(Intercept)  3.5289    0.6104   33.4  7.4e-09 ***
X             0.8701    0.0828  110.5 < 2e-16 ***
X2           0.9273    0.1316   49.6  1.8e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

```

Estimated Scale Parameters:  
 Estimate Std.err  
 (Intercept) 164392 98306

Correlation: Structure = exchangeable Link = identity

Estimated Correlation Parameters:  
 Estimate Std.err  
 alpha 0.0588 0.028  
 Number of clusters: 200 Maximum cluster size: 5

Analysis of 'Wald statistic' Table  
 Model: poisson, link: log  
 Response: Y  
 Terms added sequentially (first to last)

	Df	X2	P(> Chi )
X	1	116.2	< 2e-16 ***
X2	1	49.6	1.8e-12 ***

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

### Output model binomial

Call:  
 geeglm(formula = Y ~ X + X2, family = binomial, data =  
 Data,  
 id = ID, corstr = "exchangeable")

Coefficients:  
 Estimate Std.err Wald Pr(>|W|)  
 (Intercept) 3.449 0.720 22.92 1.7e-06 \*\*\*  
 X 1.532 2.811 0.30 0.59  
 X2 1.064 0.722 2.17 0.14

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

Analysis of 'Wald statistic' Table  
 Model: binomial, link: logit  
 Response: Y  
 Terms added sequentially (first to last)

	Df	X2	P(> Chi )
X	1	0.294	0.59
X2	1	2.174	0.14

Estimated Scale Parameters:  
                  Estimate Std.err  
(Intercept)      0.54      5.89

Correlation: Structure = exchangeable  Link = identity

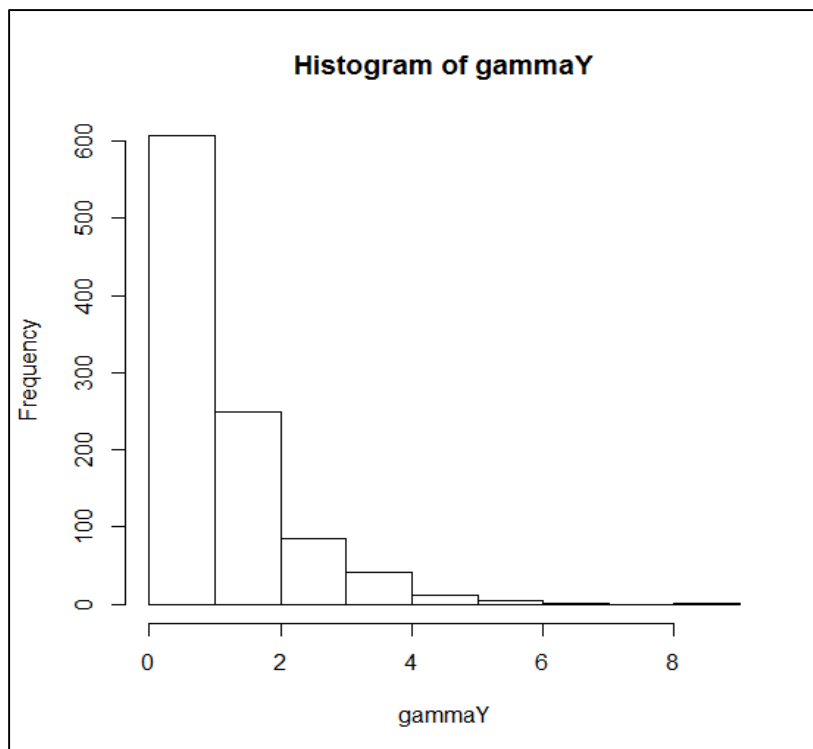
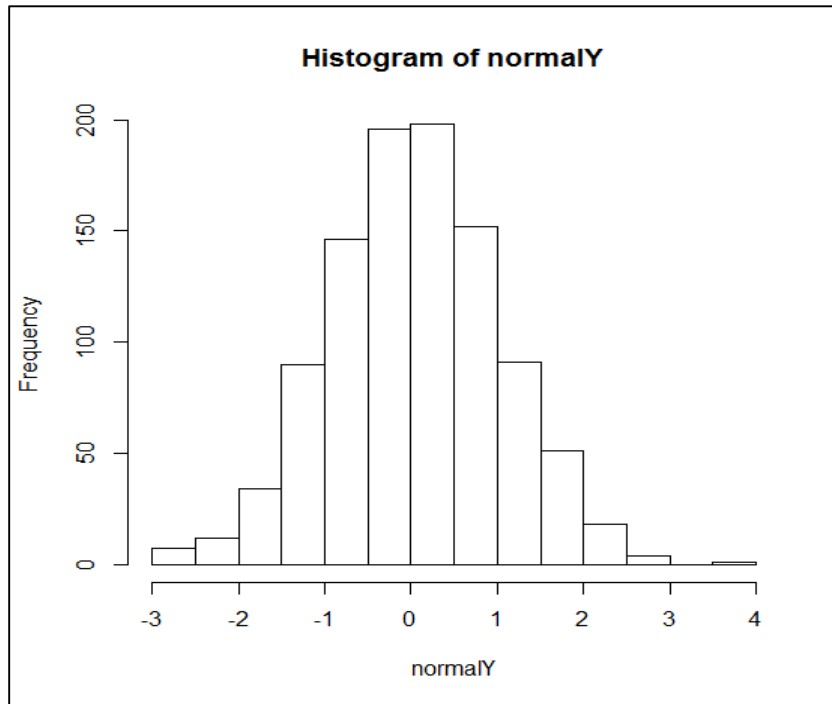
Estimated Correlation Parameters:

                  Estimate Std.err  
alpha -0.00202  0.0237

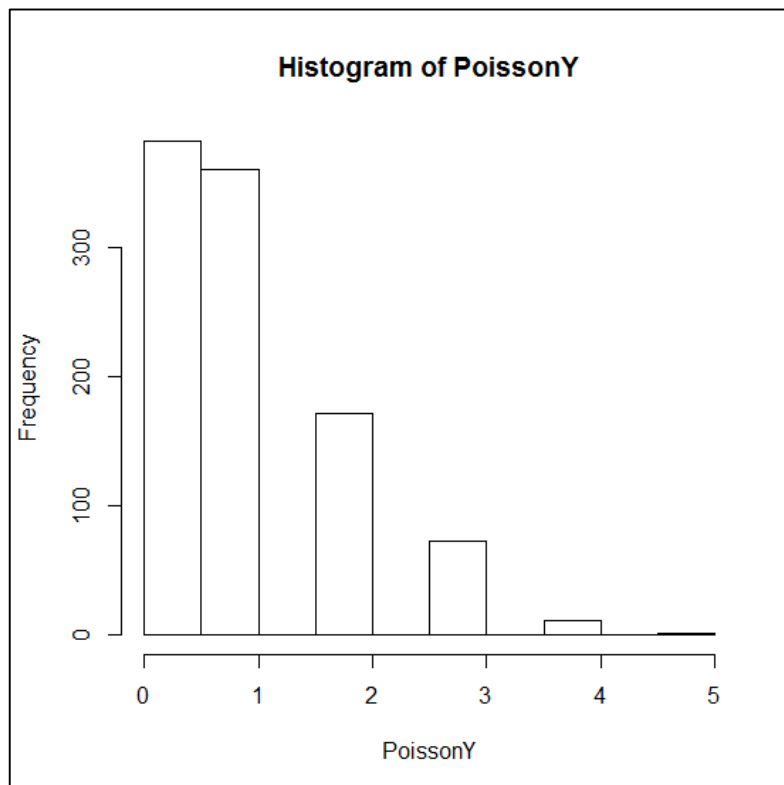
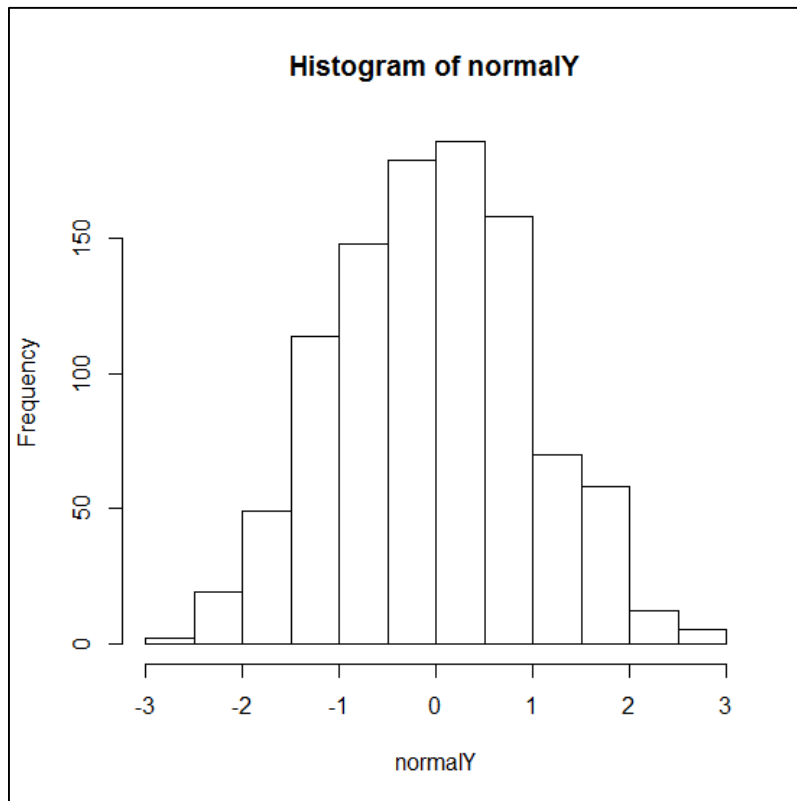
Number of clusters:  200  Maximum cluster size: 5

**Lampiran 5. Histogram data Y untuk model GEE sebelum dan sesudah konversi dengan teknik NORTA**

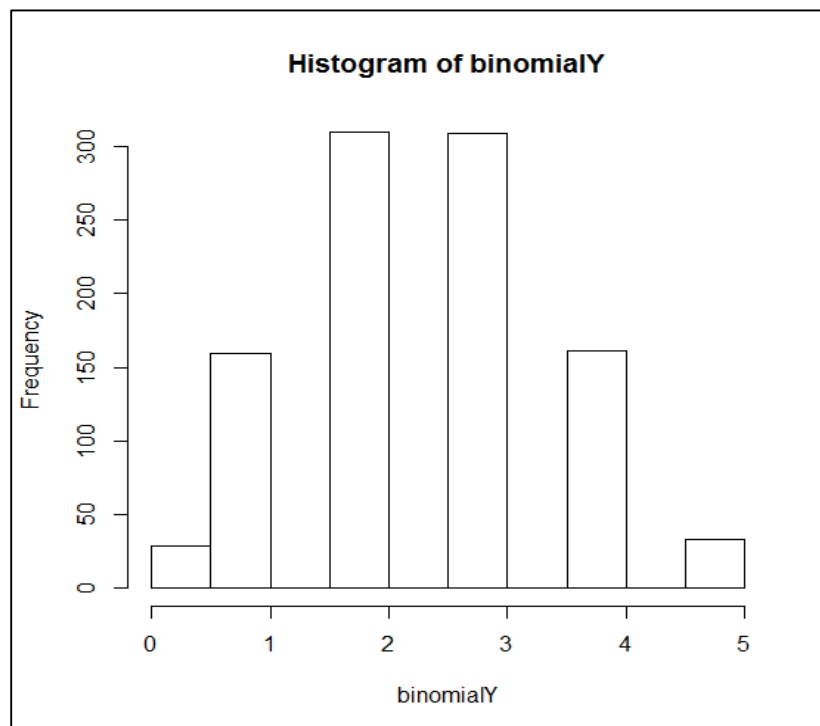
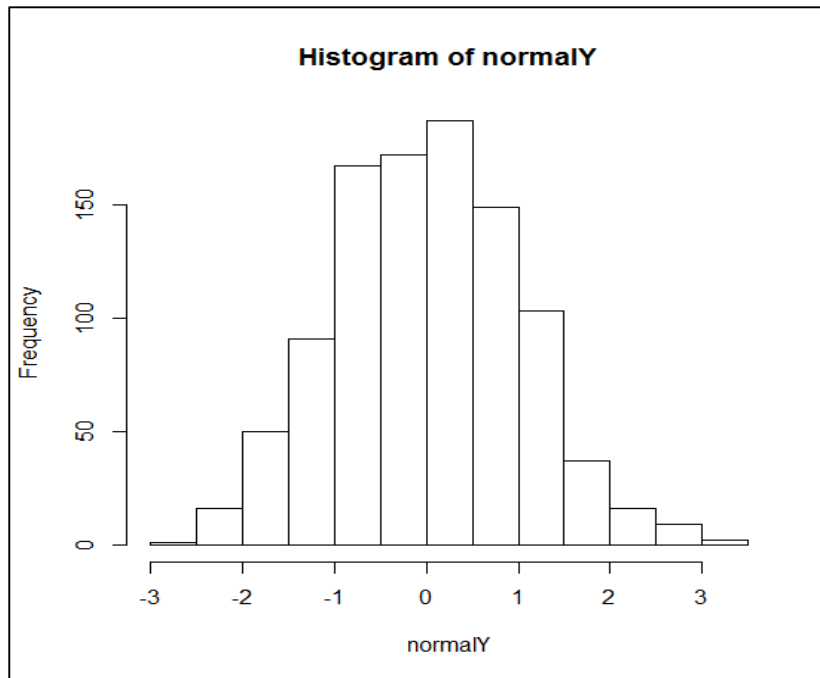
**1. Model gamma**



## 2. Model Poisson



### 3. Model binomial



**Lampiran 5. Contoh hasil simulasi dari 1000 kali replikasi untuk jumlah pencilan 1% dengan matriks korelasi *exchangeable***

	bo	b1	b2	bo IRLTS	b1 IRLTS	b2 IRLTS
[1,]	1.32507	1.18417	1.04807	1.05868	0.99364	0.99805
[2,]	1.23918	0.97527	1.27432	1.04630	1.01194	0.97394
[3,]	0.55742	1.19176	1.26334	0.78709	1.03857	1.01539
[4,]	0.69484	1.30376	1.15473	0.90255	1.03661	1.01056
[5,]	1.80505	1.07882	0.96105	1.05038	0.94368	1.01196
[6,]	0.61609	1.13257	1.32989	1.16570	0.96214	0.98284
[7,]	1.44099	1.38737	0.81726	1.12367	0.95293	1.02035
[8,]	1.25643	0.99063	1.25217	0.97723	1.01238	0.98490
[9,]	1.36440	1.17700	1.02858	1.02819	1.01108	0.97682
[10,]	0.49430	1.13067	1.38371	0.39014	1.01179	1.19158
[11,]	1.93129	1.40649	0.63938	1.01545	1.01816	0.98912
[12,]	1.01526	1.26769	1.04621	1.14218	0.94008	0.99677
[13,]	3.14230	0.76129	0.85539	1.06845	0.98909	1.01653
[14,]	2.12262	1.02434	0.95696	0.87770	1.05297	1.01312
[15,]	1.01585	0.99211	1.35300	1.12380	1.01539	0.95868
[16,]	2.25302	0.92332	1.00645	1.17667	0.98849	0.95928
[17,]	0.40852	1.17591	1.34263	1.20253	0.92296	0.99024
[18,]	2.43793	0.61198	1.23255	0.84233	0.98288	1.03351
[19,]	1.62352	0.86910	1.23307	0.84758	1.00196	1.03217
[20,]	1.52273	0.97086	1.15990	0.70726	1.05671	1.01311
[21,]	0.60904	1.44287	1.00551	0.90959	1.01903	1.00677
[22,]	2.05298	1.07847	0.89875	1.07439	0.97599	1.00177
[23,]	3.16969	0.53831	1.08679	1.05497	0.99238	0.98953
[24,]	1.90316	0.90294	1.11808	0.89207	1.00567	1.01889
[25,]	2.64374	0.95026	0.82084	0.67581	1.08442	1.01140
[26,]	0.41455	1.13741	1.40743	0.91128	1.04057	1.01514
[27,]	0.67486	1.10421	1.32406	0.71435	1.05699	1.02603
[28,]	2.10750	1.02693	0.94931	1.06316	0.99221	0.99968
[29,]	1.33558	0.96247	1.29184	1.11179	1.00463	0.99515
[30,]	1.97236	0.82241	1.19195	1.11185	0.97137	0.99168
[31,]	4.73745	0.31216	0.78135	1.26436	0.94522	0.97679
[32,]	0.88428	1.43885	0.95269	0.90145	1.03721	1.00398
[33,]	2.12125	0.68933	1.26759	0.84869	1.01616	1.01297
[34,]	2.65342	1.01774	0.77948	1.85838	1.16364	0.56896
[35,]	2.10408	1.27437	0.70349	0.93735	1.00080	1.01289
[36,]	1.28767	1.03555	1.16716	0.81456	1.00019	1.02758
[37,]	3.49421	0.67759	0.75465	1.02664	0.96298	1.00342
[38,]	0.72693	1.43869	0.98498	1.16665	0.96803	0.98596

[39,]	2.32805	1.11928	0.75698	0.99613	0.98378	1.00782
[40,]	0.68896	1.19989	1.24557	0.85190	1.06197	1.00015
[41,]	-0.37462	1.68798	1.07748	0.96717	0.99372	1.02970
[42,]	0.41007	1.25945	1.28142	1.05726	1.02527	0.97024
...						
...						
[965,]	3.10956	1.06441	0.57160	1.12484	0.93397	1.02379
[966,]	1.96191	0.90194	1.10290	1.05209	0.97510	1.00011
[967,]	0.49864	1.30423	1.18461	0.89024	1.08725	0.95740
[968,]	1.87908	1.22461	0.84493	1.14821	0.97580	1.00361
[969,]	2.66580	0.70311	1.09619	1.05979	0.98984	1.00990
[970,]	2.24759	1.27858	0.63947	0.82058	1.08676	0.98804
[971,]	2.20268	1.04184	0.89158	0.94346	1.02771	0.99389
[972,]	3.86838	0.73060	0.65231	0.73853	1.08827	1.00011
[973,]	3.60587	0.46800	0.94615	1.18416	0.94052	1.00277
[974,]	0.67688	1.07438	1.37090	1.12457	0.94129	1.01954
[975,]	1.57543	0.93846	1.20469	0.95944	1.01104	0.99618
[976,]	3.24631	0.60128	0.99905	1.07620	0.98524	0.99876
[977,]	2.75261	0.81214	0.92661	1.25962	0.94098	0.97202
[978,]	3.86251	0.73753	0.65528	1.20711	0.93415	1.00725
[979,]	3.61623	0.91959	0.56186	1.03469	0.99370	1.01403
[980,]	0.86335	1.15500	1.25208	1.20389	0.99369	0.97526
[981,]	3.08069	0.84253	0.80801	1.25729	0.90534	1.01747
[982,]	0.47531	1.14952	1.37035	1.24042	0.96867	0.96675
[983,]	3.60202	0.72365	0.72068	1.23422	0.94395	0.97307
[984,]	3.28099	0.71061	0.87981	1.29911	0.94988	0.98185
[985,]	1.92979	0.76917	1.24377	1.24193	0.92837	0.98927
[986,]	0.92471	1.40360	0.95283	0.89987	1.04038	1.00106
[987,]	2.51270	0.82983	0.99858	1.17793	0.92908	0.99497
[988,]	2.19667	0.91624	1.01847	0.94461	1.00164	1.01870
[989,]	4.16034	0.39755	0.86983	0.92024	1.01720	1.01988
[990,]	1.75679	1.22311	0.83772	0.75355	1.06569	1.00183
[991,]	1.13788	1.30031	0.98177	0.67561	1.07673	1.03238
[992,]	1.81866	1.05376	1.00710	0.93132	1.01465	1.00489
[993,]	1.25371	1.00143	1.26219	0.97860	1.01409	1.01741
[994,]	2.03346	1.34870	0.54122	0.79834	1.06443	0.98604
[995,]	1.94884	0.84798	1.17091	1.04334	0.97977	1.01449
[996,]	2.53466	1.18603	0.64358	0.89835	1.04691	0.99493
[997,]	1.62215	0.82688	1.25987	1.29149	0.85283	1.00650
[998,]	2.96390	0.75143	0.94452	1.25407	0.93973	0.99320
[999,]	1.28367	1.04042	1.18135	0.93241	1.02577	0.98409
[1000,]	2.02377	1.28980	0.69456	0.96691	1.01939	0.99173