



TESIS – SS 092304

**PEMODELAN JUMLAH HARI HUJAN PER BULAN DI
KABUPATEN INDRAMAYU DENGAN PENDEKATAN *ZERO
INFLATED MODEL (ZIM)***

**FRITS FAHRIDWS DAMANIK
NRP. 1312201716**

**DOSEN PEMBIMBING
Dr. rer. pol. Heri Kuswanto, M.Si**

**PROGRAM MAGISTER JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2014**



TESIS – SS 092304

**THE MODELING OF NUMBER OF MONTHLY RAINY DAYS
IN INDRAMAYU REGENCY USING ZERO INFLATED MODEL
(ZIM) APPROACH**

**FRITS FAHRIDWS DAMANIK
NRP. 1312201716**

**SUPERVISOR
Dr. rer. pol. Heri Kuswanto, M.Si**

**PROGRAM OF MAGISTER DEPARTEMENT OF STATISTICS
FACULTY OF MATEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2014**

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT atas segala limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan karya tulis yang berbentuk tesis ini yang berjudul:

“Pemodelan Jumlah Hari Hujan per Bulan di Kabupaten Indramayu dengan Pendekatan *Zero Inflated Model* (ZIM)”

Dalam proses penulisan tesis ini, tentunya banyak pihak yang telah memberikan bantuan demi terselesaikannya tesis ini. Oleh karena itu penulis ingin menyampaikan ucapan terima kasih yang tiada hingga kepada:

1. Ibu, Istriku Anisa Margareth dan Anak-anakku tercinta: Ata, Shafa dan Rashad, atas segala dukungan dan doanya yang tanpa henti.
2. Kepala Badan Pusat Statistik, atas kesempatan yang diberikan dalam menempuh jenjang pendidikan S2 Statistika di Institut Teknologi Surabaya.
3. Bapak Dr. rer. pol. Heri Kuswanto, M.Si selaku dosen pembimbing, atas bimbingannya yang penuh kesabaran hingga terselesaikannya penulisan tesis ini.
4. Bapak Dr. Brodjol Sutijo Suprih Ulama dan Ibu Dr. Irhamah, S.Si, M.Si selaku dosen penguji, atas segala kritik dan saran yang membangun demi terselesaikannya tesis ini.
5. Bapak Dr. Mashuri, MT selaku Ketua Jurusan Statistika ITS dan Bapak Dr. Suhartono, S.Si, M.Sc selaku Kaprodi S2 Jurusan Statistika ITS serta seluruh dosen pengajar beserta karyawan Jurusan Statistika, atas segala pemberian ilmu yang bermanfaat serta fasilitas pendidikan yang diperoleh selama menjalani proses pendidikan di Jurusan Statistika ITS.
6. Semua pihak yang telah membantu penulis dalam proses penulisan tesis ini yang tidak dapat disebutkan satu persatu.

Demi kesempurnaan tesis ini, saran dan kritik yang membangun dari berbagai pihak sangat penulis harapkan. Akhirnya hanya kepada Allah SWT kita kembalikan segala urusan dan semoga tesis ini bermanfaat basgi semua pihak,

khususnya bagi penulis sendiri dan para pembaca pada umumnya, semoga Allah SWT selalu meridhoi amal ibadah kita, Amin Ya Rabbal Alamin.

Surabaya, Juli 2014

Penulis,

**PEMODELAN JUMLAH HARI HUJAN PER BULAN DI
KABUPATEN INDRAMAYU DENGAN PENDEKATAN
ZERO INFLATED MODEL (ZIM)**

**Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)**

**di
Institut Teknologi Sepuluh Nopember
Oleh :**

**FRITS FAHRIDWS DAMANIK
NRP : 1312 201 716**


Tanggal Ujian : 27 Juni 2014

Disetujui Oleh :

Periode Wisuda : September 2014


**1. Dr. rer.pol. Heri Kuswanto, M.Si
NIP. 19820326 200312 1 004**

(Pembimbing)


**2. Dr. Brodjol Sutijo Suprih Ulama
NIP. 19660125 199002 1 001**

(Penguji)


**3. Dr. Irhamah, S.Si, M.Si
NIP. 19780406 200112 2 002**

(Penguji)



Direktur Pascasarjana

**Prof. Dr. Ir. Adi Soeprijanto, MT
NIP. 19640405 199002 1 001**

Pemodelan Jumlah Hari Hujan per Bulan di Kabupaten Indramayu dengan Pendekatan *Zero Inflated Model* (ZIM)

Nama Mahasiswa : Frits Fahridws Damanik
NRP : 1312201716
Pembimbing : Dr. rer.pol. Heri Kuswanto, M.Si

ABSTRAK

Produktivitas pertanian secara umum dipengaruhi oleh faktor genetik (jenis bibit) dan faktor lingkungan seperti jenis tanah, pengelolaan lahan, teknologi, dan iklim/cuaca. Dari faktor lingkungan, keadaan cuaca sangat memegang peranan penting terhadap keberhasilan pertanian. Kondisi ini terjadi tidak lepas dari sangat berfluktuatifnya kejadian hujan antar waktu sehingga pembudidayaan tanaman perlu disesuaikan terhadap fluktuasi curah hujan dan frekuensi hujan. Untukantisipasi terhadap perubahan curah hujan dan frekuensi hujan yang tiba-tiba ekstrim diperlukan suatu sistem peringatan dini dengan memanfaatkan model prediksi jumlah hari hujan, sehingga gambaran frekuensi hujan beberapa periode ke depan dapat di peroleh lebih awal. Pemodelan terhadap kasus kejadian diskrit secara *series* seperti jumlah hari hujan per bulandapat dilakukan dengan menggunakan *metode Zero Inflated Model* (ZIM) yang dapat mengakomodir banyaknya nilai pengamatan bernilai nol (*excess zeros*) dalam deret data *series*. Penelitian ini menggunakan data jumlah hari hujan di Kabupaten Indramayu pada periode 1965-2010 dimana sebanyak 18,30 persen dari data merupakan observasi bernilai nol. Hasil penelitian memperlihatkan bahwa model *Negative Binomial* (NB)*autoregression* adalah model terbaik yang dihasilkan pendekatan ZIM dengan nilai AIC terkecil yaitu sebesar 3.067,544. Model ini juga memberikan nilai MSE yang jauh lebih kecil yaitu sebesar 12,334dibandingkan dengan MSE yang dihasilkan dengan menggunakan pendekatan ARIMA yaitu sebesar 49,9424.

Kata Kunci : *poison regression model, negative binomial regression model, zero-inflated model, excess zeros, rainfall forecasting, ARIMA.*

The Modeling of Number of Monthly Rainy Days in Indramayu Regency Using Zero Inflated Model (ZIM) Approach

Name : Frits Fahridws Damanik
NRP : 1312201716
Supervisor : Dr. rer. pol. Heri Kuswanto, M.Si

ABSTRACT

Generally, agricultural productivity was not only influenced by genetic factor (kinds of seeds), but also environmental factor such as kinds of soil, land management, technology and climate. Among various environmental factor, weather situation plays important role on determining agricultural success. Such condition could not separate from the rainfall and its frequency that fluctuate in occurrence. Thus, an early warning system is needed for anticipating a sudden or extreme change on rainfall and its frequency. Some early warning system of rainfall is built by utilizing model of number of rainy days to predict further condition. This research aims to build a model of the number of monthly rainy days in Indramayu Regency, using the data series from 1965 to 2010. We found that the case is occurs on discrete pattern and its data have 18,30 percent observations that contain zero. Thus we use Zero Inflated Model (ZIM) method to build a monthly rainy days model because its ability to accommodate the excess zero effect. The results shows that the Negative Binomial (NB) autoregression was the finest model produced using ZIM approach. It resulting AIC value of 3.067,544 and MSE 12,334. Such value is smaller compared to the model produced using ARIMA approach.

Keyword : *poison regression model, negative binomial regression model, zero-inflated model, excess zeros, rainfall forecasting, ARIMA.*

DAFTAR ISI

	Halaman
ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	xi
BAB 1 PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah	4
BAB 2 TINJAUAN PUSTAKA	
2.1. Zero-Inflated Poisson (ZIP) <i>autoregression</i>	5
2.2. Estimasi Parameter ZIP <i>autoregression</i>	7
2.3. Uji Parameter ZIP <i>autoregression</i>	9
2.3.1. Pengujian Kesesuaian Model	9
2.3.2. Pengujian Signifikansi Parameter secara Individu	10
2.3.3. Pengujian Overdispersi	10
2.4. Zero-Inflated Negative Binomial(ZINB) <i>autoregression</i>	11
2.5. Estimasi Parameter ZINB <i>autoregression</i>	12
2.6. Kriteria Pemilihan Model Terbaik ZIM	15
2.7. Fleksibilitas ZINB <i>autoregression</i>	15
2.8. Metode ARIMA	16
2.8.1. Proses <i>Autoregressive</i> (Proses AR)	16
2.8.2. Proses <i>Moving Average</i> (Proses MA)	16
2.8.3. Proses <i>Autoregressive Integrated Moving Avarage</i>	17
2.8.4. Prosedur Box-Jenkins	17

BAB 3 METODE PENELITIAN	
3.1. Sumber Data	23
3.2. Variabel Penelitian	23
3.3. Langkah-langkah Penelitian	23
BAB 4 HASIL DAN PEMBAHASAN	
4.1. Gambaran Umum Kejadian Hujan di Kabupaten Indramayu	27
4.2. Model ZIP <i>autoregression</i>	29
4.2.1. Estimasi Parameter ZIP <i>autoregression</i> dan Poisson <i>autoregression</i> Data <i>in-sample</i>	30
4.2.2. Uji Parameter ZIP <i>autoregression</i> dan Poisson <i>autoregression</i> Data <i>in-sample</i>	31
4.2.3. Deteksi Overdispersi Pada Data <i>in-sample</i>	33
4.2.4. Estimasi dan Uji Parameter NB <i>autoregression</i> Data <i>in-sample</i>	33
4.2.5. Pemilihan Model Terbaik ZIM	34
4.3. Metode ARIMA	34
4.3.1. Identifikasi Model	35
4.3.2. Estimasi Parameter dan Signifikansi Parameter Model Dugaan	37
4.3.3. Uji White Noise dan Kenormalan Residual Model	38
4.4. Perbandingan Model Terbaik Antara Model ZIM dan Model ARIMA	43
BAB 5 KESIMPULAN DAN SARAN	
5.1. Kesimpulan	45
5.2. Saran	46
DAFTAR PUSTAKA	47
LAMPIRAN	49

DAFTAR TABEL

	Halaman
Tabel 2.1. Model observation driven untuk <i>Zero-inflated</i> dan Overdispersi pada data <i>time series</i>	15
Tabel 4.1. Statistika Deskriptif Data Jumlah Hari Hujan per Bulan	27
Tabel 4.2. Estimasi dan Uji Parameter Individual Model ZIP <i>autoregression</i> dan Poisson <i>autoregression</i> Pada Data <i>in-sample</i>	31
Tabel 4.3. Estimasi Parameter Model Poisson <i>autoregression</i> dan Negatif Binomial <i>autoregression</i> pada Data <i>in-sample</i>	33
Tabel 4.4. Nilai AIC dan MSE dari Model Poisson <i>autoregression</i> dan Negatif Binomial <i>autoregression</i>	34
Tabel 4.5. Estimasi dan Signifikansi Parameter Model Dugaan pada Data <i>in-sample</i>	38
Tabel 4.6. Uji <i>White Noise</i> dan Kenormalan Residual Model Dugaan pada Data <i>in-sample</i>	39
Tabel 4.7. Estimasi dan Signifikansi Parameter Model Dugaan pada Data <i>in-sample</i> transformasi	42
Tabel 4.8. Uji <i>White Noise</i> dan Kenormalan Residual Model Dugaan pada Data <i>in-sample</i> transformasi	43
Tabel 4.9. Nilai Ramalan Data <i>out-sample</i> Metode ARIMA dan ZIM	44
Tabel 4.10. Perbandingan Nilai MSE dari Model ZIM dan ARIMA	44

“Halaman ini sengaja dikosongkan”

DAFTAR GAMBAR

	Halaman
Gambar 3.1	Tahapan <i>Iteratif</i> Box-Jenkins untuk pembentukan model 18
Gambar 3.2	Diagram Alur metode Analisis 25
Gambar 4.1	Plot <i>Time Series</i> Data Jumlah Hari Hujan per Bulan di Kabupaten Indramayu Periode 1965 – 2010 27
Gambar 4.2	Banyaknya Jumlah Hari Hujan Per Bulan di Kabupaten Indramayu Periode 1965 – 2010 28
Gambar 4.3	<i>Goodness-of-fit</i> Jumlah Hari Hujan Per Bulan di Kabupaten Indramayu Periode 1965 – 2010 29
Gambar 4.4	Plot <i>Time Series</i> Data Jumlah Hari Hujan per Bulan di Kabupaten Indramayu Periode 1965 – 2009 30
Gambar 4.5	Plot ACF Data <i>in-sample</i> Jumlah Hari Hujan per Bulan 35
Gambar 4.6	Plot ACF Data <i>in-sample</i> Jumlah Hari Hujan per Bulan setelah <i>differencing</i> 1 reguler 35
Gambar 4.7	Plot ACF Data <i>in-sample</i> Jumlah Hari Hujan per Bulan setelah <i>differencing</i> 1 reguler dan musiman 12 36
Gambar 4.8	Plot PACF Data <i>in-sample</i> Jumlah Hari Hujan per Bulan setelah <i>differencing</i> 1 reguler dan musiman 12 37
Gambar 4.9	Box-Cox Plot Data <i>in-sample</i> + 1 39
Gambar 4.10	Plot ACF Data <i>in-sample</i> transformasi 40
Gambar 4.11	Plot ACF Data <i>in-sample</i> transformasi setelah <i>differencing</i> 1 reguler dan musiman 12 41
Gambar 4.12	Plot PACF Data <i>in-sample</i> transformasi setelah <i>differencing</i> 1 reguler dan musiman 12 41

“Halaman ini sengaja dikosongkan”

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Produktivitas pertanian secara umum dipengaruhi oleh faktor genetik (jenis bibit) dan faktor lingkungan seperti jenis tanah, pengelolaan lahan, teknologi, dan iklim/cuaca. Dari faktor lingkungan, keadaan cuaca sangat memegang peranan penting terhadap keberhasilan pertanian khususnya pada saat pergantian musim. Curah hujan yang terlalu besar pada awal masa tanam dapat mengakibatkan gagal panen dikarenakan banjir, sedangkan sebaliknya ketiadaan hujan dapat berakibat pada kondisi kekeringan pada tanaman yang juga dapat berakibat gagal panen jika persediaan air tidak mencukupi. Kondisi ini terjadi tidak lepas dari sangat berfluktuatifnya kejadian hujan antar waktu sehingga pembudidayaan tanaman perlu disesuaikan terhadap fluktuasi curah hujan. Tidak jarang terjadi, antisipasi yang terlambat terhadap fluktuasi curah hujan yang tiba-tiba ekstrem berakibat pada kegagalan hasil pertanian secara luas.

Menurut Surmaini dkk (2006) informasi ramalan curah hujan sangat berguna bagi petani dalam mengantisipasi kemungkinan kejadian iklim ekstrim dengan menyesuaikan waktu tanam dan alokasi peruntukan lahan. Mengingat pentingnya ketersediaan model untuk prediksi kejadian hujan maka dirasa perlu untuk membuat kajian yang mendalam mengenai pemodelan dimaksud. Terdapat 2 (dua) jenis pendekatan model yang dapat digunakan dalam memodelkan curah hujan, yaitu terjadinya hari hujan dan jumlah (volume) hujan itu sendiri.

Analisis runtun waktu (*time series*) yang umum digunakan untuk memperoleh angka ramalan dari suatu data *series* adalah dengan menggunakan metode *Autoregressive Integrated Moving Average* (ARIMA). Beberapa peneliti yang menggunakan metode ini antara lain Weesakul & Lowanichchai (2005) meramalkan curah hujan untuk perencanaan ketersediaan air di Thailand dengan menggunakan metode ARIMA dimana dalam prosesnya pada beberapa stasiun curah hujan dilakukan proses *differencing* untuk menstasionerkan data, Narayanan dkk (2013) menggunakan metode ARIMA untuk mencari model terbaik data

curah hujan *pre-monsoon* di bagian barat India disertai dengan *trend analysis* dari hasil model ARIMA.

Data kejadian hujan per bulan merupakan data diskrit sehingga model yang dapat digunakan dalam menganalisis hubungan kausal (sebab akibat) antara variabel penjelas dengan variabel responnya digunakan regresi Poisson. Salah satu keterbatasan dari regresi Poisson ini adalah adanya syarat equidispersi, yaitu mean dan varians dari variabel respon harus bernilai sama (Khoshgoftaar dkk dalam Wulandari dkk, 2010). Pada kenyataannya sering terjadi kasus overdispersi atau underdispersi pada variabel respon dimana nilai mean bisa lebih besar ataupun lebih kecil dari variansnya. Salah satu penanganan terhadap kasus ini adalah dengan menggunakan model regresi Binomial Negatif. Beberapa penelitian yang dilakukan dengan menerapkan metode ini antara lain Byers dkk (2003) menerapkan pendekatan Binomial Negatif dalam studi kasus penelitian penuaan untuk menangani terjadinya kasus overdispersi pada data diskrit dimana varian data 7 kali lebih besar daripada nilai tengahnya, Linden & Mantyniemi (2011) menerapkan model Binomial Negatif pada data count di bidang ekologi dimana perilaku migrasi burung menjadi objek penelitiannya dengan hasil penelitian pendekatan binomial negatif dapat menjelaskan hubungan antara varians dan mean yang tidak dapat dijelaskan oleh model Poisson dikarenakan adanya kasus overdispersi.

Salah satu penyebab overdispersi yang umum terjadi adalah banyaknya nilai pengamatan dari variabel respon yang bernilai nol (*excess zeros*). Dalam kasus seperti ini maka pemodelan yang dapat digunakan antara lain, yaitu model regresi *Zero-Inflated Poisson* (ZIP) ataupun model regresi *Zero-Inflated Negatif Binomial* (ZINB) yang dikembangkan oleh Lambert (1992). Penelitian yang menggunakan metode ini antara lain, yaitu Wulandari dkk (2010) melakukan pemodelan resiko penyakit kaki gajah (filateriasis) di Provinsi Papua dengan regresi ZIP, Ariawan dkk (2012) melakukan pemodelan dengan regresi ZINB untuk data respon diskrit dengan *excess zeros* pada kasus asuransi mobil di Semarang, Zafakali & Ahmad (2013) memodelkan kasus penyakit Thalassemia pada anak-anak dengan menggunakan model regresi ZIP sekaligus melakukan

penanganan terjadinya kasus overdispersi pada data dengan menggunakan model regresi ZINB.

Seiring dengan analisis kausal, pada analisis runtun waktu (*time series*) juga terdapat metode-metode yang dapat digunakan untuk kasus data diskrit seperti model Poisson *autoregression*. Fokianos dkk (2009) menggunakan pendekatan GARCH pada model Poisson *timeseries* untuk memodelkan jumlah transaksi per menit pada saham Ericson B. Davis & Wu (2009) menerapkan model Negatif Binomial untuk data *count time series* pada kasus jumlah penderita polio di Amerika Serikat. Pada kasus data diskrit dengan *excess zeros* pada data *time series*, Yau dkk (2004) pertama kali memodelkan ZIP *mixed autoregression* untuk melihat pengaruh intervensi partisipasi *ergonomic* pada pengurangan resiko kecelakaan kerja. Pemodelan ini masuk ke dalam kelompok *parameter driven model*. Yang (2012) membuat model statistik untuk data *count time series* dengan *excess zeros* dengan dua pendekatan yaitu *observation driven model* dan *parameter driven model*. Pemodelan ini disimulasikan pada kasus jumlah penderita penyakit spilis di Kota Maryland, Amerika Serikat dengan hasil simulasi model ZIP *autoregression* memberikan pengurangan nilai *Akaike information criterion* (AIC) yang besar (lebih dari 200) dibandingkan jika model yang dibentuk hanya model Poisson *autoregression*.

Penelitian ini akan mengaplikasikan pendekatan *observation driven model* untuk mencari model terbaik pada kasus jumlah hari hujan per bulan di Kabupaten Indramayu. Model terbaik ini nantinya akan dibandingkan dengan model *time series* klasik ARIMA.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka permasalahan yang dapat dirumuskan adalah sebagai berikut:

1. Bagaimana mencari model terbaik jumlah hari hujan per bulan di Kabupaten Indramayu dengan menggunakan *Zero-Inflated Model* (ZIM)?
2. Bagaimana perbandingan model terbaik yang dihasilkan ZIM dibandingkan dengan model terbaik yang dihasilkan metode klasik ARIMA?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh model terbaik jumlah hari hujan per bulan di Kabupaten Indramayu dengan menggunakan ZIM.
2. Melihat hasil perbandingan antara model terbaik yang dihasilkan ZIM dengan model terbaik yang dihasilkan metode klasik ARIMA.

1.4. Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan manfaat yang luas tidak hanya untuk peneliti saja tetapi juga bagi pihak lain yang melakukan kajian di bidang ini baik dalam hal metode statistika yang digunakan ataupun dari latar belakang kasus penelitian yaitu bidang klimatologi dan pertanian. Beberapa manfaat yang bisa didapatkan dari penelitian ini adalah sebagai berikut:

1. Memperkaya literatur tentang iklim khususnya kejadian hari hujan per bulan di Kabupaten Indramayu.
2. Memberikan informasi kajian jumlah hari hujan per bulan secara analisis *timeseries* melalui model dan angka ramalan yang dapat digunakan sebagai bahan evaluasi perencanaan bagi Pemerintah Daerah khususnya Dinas Pertanian dalam menyusun kebijakan dan penyuluhan pertanian khususnya berkenaan dengan masa tanam petani di masa yang akan datang.
3. Menambah pemahaman dan pengetahuan lebih dalam bagi peneliti mengenai penerapan metode statistika yang berkaitan dengan data diskrit *time series* dengan *excess zeros* dibidang klimatologi dan pertanian.

1.5. Batasan Masalah

Penelitian ini dilakukan terhadap kasus jumlah hari hujan per bulan di Kabupaten Indramayu pada periode Januari 1965 sampai dengan Desember 2010. Fokus penelitian ini adalah pemodelan ZIM dengan pendekatan *observation driven model* yang kemudian akan dibandingkan dengan model ARIMA.

BAB 2

TINJAUAN PUSTAKA

Pada bagian ini akan diuraikan secara umum metode yang digunakan ZIM dan tahapan-tahapan yang digunakan pada metode ARIMA dengan prosedur Box Jenkins.

2.1. Zero-Inflated Poisson (ZIP) autoregression

Menurut Yang (2012), misalkan $\{Y_t\}$ merupakan variabel respon *series* yang berupa data count, diasumsikan bahwa data series $\{Y_t\}$ mengikuti distribusi ZIP (λ_t, ω_t) dengan *probability mass function* (p.m.f) sebagai berikut:

$$f_{Y_t}(y_t | \mathbf{F}_{t-1}) = \omega_t I_{(y_t=0)} + (1 - \omega_t) \exp(-\lambda_t) \lambda_t^{y_t} / y_t! \quad (2.1)$$

atau ekuivalen dengan :

$$f_{Y_t}(y_t | \mathbf{F}_{t-1}) = \begin{cases} \omega_t I_{(y_t=0)} + (1 - \omega_t) \exp(-\lambda_t), & \text{jika } y_t = 0 \\ (1 - \omega_t) \exp(-\lambda_t) \lambda_t^{y_t} / y_t!, & \text{jika } y_t > 0 \end{cases} \quad (2.2)$$

dimana λ_t adalah parameter dari distribusi Poisson dan ω_t sering disebut sebagai parameter *zero-inflation*. Nilai ω_t bernilai $0 \leq \omega_t \leq 1$. Untuk pembahasan selanjutnya, notasi $I_{y_t=0}$ akan dinotasikan dengan $y_{0,t}$.

Untuk setiap bilangan non negatif m , maka bentuk *cumulative distribution function* (c.d.f) untuk $Y_t | \mathbf{F}_{t-1}$ dapat ditentukan dengan

$$\begin{aligned} F_{Y_t}(m | \mathbf{F}_{t-1}) &= \Pr(Y_t \leq m | \mathbf{F}_{t-1}) \\ &= \sum_{y_t=0}^m f_{Y_t}(y_t | \mathbf{F}_{t-1}) \\ &= \omega_t + (1 - \omega_t) \exp(-\lambda_t) \sum_{y_t=0}^m \lambda_t^{y_t} / y_t!. \end{aligned} \quad (2.3)$$

Secara umum, distribusi ZIP pada persamaan (2.1) dan (2.2) dapat dipandang sebagai distribusi *mixture* (campuran) dari distribusi Poisson yang terdiri dari dua komponen yaitu komponen asli distribusi Poisson dan komponen

distribusi turunan (*degenerate distribution*) berupa parameter *zero-inflation*. Jika diasumsikan variabel *dichotomous* u_t merupakan data count y_t yang berasal dari distribusi turunan ($u_t = 1$) atau dari distribusi asli Poisson ($u_t = 0$) dimana u_t dan u_s saling independen (untuk $s < t$), maka dapat ditentukan model untuk F_{t-1} yaitu :

$$u_t | F_{t-1} \sim \text{Bernoulli}(\omega_t) \quad (2.4)$$

dan

$$Y_t | u_t, F_{t-1} \sim \text{Poisson}((1-u_t)\lambda_t) \quad (2.5)$$

Jika dari persamaan (2.4) dan (2.5) diterapkan pada persamaan (2.1) dan (2.2) akan terlihat bahwa pada saat parameter $\omega_t = 0$ maka distribusi ZIP akan kembali ke bentuk asli distribusi Poisson.

Berdasarkan persamaan (2.1), (2.2), (2.3) dan (2.4) dapat diturunkan mean dan varian dari $Y_t | F_{t-1}$ sebagai berikut :

$$\begin{aligned} E(Y_t | F_{t-1}) &= E\{E(Y_t | u_t, F_{t-1})\} \\ &= E\{(1-u_t)\lambda_t | F_{t-1}\} \\ &= \lambda_t(1-\omega_t) \end{aligned} \quad (2.6)$$

dan

$$\begin{aligned} \text{Var}(Y_t | F_{t-1}) &= E\{\text{Var}(Y_t | u_t, F_{t-1})\} + \text{Var}\{E(Y_t | u_t, F_{t-1})\} \\ &= E\{(1-u_t)\lambda_t | F_{t-1}\} + \text{Var}\{(1-u_t)\lambda_t | F_{t-1}\} \\ &= \lambda_t(1-\omega_t) + \lambda_t^2\omega_t(1-\omega_t) \\ &= \lambda_t(1-\omega_t)(1+\lambda_t\omega_t) \end{aligned} \quad (2.7)$$

sehingga indeks dispersi untuk distribusi ZIP berupa rasio dari varian terhadap mean adalah sebagai berikut :

$$\frac{\text{Var}(Y_t | F_{t-1})}{E(Y_t | F_{t-1})} = 1 + \lambda_t\omega_t. \quad (2.8)$$

Indeks dispersi akan bernilai lebih atau sama dengan satu, sehingga pada distribusi ZIP memungkinkan terjadinya overdispersi yang tidak dapat diakomodasi oleh distribusi Poisson secara umum.

Adapun model untuk ZIP *autoregression* dengan parameter asli λ_t dan parameter *zero-inflation* ω_t menurut Yang (2012) adalah :

$$\eta_t = \log \lambda_t = \mathbf{x}_{t-1}^T \boldsymbol{\beta} \quad (2.9)$$

dan

$$\xi_t = \text{logit}(\omega_t) = \mathbf{z}_{t-1}^T \boldsymbol{\gamma}, \quad (2.10)$$

dimana $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ dan $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ adalah koefisien regresi untuk fungsi log linier dan logistik sedangkan $\mathbf{x}_{t-1} = (x_{t-1,1}, \dots, x_{t-1,p})^T$ dan $\mathbf{z}_{t-1} = (z_{t-1,1}, \dots, z_{t-1,q})^T$ adalah vektor variabel penjelas sebelumnya yang menerangkan korelasi antara lag respon *series*. Untuk lebih memudahkan, dituliskan notasi $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ sebagai vektor parameter yang tidak diketahui dengan dimensi $(p + q)$.

2.2. Estimasi Parameter ZIP *autoregression*

Berdasarkan definisi umum tentang bentuk parsial likelihood yang diberikan oleh Kedem dan Fokianos (2002) dalam Yang (2012)

$$\text{PL}(\boldsymbol{\theta}) = \prod_{t=1}^N f_{Y_t}(y_t | \mathbf{F}_{t-1}) \quad (2.11)$$

dan persamaan (2.1) maka log partial likelihood untuk ZIP *autoregression* adalah:

$$\begin{aligned} \log \text{PL}(\boldsymbol{\theta}) &= \sum_{t=1}^N \log f_{Y_t}(y_t | \mathbf{F}_{t-1}) \\ &= \sum_{t=1}^N \log \{ \omega_t y_{0,t} + (1 - \omega_t) \exp(-\lambda_t) \lambda_t^{y_t} / y_t! \}. \end{aligned} \quad (2.12)$$

Estimasi parameter dilakukan dengan cara memaksimalkan fungsi log PL($\boldsymbol{\theta}$) dengan metode *maximum partial likelihood estimator* (MPLE), dimana

$$\mathbf{S}_N(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log \text{PL}(\boldsymbol{\theta}) = \sum_{t=1}^N \mathbf{C}_{t-1} \mathbf{v}_t(\boldsymbol{\theta}) \quad (2.13)$$

dengan

$$\mathbf{C}_{t-1} = \begin{bmatrix} x_{t-1} & 0 \\ 0 & z_{t-1} \end{bmatrix}$$

dan

$$\mathbf{v}_t(\boldsymbol{\theta}) = \begin{bmatrix} v_{1,t}(\boldsymbol{\theta}) \\ v_{2,t}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} y_t - \lambda_t (1 - \omega_t y_{0,t} / p_{0,t}) \\ \omega_t (y_{0,t} / p_{0,t} - 1) \end{bmatrix}$$

$p_{0,t} = \omega_t + (1 - \omega_t) \exp(-\lambda_t)$ adalah p.m.f dari $Y_t | \mathbf{F}_{t-1}$ pada saat bernilai nol.

Bentuk $\mathbf{v}_t(\boldsymbol{\theta})$ di atas adalah bentuk non linier sehingga diperlukan pendekatan iterative algoritma untuk mengestimasi parameter. Yang (2012) melakukan beberapa pendekatan yaitu: Newton-Rapshson (NR) dan Fisher Scoring (FS) *algorithms* serta teorema Bayes sehingga didapatkan matriks *observation information* negatif Hessian

$$\mathbf{H}_N(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log \text{PL}(\boldsymbol{\theta}) = \sum_{t=1}^N \mathbf{C}_{t-1} \mathbf{D}_t(\boldsymbol{\theta}) \mathbf{C}_{t-1}^T, \quad (2.14)$$

dimana $\mathbf{D}_t(\boldsymbol{\theta})$ adalah matriks simetris 2 x 2 dengan elemen-elemen :

$$d_{11,t}(\boldsymbol{\theta}) = \lambda_t [1 - y_{0,t} \omega_t \{ \omega_t + (1 - \omega_t)(1 + \lambda_t) \exp(-\lambda_t) \} / p_{0,t}^2]$$

$$d_{12,t}(\boldsymbol{\theta}) = -y_{0,t} \omega_t (1 - \omega_t) \lambda_t \exp(-\lambda_t) / p_{0,t}^2$$

$$d_{22,t}(\boldsymbol{\theta}) = \omega_t (1 - \omega_t) \{ (1 - y_{0,t} \exp(-\lambda_t)) / p_{0,t}^2 \}.$$

Adapun untuk matriks *conditional information* Fisher didapatkan

$$\mathbf{G}_N(\boldsymbol{\theta}) = \sum_{t=1}^N \text{Var}\{ \mathbf{C}_{t-1} \mathbf{v}_t(\boldsymbol{\theta}) | \mathbf{F}_{t-1} \} = \sum_{t=1}^N \mathbf{C}_{t-1} \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{C}_{t-1}^T, \quad (2.15)$$

dimana $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \text{Var}\{ \mathbf{v}_t(\boldsymbol{\theta}) | \mathbf{F}_{t-1} \}$ adalah matriks simetris 2 x 2 dengan elemen-elemen :

$$\sigma_{11,t}(\boldsymbol{\theta}) = (1 - \omega_t) \lambda_t [\exp(-\lambda_t) + \omega_t \{ 1 - (1 + \lambda_t) \exp(-\lambda_t) \}] / p_{0,t}^2$$

$$\sigma_{12,t}(\boldsymbol{\theta}) = -\omega_t (1 - \omega_t) \lambda_t \exp(-\lambda_t) / p_{0,t}^2$$

$$\sigma_{22,t}(\boldsymbol{\theta}) = \omega_t^2 (1 - \omega_t) \{ (1 - \exp(-\lambda_t)) \} / p_{0,t}^2.$$

Dengan menggunakan metode *iteratively reweighted least squares* (IRLS) didapatkan vektor *partial score* untuk ZIP autoregression:

$$\mathbf{S}_N(\boldsymbol{\theta}) = \sum_{t=1}^N \mathbf{C}_{t-1} \mathbf{v}_t(\boldsymbol{\theta}) = \mathbf{C}' \mathbf{v}(\boldsymbol{\theta}) \quad (2.16)$$

serta bentuk dekomposisi orthogonal dari

$$\log \text{PL}^c(\boldsymbol{\theta}) = \sum_{t=1}^N (1-u_t)(y_t \log \lambda_t - \lambda_t) + \sum_{t=1}^N \{u_t \log \omega_t + (1-u_t) \log(1-\omega_t)\} \quad (2.17)$$

untuk kemudian dilakukan tahapan *Expectation Maximization (EM) algorithm* yang terdiri dari dua tahap, yaitu :

- Tahap ekspektasi (*E-step*)

Mengganti variabel u_t dengan $\hat{u}_t^{(j)}$ sehingga didapatkan

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}_t^{(j)}) &= E\{\log \text{PL}^c \boldsymbol{\theta} | y, \boldsymbol{\theta}^{(j)}\} \\ &= \sum_{t=1}^N (1-\hat{u}_t^{(j)})(y_t \log \lambda_t - \lambda_t) + \sum_{t=1}^N \{\hat{u}_t^{(j)} \log \omega_t + (1-\hat{u}_t^{(j)}) \log(1-\omega_t)\}, \end{aligned}$$

dimana $\hat{u}_t^{(j)}$ merupakan ekspektasi u_t pada iterasi ke- j .

- Tahap maksimalisasi (*M-step*)

Memaksimalkan nilai β dan γ dengan menghitung $\beta^{(j+1)}$ dan $\gamma^{(j+1)}$ melalui

$$\sum_{t=1}^N (1-\hat{u}_t^{(j)})(y_t \log \lambda_t - \lambda_t)$$

dan

$$\sum_{t=1}^N \{\hat{u}_t^{(j)} \log \omega_t + (1-\hat{u}_t^{(j)}) \log(1-\omega_t)\}.$$

2.3. Uji Parameter ZIP autoregression

Pengujian parameter model hasil estimasi dilakukan dengan dua tahapan, yaitu pengujian kesesuaian model dan uji signifikansi parameter secara individu.

2.3.1. Pengujian Kesesuaian Model

Pengujian kesesuaian model dilakukan dengan menggunakan *Likelihood Ratio (LR) Test* dengan menggunakan hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = \gamma_1 = \gamma_2 = \dots = \gamma_q = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_i \neq 0 \text{ atau } \gamma_j \neq 0$$

untuk setiap $i = 1, 2, \dots, p$ dan $j = 1, 2, \dots, q$ dengan statistik uji yang digunakan yaitu:

$$\begin{aligned} G &= -2(\log L_0 - \log L_1) \\ &= -2\{[\log L(\beta_0|y_t, \omega_t) + \log L(\gamma_0|y_t, \omega_t)] \\ &\quad - [\log L(\beta|y_t, \omega_t) + \log L(\gamma|y_t, \omega_t)]\} \end{aligned}$$

dimana H_0 ditolak pada taraf signifikansi α jika $G_{\text{hitung}} > \chi_{\alpha; 2p}^2$.

2.3.2. Pengujian Signifikansi Parameter secara Individu

Pengujian signifikansi parameter model $\log(\lambda_t) = \mathbf{x}_{t-1}^T \boldsymbol{\beta}$ dimana $t = 1, 2, \dots, n$ dilakukan dengan menggunakan hipotesis:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

untuk setiap $i = 1, 2, \dots, p$ dengan statistik uji $W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2$ dimana H_0 ditolak pada taraf signifikansi α jika $W_j > \chi_{\alpha; 1}^2$, sedangkan untuk pengujian signifikansi parameter model $\text{logit}(\omega_t) = \mathbf{z}_{t-1}^T \boldsymbol{\gamma}$ dimana $t = 1, 2, \dots, n$ dilakukan dengan menggunakan hipotesis:

$$H_0 : \gamma_j = 0$$

$$H_1 : \gamma_j \neq 0$$

untuk setiap $j = 1, 2, \dots, q$ dengan statistik uji $W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2$ dimana H_0 ditolak pada taraf signifikansi α jika $W_j > \chi_{\alpha; 1}^2$.

2.3.3. Pengujian Overdispersi

Pengujian overdispersi pada model dilakukan dengan menggunakan uji Pearson chi-square dengan hipotesis:

$$H_0 : \kappa \leq 1 \text{ (Data tidak overdispersi)}$$

$$H_1 : \kappa > 1 \text{ (Data overdispersi)}$$

$$\text{dimana } \hat{\kappa} = \frac{\chi^2}{n-p-1} \text{ dan } \chi^2 = \sum_{t=1}^n \frac{(y_t - \hat{\lambda}_t)^2}{\text{var}(\hat{\lambda}_t)}.$$

2.4. Zero-Inflated Negative Binomial (ZINB) autoregression

Pada bagian ini akan dibahas tentang *zero inflated negative binomial* (ZINB) *autoregression* yang dapat digunakan sebagai alternatif untuk menangani masalah over dispersi yang terjadi pada ZIP *autoregression*.

Menurut Yang (2012), jika dimisalkan data series $\{Y_t\}$ diasumsikan mengikuti distribusi ZINB dengan p.m.f sebagai berikut:

$$f_{Y_t}(y_t | \mathbf{F}_{t-1}) = \begin{cases} \omega_t + (1 - \omega_t) \left(\frac{k_t}{k_t + \lambda_t} \right)^{k_t}, & \text{jika } y_t = 0 \\ (1 - \omega_t) \frac{\Gamma(k_t + y_t)}{\Gamma(k_t) y_t!} \left(\frac{k_t}{k_t + \lambda_t} \right)^{k_t} \left(\frac{\lambda_t}{k_t + \lambda_t} \right)^{y_t}, & \text{jika } y_t > 0 \end{cases} \quad (2.18)$$

dimana k_t adalah parameter dispersi dari distribusi *Negative Binomial* (NB).

Seperti halnya pada distribusi ZIP, distribusi ini dapat dipandang juga sebagai distribusi *mixture* (campuran) dari distribusi NB yang terdiri dari dua komponen yaitu komponen asli distribusi NB dan komponen distribusi turunan (*degenerate distribution*) pada saat bernilai nol. Jika diasumsikan variabel *dichotomous* u_t merupakan data count y_t yang berasal dari distribusi turunan ($u_t = 1$) atau dari distribusi asli NB ($u_t = 0$) maka dapat ditentukan model untuk \mathbf{F}_{t-1} yaitu :

$$u_t | \mathbf{F}_{t-1} \sim \text{Bernoulli}(\omega_t) \quad (2.19)$$

dan

$$Y_t | u_t, \mathbf{F}_{t-1} \sim \text{NB}(k_t, (1 - u_t)\lambda_t) \quad (2.20)$$

dengan

$$\begin{aligned} E(Y_t | \mathbf{F}_{t-1}) &= E\{E(Y_t | u_t, \mathbf{F}_{t-1})\} \\ &= E\{(1 - u_t)\lambda_t | \mathbf{F}_{t-1}\} \\ &= \lambda_t(1 - \omega_t) \end{aligned} \quad (2.21)$$

dan

$$\begin{aligned} \text{Var}(Y_t | \mathbf{F}_{t-1}) &= E\{\text{Var}(Y_t | u_t, \mathbf{F}_{t-1})\} + \text{Var}\{E(Y_t | u_t, \mathbf{F}_{t-1})\} \\ &= E\{(1 - u_t)\lambda_t + (1 - u_t)^2 \lambda_t^2 / k_t | \mathbf{F}_{t-1}\} + \text{Var}\{(1 - u_t)\lambda_t | \mathbf{F}_{t-1}\} \end{aligned}$$

$$\begin{aligned}
&= \lambda_t(1 - \omega_t) + \lambda_t^2(1 - \omega_t) / k_t + \lambda_t^2 \omega_t(1 - \omega_t) \\
&= \lambda_t(1 - \omega_t)(1 + \lambda_t \omega_t + \lambda_t / k_t)
\end{aligned} \tag{2.22}$$

sehingga indeks dispersi

$$\frac{\text{Var}(Y_t | \mathbf{F}_{t-1})}{E(Y_t | \mathbf{F}_{t-1})} = 1 + \lambda_t \omega_t + \lambda_t / k_t. \tag{2.23}$$

2.5. Estimasi Parameter ZINB autoregression

Dalam ZINB terdapat tiga komponen parameter sistemik, yaitu:

$$\log k_t = \mathbf{s}_{t-1}^T \boldsymbol{\alpha} \tag{2.24}$$

$$\log \lambda_t = \mathbf{x}_{t-1}^T \boldsymbol{\beta} \tag{2.25}$$

$$\text{logit}(\omega_t) = \mathbf{z}_{t-1}^T \boldsymbol{\gamma}, \tag{2.26}$$

dimana $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ dan $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ adalah koefisien regresi sedangkan $\mathbf{s}_{t-1} = (s_{t-1,1}, \dots, s_{t-1,1})^T$, $\mathbf{x}_{t-1} = (x_{t-1,1}, \dots, x_{t-1,p})^T$ dan $\mathbf{z}_{t-1} = (z_{t-1,1}, \dots, z_{t-1,q})^T$ adalah vektor variabel penjelas sebelumnya yang menerangkan korelasi antara lag respon *series*. Untuk lebih memudahkan, dituliskan notasi $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ sebagai vektor parameter yang tidak diketahui dengan dimensi $(1 + p + q)$.

Log-partial likelihood untuk ZINB dalam Yang (2012) sebagai berikut :

$$\begin{aligned}
\log \text{PL}(\boldsymbol{\theta}) &= \sum_{y_t=0} \log \{ \omega_t + (1 - \omega_t) p_t^{k_t} \} \\
&+ \sum_{y_t > 0} \{ \log(1 - \omega_t) + \log \Gamma(k_t + y_t) - \log \Gamma(k_t) \\
&- \log(y_t!) + k_t \log p_t + y_t \log(1 - p_t) \}
\end{aligned} \tag{2.27}$$

yang mana dapat diturunkan menjadi

$$\frac{\partial p_t}{\partial k_t} = p_t(1 - p_t) / k_t$$

$$\frac{\partial p_t}{\partial \lambda_t} = -p_t^2 / k_t$$

$$\frac{\partial p_t^{k_t}}{\partial k_t} = p_t^{k_t} (\log p_t + 1 - p_t)$$

$$\frac{\partial p_t^{k_t}}{\partial \lambda_t} = -p_t^{k_t+1},$$

sedangkan vektor *partial score* dari ZINB *autoregression* ditentukan

$$\mathbf{S}_N(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log \text{PL}(\boldsymbol{\theta}) = \sum_{t=1}^N \mathbf{C}_{t-1} \mathbf{v}_t(\boldsymbol{\theta}) \quad (2.28)$$

dimana

$$\mathbf{C}_{t-1} = \begin{bmatrix} s_{t-1} & 0 & 0 \\ 0 & x_{t-1} & 0 \\ 0 & 0 & z_{t-1} \end{bmatrix} \text{ dan } \mathbf{v}_t(\boldsymbol{\theta}) = \begin{bmatrix} v_{1,t}(\boldsymbol{\theta}) \\ v_{2,t}(\boldsymbol{\theta}) \\ v_{3,t}(\boldsymbol{\theta}) \end{bmatrix}$$

dengan elemen-elemen $\mathbf{v}_t(\boldsymbol{\theta})$:

$$v_{1,t}(\boldsymbol{\theta}) = k_t (1 - \omega_t y_{0,t} / p_{0,t}) (\log p_t + 1 - p_t) + k_t (1 - y_{0,t}) \{ \Psi_0(k_t + y_t) - \Psi_0(k_t) \} - y_t p_t$$

$$v_{2,t}(\boldsymbol{\theta}) = p_t y_t - k_t (1 - p_t) (1 - \omega_t y_{0,t} / p_{0,t})$$

$$v_{3,t}(\boldsymbol{\theta}) = \omega_t (y_{0,t} / p_{0,t} - 1).$$

adapun $\Psi_0(\cdot)$ adalah fungsi digamma dan $p_{0,t} = \omega_t + (1 - \omega_t) p_t^{k_t}$ adalah p.m.f dari $Y_t | \mathbf{F}_{t-1}$ pada saat bernilai nol.

Yang (2012) juga menjabarkan matriks *observation information* negatif Hessian untuk ZINB *autoregression*

$$\mathbf{H}_N(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log \text{PL}(\boldsymbol{\theta}) = \sum_{t=1}^N \mathbf{C}_{t-1} \mathbf{D}_t(\boldsymbol{\theta}) \mathbf{C}_{t-1}^T, \quad (2.29)$$

dimana $\mathbf{D}_t(\boldsymbol{\theta})$ adalah matriks simetris 3 x 3 dengan elemen-elemen :

$$d_{11,t}(\boldsymbol{\theta}) = k_t (1 - p_t) \{ y_t p_t / k_t - (1 - p_t) (1 - \omega_t y_{0,t} / p_{0,t}) \} - k_t (1 - y_{0,t}) \{ \Psi_0(k_t + y_t) - \Psi_0(k_t) + k_t \Psi_1(k_t + y_t) - k_t \Psi_1(k_t) \} - k_t (\log p_t + 1 - p_t) \{ 1 - \omega_t y_{0,t} / p_{0,t} + k_t \omega_t y_{0,t} (\log p_t + 1 - p_t) (p_{0,t} - \omega_t) / p_{0,t}^2 \}$$

$$d_{12,t}(\theta) = k_t(1-p_t)\{k_t\omega_t y_{0,t}(\log p_t + 1 - p_t)(p_{0,t} - \omega_t) / p_{0,t}^2 + (1-p_t)(1-\omega_t y_{0,t} / p_{0,t}) - y_t p_t / k_t\}$$

$$d_{13,t}(\theta) = k_t\omega_t y_{0,t}(\log p_t + 1 - p_t)(p_{0,t} - \omega_t) / p_{0,t}^2$$

$$d_{22,t}(\theta) = k_t(1-p_t)\{p_t y_t / k_t + p_t(1-\omega_t y_{0,t} / p_{0,t}) - k_t\omega_t y_{0,t}(1-p_t)(p_{0,t} - \omega_t) / p_{0,t}^2\}$$

$$d_{23,t}(\theta) = -k_t\omega_t(1-\omega_t)(1-p_t)p_t^{k_t} y_{0,t} / p_{0,t}^2$$

$$d_{33,t}(\theta) = \omega_t(1-\omega_t)(1-p_t^{k_t} y_{0,t} / p_{0,t}^2), \text{ dengan } \Psi_1(\cdot) \text{ adalah fungsi trigamma.}$$

Adapun untuk matriks *conditional information* Fisher didapatkan

$$\mathbf{G}_N(\boldsymbol{\theta}) = \sum_{t=1}^N \text{Var}\{\mathbf{C}_{t-1} \mathbf{v}_t(\boldsymbol{\theta}) | \mathbf{F}_{t-1}\} = \sum_{t=1}^N \mathbf{C}_{t-1} \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \mathbf{C}_{t-1}^T, \quad (2.30)$$

dimana $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ adalah matriks simetris 3 x 3 dengan elemen-elemen :

$$\sigma_{11,t}(\theta) = k_t\{k_t c_t - (1-\omega_t)(1-p_t)\} - \omega_t(1-\omega_t)k_t^2(\log p_t + 1 - p_t)^2 p_t^{k_t} / p_{0,t}$$

$$\sigma_{12,t}(\theta) = \omega_t(1-\omega_t)k_t^2(1-p_t)(\log p_t + 1 - p_t)p_t^{k_t} / p_{0,t}$$

$$\sigma_{13,t}(\theta) = \omega_t(1-\omega_t)k_t(\log p_t + 1 - p_t)p_t^{k_t} / p_{0,t}$$

$$\sigma_{22,t}(\theta) = (1-\omega_t)k_t(1-p_t)[p_t^{k_t} + \omega_t\{1-p_t^{k_t} - k_t(1-p_t)p_t^{k_t}\}] / p_{0,t}$$

$$\sigma_{23,t}(\theta) = -\omega_t(1-\omega_t)k_t(1-p_t)p_t^{k_t} / p_{0,t}$$

$$\sigma_{33,t}(\theta) = \omega_t^2(1-\omega_t)(1-p_t^{k_t}) / p_{0,t}$$

dengan

$$c_t = E\{(1-y_{0,t})[\Psi_1(k_t + y_t) | \mathbf{F}_{t-1}]\} = \sum_{j=0}^{\infty} P_r(Y_t > j | \mathbf{F}_{t-1}) / (k_t + j)^2 \quad (2.31)$$

yang mana bisa didekati dengan

$$\tilde{c}_t = \sum_{j=0}^M P_r(Y_t > j | \mathbf{F}_{t-1}) / (k_t + j)^2 \quad (2.32)$$

dengan M adalah bilangan yang cukup besar (misalnya 100).

Untuk uji parameter dan overdispersi pada model ZINB *autoregression* sama prosedurnya seperti pada model ZIP *autoregression*.

2.6. Kriteria Pemilihan Model Terbaik ZIM

Pemilihan model terbaik pada ZIM dilakukan dengan cara, yaitu :

- Untuk data *in-sample* berdasarkan nilai *Akaike information criterion* (AIC) terkecil.

Nilai AIC untuk parsial likelihood mengikuti definisi dari Kedem dan Fokianos dalam Yang (2012):

$$AIC = -2\log PL(\hat{\theta}) + 2\dim(\theta), \quad (2.33)$$

dimana $\hat{\theta}$ adalah MPLE dan $\dim(\theta)$ adalah jumlah parameter bebas dalam model.

- Untuk data *out-sample* berdasarkan nilai *Mean Squared Error* (MSE) terkecil.

Nilai MSE ditentukan :

$$MSE = \frac{1}{n} \sum_t^n (Y_t - \hat{Y}_t)^2, \quad (2.34)$$

2.7. Fleksibilitas ZINB *autoregression*

Yang (2012) memaparkan keterkaitan antara ZINB *autoregression* dengan ZIP *autoregression* melalui Tabel 2.1. Dari Tabel 2.1 terlihat bahwa model ZINB *autoregression* merupakan model yang sangat umum dan sangat fleksibel dimana di dalamnya terkandung banyak model yang dapat digunakan pada kasus-kasus tertentu.

Tabel 2.1. Model *Observation Driven* untuk Zero-inflated dan Overdispersi pada Data *Time Series*

Zero-Inflated	Overdispersi	Autokorelasi	Model
Tidak	Tidak	Tidak	Regresi Poisson
Tidak	Tidak	Ya	Poisson <i>autoregression</i>
Tidak	Ya	Tidak	Regresi NB
Tidak	Ya	Ya	NB <i>autoregression</i>
Ya	Tidak	Tidak	Regresi ZIP
Ya	Tidak	Ya	ZIP <i>autoregression</i>
Ya	Ya	Tidak	Regresi ZINB
Ya	Ya	Ya	ZINB <i>autoregression</i>

Sumber: Yang, 2012

Secara khusus dapat diamati bahwa model ZINB autoregression dapat berubah bentuk menjadi:

- Model NB *autoregression* pada saat $\omega_t = 0$
- Model ZIP *autoregression* pada saat $1/k_t \rightarrow 0$
- Model Poisson *autoregression* pada saat $\omega_t = 0$ dan $1/k_t \rightarrow 0$

selain itu, jika tidak terdapat kasus autokorelasi pada data *series* mana keempat model *autoregression* tersebut akan berubah menjadi bentuk model regresi umum masing-masing distribusi asalnya.

2.8. Metode ARIMA

Model ARIMA dibangun dari 3 (tiga) proses, yaitu: proses *Autoregressive*, proses *Integrated* dan proses *Moving Average*.

2.8.1. Proses *Autoregressive* (Proses AR)

Proses *autoregressive* sesuai dengan namanya adalah proses hasil regresi dengan dirinya sendiri. Secara matematis proses AR(p) dapat dituliskan:

$$\begin{aligned}\phi_p(B)Y_t &= a_t \\ (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)Y_t &= a_t \\ Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t\end{aligned}\tag{2.35}$$

yang menunjukkan bahwa nilai saat ini merupakan jumlah tertimbang dari nilai-nilai p yang lalu ditambah satu error yang random saat ini.

2.8.2. Proses *Moving Average* (Proses MA)

Yaitu proses *time series* di mana nilai suatu data *time series* pada suatu waktu merupakan fungsi dari *error/residual* pada waktu-waktu sebelumnya. Secara matematik proses MA (q) dapat dituliskan sebagai berikut :

$$\begin{aligned}Y_t &= \theta_q(B)a_t \\ Y_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t \\ Y_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}\end{aligned}\tag{2.36}$$

2.8.3. Proses Autoregressive Integrated Moving Average

Bila data yang digunakan sudah stationer dan kedua proses antara *autoregressive* dan *moving average* digabung maka akan menghasilkan proses *autoregressive moving average*. Proses ini biasa disebut dengan model ARMA (p,q). Secara matematik model ARMA (p,q) dapat dituliskan sebagai berikut:

$$\phi_p(B)Y_t = \theta_q(B)a_t \quad (2.37)$$

Apabila data yang digunakan belum stasioner biasanya diatasi dengan melakukan *differencing* terlebih dahulu. Proses *differencing* yaitu membuat data baru dengan cara mengurangi nilai observasi pada satu periode dengan nilai observasi periode sebelumnya. Proses *differencing* dalam model ARIMA disebut juga dengan proses *Integrated*. Secara matematik proses *differencing/integrated* dapat dituliskan sebagai berikut :

$$(1-B)^d Y_t = \omega_t \quad (2.38)$$

dimana d menunjukkan derajat *differencing*. Derajat *differencing* sebesar d berarti sebanyak d kali proses *differencing* dilakukan untuk mendapatkan *series* data yang stationer. Bila ketiga proses time series diatas dilakukan penggabungan, maka proses time series tersebut disebut dengan proses *Autoregressive Integrated Moving Average* atau disingkat dengan ARIMA. Biasanya model ARIMA dituliskan sebagai ARIMA (p,d,q) dimana p adalah orde dari *autoregressive*, d adalah orde *differencing* dan q adalah orde dari *moving average*.

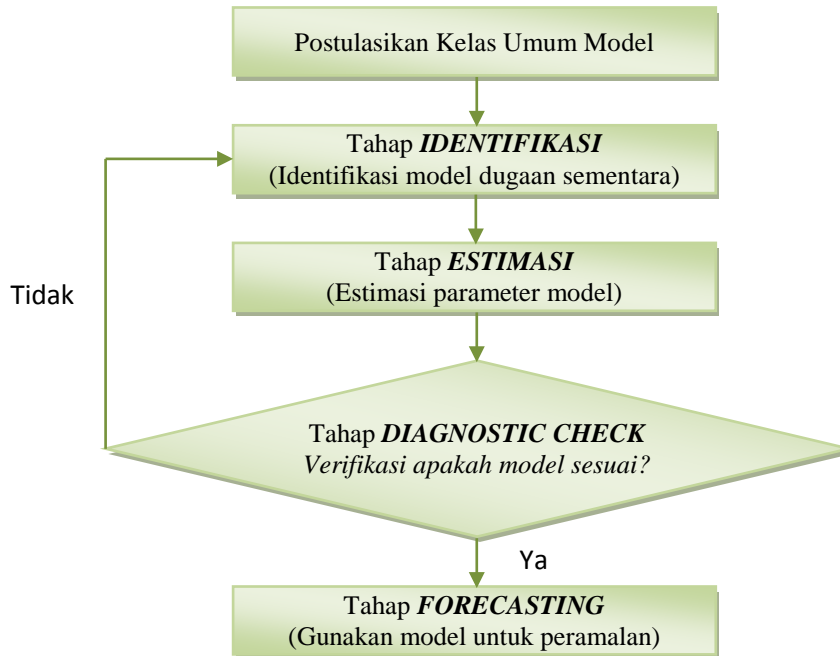
Dari ketiga proses tersebut, maka persamaan model ARIMA dapat diturunkan sebagai berikut:

$$\begin{aligned} \phi_q(B)(1-B)^d Y_t &= \theta_q(B)a_t \\ (1-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1-B)^d Y_t &= (1-\theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t \end{aligned} \quad (2.39)$$

2.8.4. Prosedur Box-Jenkins

Metode ARIMA yang mengikuti prosedur Box-Jenkins (1970) mempunyai 4 (empat) tahapan yaitu: identifikasi, estimasi, *diagnostic check* dan *forecasting*. Proses *iteratif* terjadi pada tahap *diagnostic check*. Bila pada tahap

ini, uji parameter dan uji asumsi tidak terpenuhi maka tahapan kembali pada tahap pertama yaitu identifikasi.



Gambar 2.1. Tahapan *Iteratif* Box-Jenkins untuk pembentukan model ARIMA (Suhartono, 2009)

2.8.4.1. Identifikasi Model

Tahap 1. Plot data *time series*

Pada tahap ini dilakukan plot data sebagai gambaran untuk melihat pola data yang mencakup: *trend*, musiman, *outlier*, varians yang tidak konstan dan masalah stasioneritas data.

Tahap 2. Menghitung nilai ACF dan PACF dari *series data* yang asli

Tahap ini dilakukan untuk melihat apakah series data yang digunakan sudah stasioner atau belum. Jika nilai ACF menurun secara lambat menuju nol dan nilai PACF *cut off* setelah *lag* 1, maka dapat diartikan bahwa data yang digunakan belum stasioner dan perlu dilakukan *differencing*.

Tahap 3. Menghitung nilai ACF dan PACF dari *series data* yang sudah stasioner

Nilai ACF dan PACF pada *series data* yang sudah stasioner pada tahap ini dilakukan untuk menentukan orde dari p dan q pada model ARIMA (p,d,q) .

Nilai p adalah orde tertinggi pada *polynomial autoregressive* dan q adalah orde tertinggi pada *polynomial moving average*.

Orde p pada model ARIMA (p,d,q) ditentukan berdasarkan nilai PACF yang tinggi (keluar dari garis batas) dan turun langsung (*cut-off*) setelah *lag* ke- p . Hal yang sama juga dilakukan dalam menentukan orde q pada model ARIMA.

Setelah orde p , d dan q diketahui, maka dibentuk model-model ARIMA(p,d,q) yang memungkinkan untuk selanjutnya dilakukan estimasi parameter.

2.8.4.2. Estimasi Parameter

Menurut Wei (2006), estimasi parameter dalam analisis *time series* dapat dilakukan dengan menggunakan beberapa metode antara lain: Metode *Moment*, Metode *Maximum Likelihood* dan Metode *Least Square*. Penjabaran metode tidak diuraikan karena bukan fokus utama dalam penelitian ini.

2.8.4.3. Cek Diagnosa (*Diagnostic Check*)

Cek diagnosa dilakukan untuk mengetahui apakah suatu model layak untuk digunakan. Cek diagnosa meliputi uji signifikansi parameter, uji asumsi *white noise* dan kenormalan. Uji signifikansi parameter dilakukan pada parameter dalam model, sedangkan uji asumsi *white noise* dan kenormalan dilakukan pada *residual* dari model.

a. Uji Signifikansi Parameter

Uji signifikansi parameter digunakan untuk menguji apakah parameter tersebut mempunyai pengaruh terhadap model. Uji ini dilakukan menggunakan uji *t student*. Perumusan hipotesis pada uji ini adalah :

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

dimana β adalah parameter model. Adapun statistik uji yang digunakan adalah :

$$t_{hitung} = \frac{\hat{\beta}}{se(\hat{\beta})}$$

Keputusan akan menolak H_0 jika $|t_{hitung}| > t_{(\frac{\alpha}{2}, n-n_p)}$ di mana n_p adalah banyaknya parameter dalam model.

b. Pemeriksaan Uji Asumsi *White Noise*

Wei (2006) menyatakan bahwa asumsi yang mendasar adalah *residual* (a_t) bersifat *white noise*. *White noise* adalah suatu asumsi yang menyatakan bahwa antar *residual* (a_1, a_2, \dots, a_t) tidak terdapat korelasi dan mempunyai *mean* sama dengan nol dan *variance* konstan. Untuk melakukan uji asumsi ini, digunakan statistik uji Ljung-Box. Perumusan hipotesis pada statistik uji Ljung-Box adalah:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$$

$$H_1 : \text{minimal ada 1 nilai } \rho_i \neq 0, i = 1, 2, \dots, k$$

Adapun statistik uji yang digunakan yaitu:

$$Q = n(n+2) \sum_{i=1}^k \frac{\hat{r}_i^2}{n-1}$$

dimana \hat{r}_i^2 adalah nilai taksiran ACF *residual* dan n adalah banyaknya observasi.

Keputusan akan menolak H_0 jika $Q > \chi_{(\alpha; k-p-q)}^2$.

c. Uji Kenormalan

Untuk menguji apakah *residual* berdistribusi normal, maka digunakan uji KS. Perumusan hipotesis pada uji KS adalah :

$$H_0 : \text{residual berdistribusi normal}$$

$$H_1 : \text{residual tidak berdistribusi normal}$$

Sedangkan statistik uji yang digunakan adalah:

$$D = \text{Sup} |S(x) - F_0(x)|$$

dimana $S(x)$ adalah proporsi kumulatif dari *residual* yang sudah diurutkan dan

$F_0(x)$ adalah nilai peluang kumulatif kurva normal. Keputusan akan menolak H_0

jika $D \geq D_{(n; 1-\alpha)}$.

2.8.4.4. Kriteria Pemilihan Model Terbaik

Wei (2006) mengemukakan beberapa kriteria pemilihan model, salah satunya adalah AIC (*Akaike's Information Criterion*).

Model terbaik yang diperoleh adalah model yang memiliki nilai AIC terkecil untuk *data in-sample*, dimana:

$$AIC(M) = n \ln \hat{\sigma}_a^2 + 2M$$

dengan

M = jumlah parameter dalam model

n = jumlah pengamatan efektif

$\hat{\sigma}_a^2$ = varians residual

serta mempunyai nilai MSE terkecil untuk data *out-sampalnya*.

2.8.4.5. Peramalan (*Forecasting*)

Berdasarkan model ARIMA terbaik yang sudah didapatkan maka penghitungan angka ramalan dapat dilakukan.

“Halaman ini sengaja dikosongkan”

BAB 3

METODE PENELITIAN

3.1. Sumber Data

Data penelitian ini menggunakan data sekunder yang bersumber dari Dinas Pertanian dan Badan Pusat Statistik Kabupaten Indramayu.

3.2. Variabel Penelitian

Variabel penelitian yang digunakan adalah data *series* jumlah hari hujan per bulan di Kabupaten Indramayu pada periode waktu Januari 1965 sampai dengan Desember 2010.

3.3. Langkah-langkah Penelitian

Metode penelitian ini menggunakan dua metode pemodelan yang nantinya akan dibandingkan sejauh mana kebaikan model yang dihasilkan sesuai dengan kriteria kebaikan model yang digunakan. Adapun langkah-langkah penelitian yang dilakukan adalah sebagai berikut:

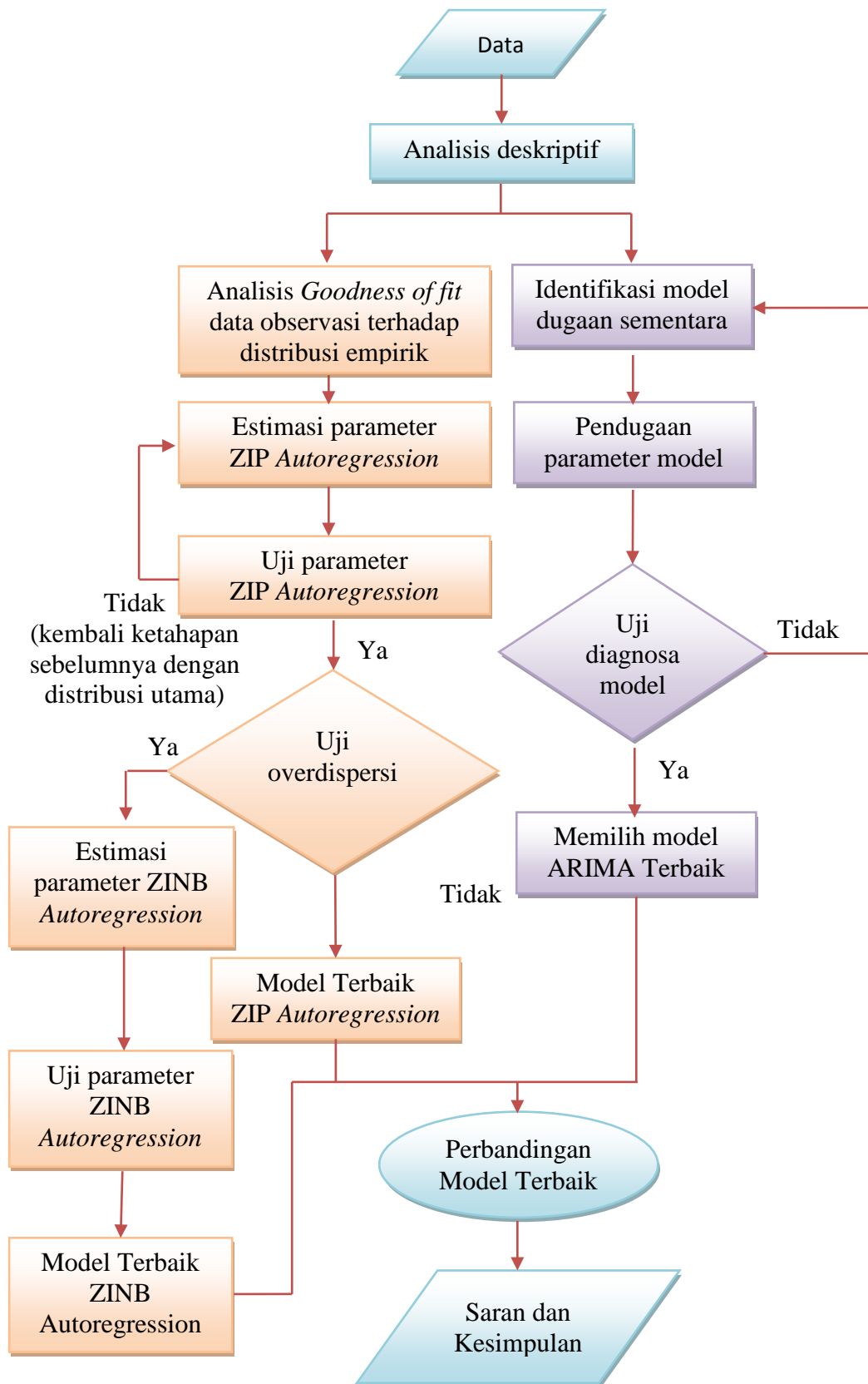
1. Melakukan analisis deskriptif terhadap data jumlah hari hujan per bulan di Kabupaten Indramayu.
2. Membandingkan *Goodness of fit* dari probabilitas observasi sampel terhadap probabilitas distribusi empirik.
3. Dengan menggunakan model ZIM, maka:
 - a. Melakukan estimasi parameter ZIP *autoregression*.
 - b. Melakukan uji kesesuaian model dan signifikansi parameter ZIP *autoregression* secara individu untuk melihat kelayakan model.
 - c. Melakukan uji overdispersi.
 - d. Menentukan model ZIP *autoregression* jika tidak terjadi overdispersi.
 - e. Jika terjadi kasus overdispersi, dilakukan estimasi parameter ZINB *autoregression*.
 - f. Melakukan uji kesesuaian model dan signifikansi parameter ZINB *autoregression* secara individu untuk melihat kelayakan model.
 - g. Menentukan model ZINB *autoregression*.

- h. Berdasarkan model terbaik dilakukan peramalan terhadap jumlah hari hujan per bulan.
4. Dengan menggunakan metode ARIMA berdasarkan prosedur Box Jenkins dilakukan tahapan-tahapan:
 - a. Melakukan identifikasi bentuk model dugaan sementara.

Tahap ini meliputi penghitungan ACF dan PACF, melakukan *differencing* bila diperlukan serta penentuan orde p dan q untuk ARIMA (p,d,q).
 - b. Melakukan pendugaan parameter model.

Tahap ini meliputi pendugaan nilai parameter.
 - c. Melakukan uji diagnosa model ARIMA (p,d,q).

Tahap ini meliputi uji asumsi signifikansi parameter, residual bersifat *white noise* dan uji asumsi residual berdistribusi normal.
 - d. Jika semua asumsi telah terpenuhi maka dilakukan peramalan terhadap jumlah hari hujan per bulan berdasarkan metode terbaik yang diperoleh.
 - e. Jika asumsi tidak terpenuhi maka kembali ketahapan a.
5. Melakukan perbandingan antara model terbaik yang dihasilkan model ZIM dengan model terbaik yang dihasilkan metode ARIMA dengan melihat nilai MSE dan MAPE terkecil yang dihasilkan dari data *out-sample*.
6. Menghitung angka ramalan jumlah hari hujan per bulan berdasarkan model terbaik yang dihasilkan.
7. Mengambil kesimpulan dari hasil penelitian dan memberikan saran yang dapat membantu peneliti selanjutnya yang menekuni bidang yang sama seperti penelitian ini.



Gambar 3.1. Diagram Alur Metode Analisis (Hasil olah)

“Halaman ini sengaja dikosongkan”

BAB 4

HASIL DAN PEMBAHASAN

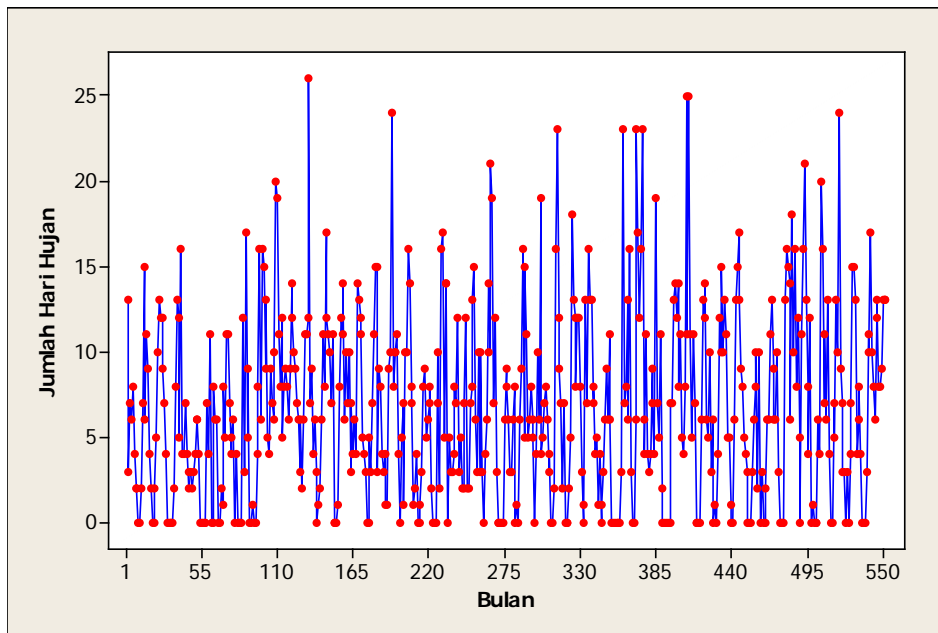
4.1. Gambaran Umum Kejadian Hujan di Kabupaten Indramayu

Data jumlah hari hujan per bulan di Kabupaten Indramayu pada periode 1965 – 2010 terdiri dari 552 data dimana jumlah hari hujan perbulan terkecil adalah sebanyak 0 hari sedangkan jumlah hari hujan perbulan terbanyak adalah 26 hari. Nilai varian yang cukup besar yang terlihat pada Tabel 4.1 menunjukkan bahwa pola data jumlah hari hujan per bulan cukup berfluktuatif antar bulannya. Gambar 4.1 memberikan gambaran yang lebih jelas mengenai fluktuasi data antar waktu pada periode pengamatan.

Tabel 4.1. Statistika Deskriptif Data Jumlah Hari Hujan per Bulan

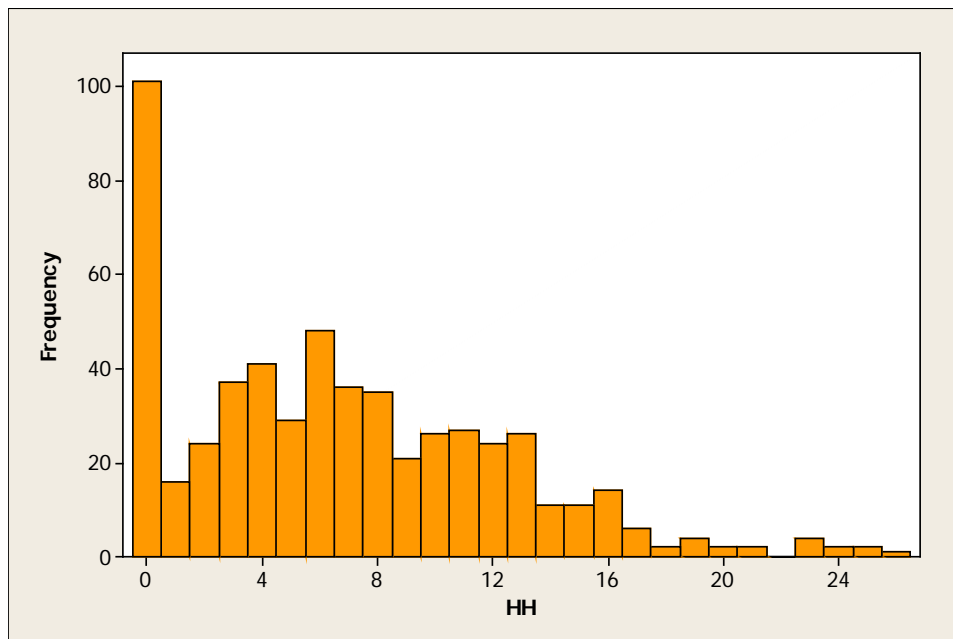
Variabel	Total Observasi	Mean	Varian	Minimum	Maksimum
Jumlah Hari Hujan per Bulan	552	6,759	30,263	0	26

Sumber: Hasil olah Minitab



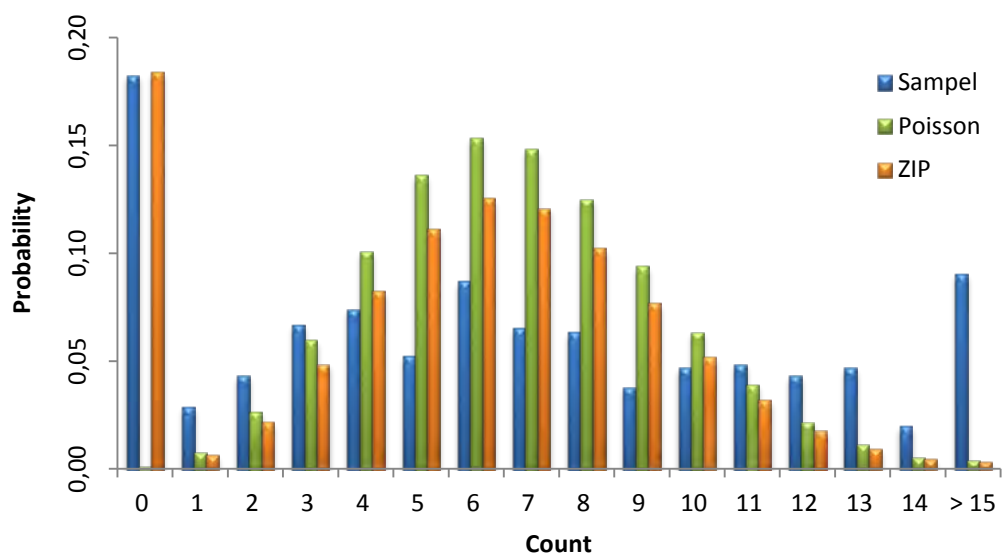
Gambar 4.1. Plot Time Series Data Jumlah Hari Hujan per Bulan di Kabupaten Indramayu Periode 1965 – 2010 (Hasil olah Minitab)

Gambar 4.2 memperlihatkan bahwa jumlah kejadian hujan per bulan sebanyak 0 hari pada periode pengamatan terdapat sebanyak 101 bulan atau 18,30 persen dari total data series yang tersedia. Dari 101 bulan pengamatan yang bernilai 0 tersebut, sebanyak 8,9 persen terjadi pada bulan Juni, sebanyak 21,8 persen terjadi pada bulan Juli, sebanyak 22,8 persen terjadi pada bulan Agustus, sebanyak 25,7 persen terjadi pada bulan September, dan sebanyak 12,9 persen terjadi pada bulan Oktober. Sedangkan sisanya terjadi pada bulan-bulan lainnya dengan persentase kejadian kurang dari 3 persen. Banyaknya *excess zeros* ini sangat mungkin dapat menyebabkan terjadinya kasus overdispersi pada data series.



Gambar 4.2. Banyaknya Jumlah Hari Hujan Per Bulan di Kabupaten Indramayu Periode 1965 – 2010 (Hasil olah Minitab)

Langkah awal yang dilakukan untuk menduga pengaruh *excess zeros* pada data series di atas adalah dengan membuat perbandingan antar nilai probabilitas dari data sampel dibandingkan dengan nilai dari probabilitas yang dihasilkan dari distribusi empirik dalam hal ini digunakan distribusi Poisson dan distribusi ZIP.



Gambar 4.3. *Goodness-of-fit* Jumlah Hari Hujan Per Bulan di Kabupaten Indramayu Periode 1965 – 2010 (Hasil olah Ms. Excel dan Minitab)

Dari Gambar 4.3 terlihat secara umum bahwa nilai probabilitas data yang dihasilkan oleh distribusi Poisson cenderung *over* atau *under estimate* dibandingkan data sampel. Hal yang sama terlihat pada nilai probabilitas yang dihasilkan oleh distribusi ZIP tetapi terlihat lebih baik dibandingkan hasil distribusi Poisson terutama untuk nilai amatan yang bernilai nol. Dari hasil ini dirasakan layak untuk melakukan pemodelan data dengan menggunakan metode ZIP untuk menangani kasus *excess-zeros* pada data ataupun menggunakan pendekatan NB dan ZINB jika nantinya terjadi kasus overdispersi.

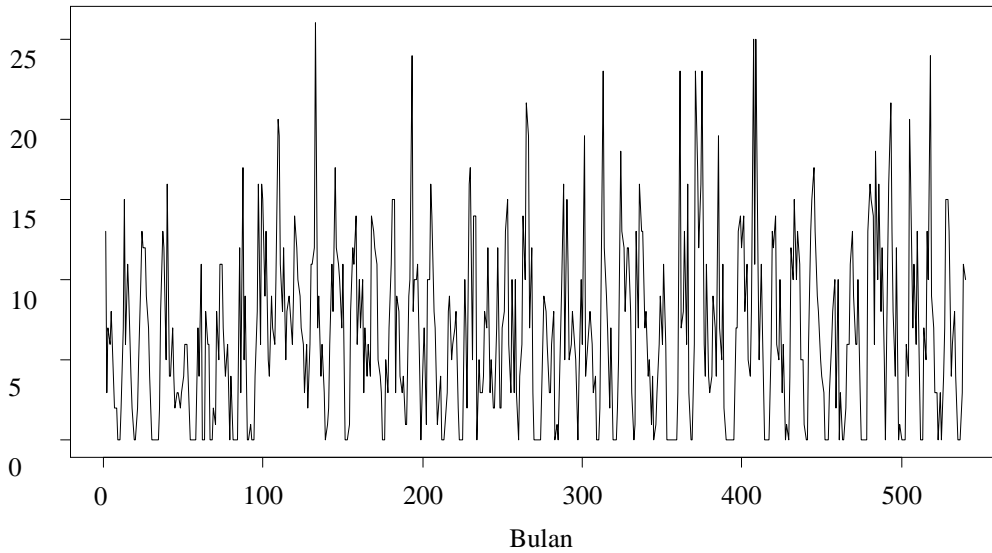
4.2. Model ZIP autoregression

Untuk mencari model terbaik dan membandingkan hasil dari 2 (dua) metode yang berbeda diperlukan pemisahan data *in-sample* dan data *out-sample*. Data *in-sample* akan digunakan sebagai bahan untuk mencari model terbaik pada masing-masing metode sedangkan dari data *out-sample* digunakan untuk melihat model terbaik yang dihasilkan antar dua metode. Data *in-sample* yang digunakan pada penelitian ini yaitu data jumlah hari hujan di Kabupaten Indramayu pada periode Januari 1965 s.d. Desember 2009 yaitu sebanyak 540 data sedangkan

sisanya sebanyak 12 data pada periode januari 2010 s.d. Desember 2010 akan digunakan sebagai data *out-sample*.

4.2.1. Estimasi Parameter ZIP autoregression dan Poisson autoregression Data *in-sample*

Plot data *time series* untuk data *in-sample* dari data jumlah hari hujan per bulan yang terlihat pada Gambar 4.4 tidak menunjukkan hal yang berbeda dari plot data secara keseluruhan dimana terlihat pola data yang sangat bervariasi antar bulannya.



Gambar 4.4. Plot Time Series Data *in-sample* Jumlah Hari Hujan Per Bulan di Kabupaten Indramayu Periode 1965 – 2009 (Hasil olah R)

Estimasi parameter ZIP *autoregression* menggunakan MPLE dengan prosedur Algoritma EM, Newton Rhapsion dan Fisher Scoring. Hasil pengolahan data *in-sample* pada Tabel 4.2 memperlihatkan hasil estimasi dan uji parameter dengan pendekatan model ZIP *autoregression* dan Poisson *autoregression*.

Pada model ZIP *autoregression*, model data diskrit yang terbentuk adalah:

$$\log(\hat{\lambda}_t) = 1,1244 + 0,8544X_tI_{(Y_t=0)} + 0,6463Z_t$$

$$\text{dimana } \hat{\lambda}_t = e^{(1,1244+0,8544X_tI_{(Y_t=0)}+0,6463Z_t)},$$

sedangkan model *zero-inflation* yang terbentuk adalah:

$$\text{logit}(\hat{\omega}_t) = -1,8005 + 0,89973Z_t$$

$$\text{dimana } \hat{\omega}_t = \frac{e^{(-1,8005+0,89973Z_t)}}{1+e^{(-1,8005+0,89973Z_t)}}.$$

Untuk model Poisson *autoregression* model data diskrit yang terbentuk adalah:

$$\log(\hat{\lambda}_t) = 0,4701 + 1,4367X_tI_{(Y_t=0)} + 0,5120Z_t$$

$$\text{dimana } \hat{\lambda}_t = e^{(0,4701+1,4367X_tI_{(Y_t=0)}+0,5120Z_t)}.$$

Tabel 4.2. Estimasi dan Uji Parameter Individual Model ZIP *autoregression* dan Poisson *autoregression* Pada Data *in-sample*

θ	ZIP Model (AIC = 3315.6)			$G_{hitung}/$ (<i>p.value</i>)	Poisson Model (AIC = 3843.9)			$G_{hitung}/$ (<i>p.value</i>)
	<i>Estimate</i>	<i>SE</i>	<i>P-Value</i>		<i>Estimate</i>	<i>SE</i>	<i>P-Value</i>	
β_0	1,1244	0,1031	< 2,2e-16		0,4701	0,0798	3,88e-09	
β_1	0,8544	0,0973	< 2,2e-16		1,4367	0,0750	<2e-16	
β_2	0,6463	0,1117	7,28e-09	151,75 (< 2,2e-16)	0,5120	0,1088	1,76e-06	588,73 (< 2,2e-16)
γ_0	-1,8005	0,2594	3,914e-12					
γ_1	0,8997	0,7776	0,2473					

Sumber: Hasil olah R

4.2.2. Uji Parameter ZIP *autoregression* dan Poisson *autoregression* Data *in-sample*

Langkah selanjutnya yang dilakukan setelah melakukan estimasi parameter model adalah melakukan pengujian kesesuaian model dan pengujian signifikansi parameter model secara individu.

4.2.2.1. Pengujian Kesesuaian Model

Pengujian ini dilakukan dengan menggunakan *Likelihood Ratio* (LR)

Test dengan menggunakan hipotesis:

$$H_0 : \beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_i \neq 0 \text{ atau } \gamma_j \neq 0$$

untuk setiap $i = 1,2$ dan $j = 1,2$ dengan statistik uji yang digunakan yaitu

$$G = -2(\log L_0 - \log L_1)$$

$$= -2\{[\log L(\beta_0|y_t, \omega_t) + \log L(\gamma_0|y_t, \omega_t)] - [\log L(\beta|y_t, \omega_t) + \log L(\gamma|y_t, \omega_t)]\}$$

dimana H_0 ditolak pada taraf signifikansi α jika $G_{hitung} > \chi_{\alpha; 2p}^2$.

Dari Tabel 4.2 terlihat G_{hitung} dari kedua model masing-masing untuk model ZIP *autoregression* sebesar $151,75 > \chi^2_{(0,05;10)} = 18.307$ dan sebesar $588.73 > \chi^2_{(0,05;6)} = 12.592$ untuk model Poisson *autoregression*, demikian pula untuk nilai $p\text{-value} < \alpha = 0,05$. Sehingga dapat disimpulkan kedua model layak untuk digunakan.

4.2.2.2. Pengujian Signifikansi Model secara Individu

Pengujian signifikansi parameter model $\log(\lambda_t) = \mathbf{x}_{t-1}^T \boldsymbol{\beta}$ dimana $t = 1, 2, \dots, n$ dilakukan dengan menggunakan hipotesis:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

untuk setiap $i = 1, 2$ dengan statistik uji $W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2$ dimana H_0 ditolak pada taraf signifikansi α jika $W_j > \chi^2_{\alpha;1}$, sedangkan untuk pengujian signifikansi parameter model $\text{logit}(\omega_t) = \mathbf{z}_{t-1}^T \boldsymbol{\gamma}$ dimana $t = 1, 2, \dots, 540$ dilakukan dengan menggunakan hipotesis:

$$H_0 : \gamma_j = 0$$

$$H_1 : \gamma_j \neq 0$$

untuk setiap $j = 1, 2$ dengan statistik uji $W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2$ dimana H_0 ditolak pada taraf signifikansi α jika $W_j > \chi^2_{\alpha;1}$.

Berdasarkan Tabel 4.2 terlihat bahwa pada kedua model ZIP *autoregression* dan Poisson *autoregression* semua parameter $\boldsymbol{\beta}$ (β_0, β_1 , dan β_2) mempunyai nilai $p\text{-value} < \alpha = 0,05$ yang berarti jumlah hari hujan ke- t dipengaruhi oleh jumlah hari hujan ($t-1$) dan adanya pengaruh positif perubahan bulan terhadap kenaikan nilai harapan terjadinya jumlah hari hujan. Sedangkan untuk parameter $\gamma_1 = \text{trend}$ pada model ZIP *autoregression* tidak signifikan yang berarti perubahan bulan tidak menyebabkan perubahan peluang tidak terjadinya hujan pada bulan tersebut. Sehingga secara umum dapat disimpulkan model ZIP *autoregression* kembali ke model Poisson *autoregression* disebabkan tidak signifikannya pengaruh $\text{logit}(\omega_t) = \mathbf{z}_{t-1}^T \boldsymbol{\gamma}$ terhadap model.

4.2.3. Deteksi Overdispersi Pada Data *in-sample*

Setelah dilakukan pengujian terhadap kesesuaian model dan signifikansi model maka dilakukan deteksi terhadap kemungkinan terjadinya kasus overdispersi pada data *series* sehingga dapat ditangani dengan menggunakan pendekatan model NB *autoregression*. Pengujian overdispersi pada model dilakukan dengan menggunakan uji Pearson chi-square dengan hipotesis:

$H_0 : \kappa \leq 1$ (Data tidak overdispersi)

$H_1 : \kappa > 1$ (Data overdispersi)

dimana $\hat{\kappa} = \frac{\chi^2}{n-p-1}$ dan $\chi^2 = \sum_{t=1}^n \frac{(y_t - \hat{\lambda}_t)^2}{var(\hat{\lambda}_t)}$.

Berdasarkan hasil pengolahan didapatkan nilai sebesar $4,552 > 1$ yang berarti terjadi overdispersi pada data amatan sehingga diperlukan alternatif model NB *autoregression* untuk penanganannya.

4.2.4. Estimasi dan Uji Parameter NB *autoregression* Data *in-sample*

Untuk menangani kasus overdispersi yang terjadi pada model maka dilakukan pendekatan dengan menggunakan model NB *autoregression*, Tabel 4.3 memberikan gambaran hasil estimasi dan uji parameter dengan menggunakan pendekatan model NB *autoregression*.

Tabel 4.3. Estimasi Parameter Model Poisson *autoregression* dan *Negatif Binomial autoregression* pada Data *in-sample*

θ	Poisson Model (AIC = 3843.9)			$G_{hitung}/$ (<i>p.value</i>)	NB Model (AIC = 3067.5)			$G_{hitung}/$ (<i>p.value</i>)
	<i>Estimate</i>	<i>SE</i>	<i>P-Value</i>		<i>Estimate</i>	<i>SE</i>	<i>P-Value</i>	
β_0	0,4701	0,0798	3,88e-09	588.73 ($< 2,2e-16$)	0.4558	0.1270	0.0003	134,81 ($< 2,2e-16$)
β_1	1,4367	0,0750	$< 2e-16$		1.4404	0.1129	$< 2e-16$	
β_2	0,5120	0,1088	1,76e-06		0.5597	0.2434	0.0215	

Sumber: Hasil olah R

Dengan menggunakan pendekatan Model NB *autoregression* didapatkan kesimpulan yang tidak berbeda dengan pendekatan model sebelumnya dimana semua parameter β (β_0, β_1 , dan β_2) signifikan. Selain itu model NB *autoregression* menghasilkan penurunan nilai AIC yang cukup besar (sekitar 800)

jika dibandingkan dengan pendekatan Poisson *autoregression*. Dengan hasil ini secara umum dapat disimpulkan bahwa model yang dihasilkan dari pendekatan model NB *autoregression* adalah model terbaik yang dihasilkan dari data *in-sample*.

4.2.5. Pemilihan Model Terbaik ZIM

Model terbaik adalah model yang mempunyai nilai AIC terkecil pada data *in-sample* serta mempunyai nilai MSE terkecil pada data *out-samplynya* dibandingkan dengan model lainnya. Nilai MSE mengindikasikan sejauh mana keakuratan hasil peramalan dari model yang sudah dihasilkan pada data *in-sample* dibandingkan dengan data sebenarnya (data *out-sample*) dilihat dari jumlah kesalahan yang dihasilkan.

Berdasarkan Tabel 4.4 terlihat bahwa model NB *autoregression* mempunyai nilai AIC yang jauh lebih kecil (sekitar 800) dibandingkan dengan model Poisson *autoregression*. Demikian pula untuk nilai MSE, model NB *autoregression* menghasilkan rata-rata kesalahan hasil estimasi yang relatif lebih kecil dibandingkan dengan model Poisson *autoregression*. Sehingga dapat disimpulkan bahwa model NB *autoregression* adalah sebagai model terbaik.

Tabel 4.4. Nilai AIC dan MSE dari Model Poisson *autoregression* dan *Negatif Binomial autoregression*

	Kriteria Kebajikan Model	Poisson <i>autoregression</i>	NB <i>autoregression</i>
<i>in-sample</i>	AIC	3.843,857	3.067,544
<i>out-sample</i>	MSE	12,704	12,334

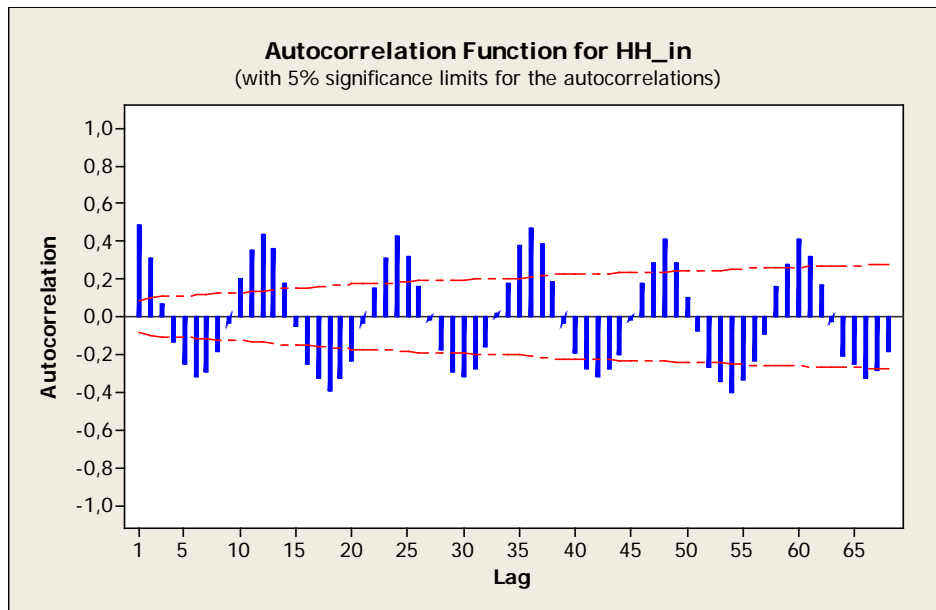
Sumber: Hasil olah R

4.3. Model ARIMA

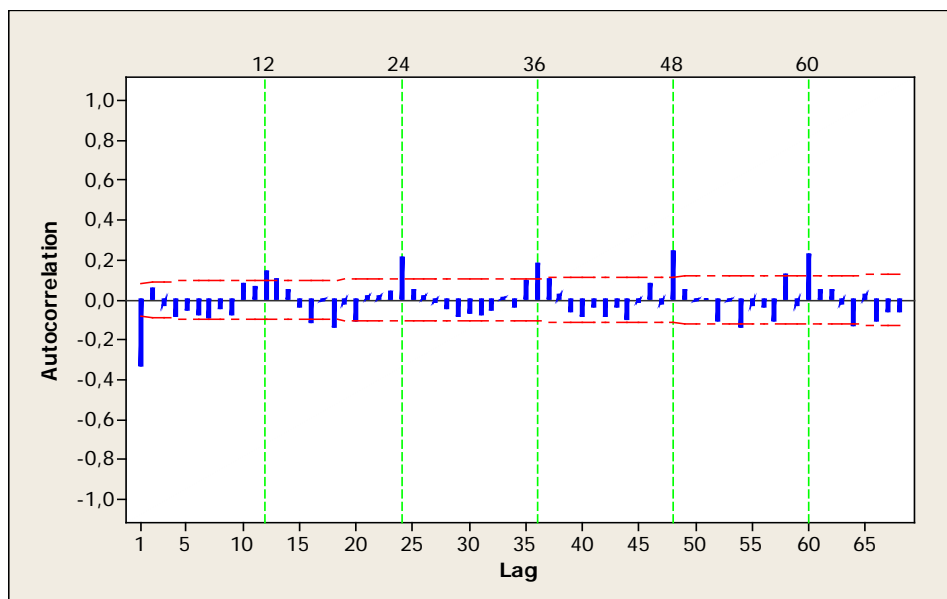
Salah satu metode yang paling umum digunakan untuk memodelkan dan meramalkan data *time series* adalah dengan menggunakan metode ARIMA. Pada penelitian ini model ARIMA digunakan sebagai bahan pembandingan untuk melihat sejauh mana kebaikan model yang dapat dihasilkan dari pendekatan model ZIM.

4.3.1. Identifikasi Model

Berdasarkan Gambar 4.4 dapat dilihat plot *time series* dari data *in-sample* sangat berfluktuatif sehingga secara kasat mata terlihat bahwa data belum stasioner terhadap mean.



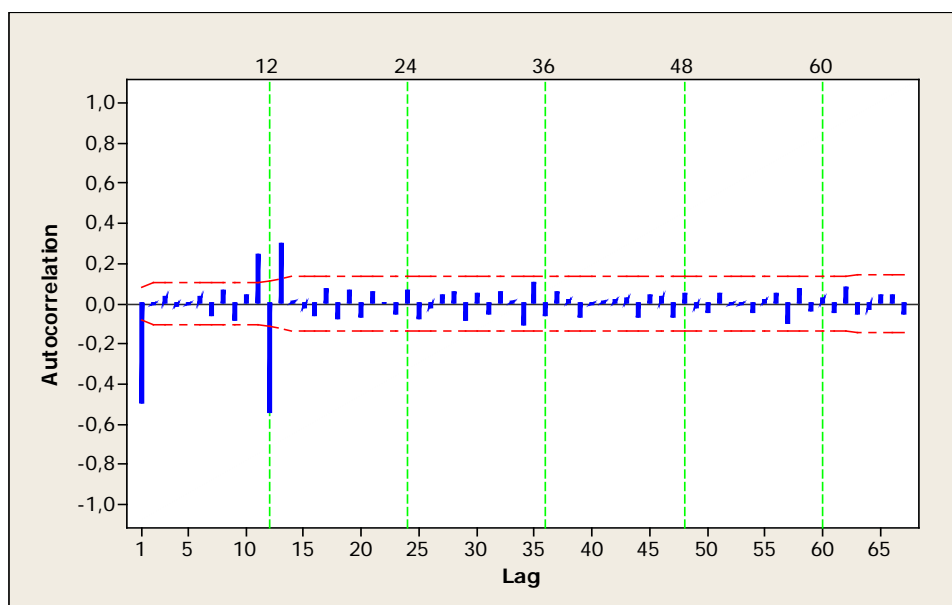
Gambar 4.5. Plot ACF Data *in-sample* Jumlah Hari Hujan per Bulan (Hasil olah Minitab)



Gambar 4.6 Plot ACF Data *in-sample* Jumlah Hari Hujan per Bulan setelah *differencing* 1 reguler (Hasil olah Minitab)

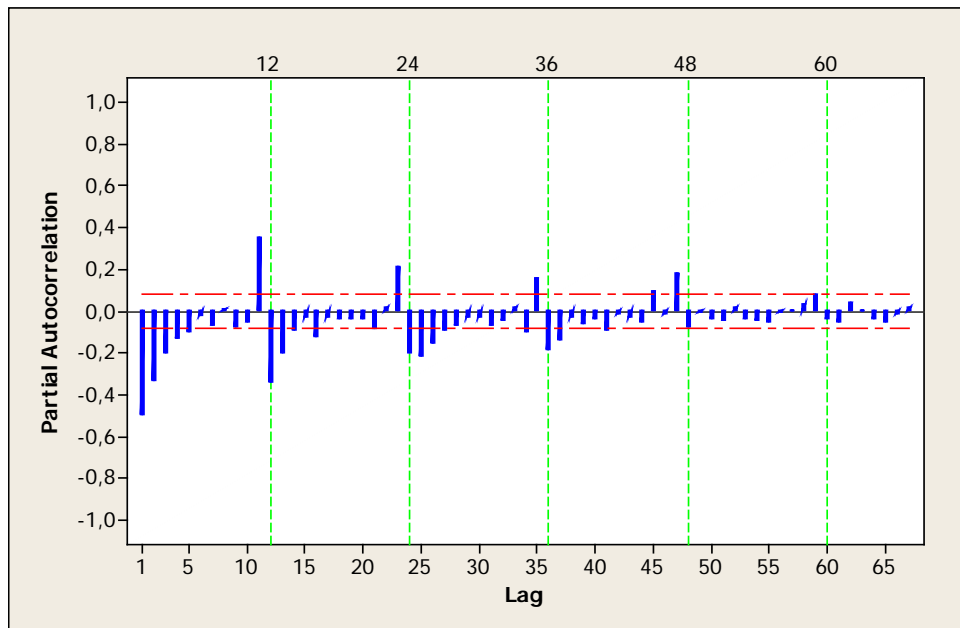
Gambar 4.5 memberikan gambaran yang lebih rinci dari plot ACF yang mendukung dugaan tersebut dimana pola *dies down* ACF yang lambat dan berulang pada periode tertentu yang dapat diduga adanya faktor musiman sehingga diperlukan *differencing* 1 reguler untuk menstasionerkan data terhadap mean kemudian dilihat kembali pola ACF nya untuk memastikan apakah memang terdapat faktor musiman pada data.

Gambar 4.6 memperlihatkan plot data ACF yang sudah di *differencing* 1 reguler dimana terlihat hasilnya pola ACF sudah *dies down* cepat tetapi masih ada lag-lag yang keluar pada periode kelipatan 12 sehingga dugaan adanya faktor musiman menjadi lebih meyakinkan. Oleh karena itu diperlukan *differencing* musiman 12 agar data stasioner terhadap mean dan bebas dari faktor musiman.



Gambar 4.7. Plot ACF Data *in-sample* Jumlah Hari Hujan per Bulan setelah *differencing* 1 reguler dan musiman 12 (Hasil olah Minitab)

Dari Gambar 4.7 terlihat pola data sudah stasioner terhadap mean, sehingga langkah selanjutnya adalah menentukan dugaan awal orde model ARIMA yang memungkinkan. Berdasarkan pada plot ACF dan PACF pada Gambar 4.7 dan 4.8 maka beberapa dugaan model ARIMA yang mungkin adalah $ARIMA(5,1,0)(0,1,1)^{12}$, $ARIMA(0,1,1)(0,1,1)^{12}$, dan $ARIMA(5,1,1)(0,1,1)^{12}$.



Gambar 4.8. Plot PACF Data *in-sample* Jumlah Hari Hujan per Bulan setelah *differencing* 1 reguler dan musiman 12 (Hasil olah Minitab)

4.3.2. Estimasi Parameter dan Signifikansi Parameter Model Dugaan

Langkah berikutnya yang dilakukan setelah menentukan model dugaan adalah melakukan estimasi parameter untuk kemudian dilakukan uji signifikansi parameter dugaan.

Hasil pengolahan pada Tabel 4.5 memperlihatkan bahwa dari 3 model dugaan awal yang diestimasi, model $ARIMA(5,1,0)(0,1,1)^{12}$ dan $ARIMA(0,1,1)(0,1,1)^{12}$ signifikan untuk semua parameter pada $\alpha = 0,05$, sedangkan untuk model $ARIMA(5,1,1)(0,1,1)^{12}$ ada 3 parameter yaitu ϕ_3, ϕ_4, ϕ_5 tidak signifikan sehingga dirasa perlu membuat model dugaan tambahan dengan mengeluarkan parameter yang tidak signifikan dimana hasilnya model $ARIMA(2,1,1)(0,1,1)^{12}$ memberikan hasil yang signifikan untuk semua parameter.

Tabel 4.5. Estimasi dan Signifikansi Parameter Model Dugaan pada Data *in-sample*

Parameter	Estimasi/ <i>p-value</i> Model Dugaan			
	ARIMA (5,1,0)(0,1,1) ¹²	ARIMA (0,1,1)(0,1,1) ¹²	ARIMA (5,1,1)(0,1,1) ¹²	ARIMA (2,1,1)(0,1,1) ¹²
ϕ_1	-0.74586 (<i><.0001</i>)		0.14853 (<i>0.0035</i>)	0.13293 (<i>0.0093</i>)
ϕ_2	-0.50139 (<i><.0001</i>)		0.15506 (<i>0.0016</i>)	0.14739 (<i>0.0029</i>)
ϕ_3	-0.34357 (<i><.0001</i>)		0.08340 (<i>0.0827</i>)	
ϕ_4	-0.21925 (<i><.0001</i>)		0.05331 (<i>0.2585</i>)	
ϕ_5	-0.10961 (<i>0.0122</i>)		0.03111 (<i>0.5075</i>)	
θ_1		0.79052 (<i><.0001</i>)	0.94527 (<i><.0001</i>)	0.91224 (<i><.0001</i>)
Θ_1	0.90242 (<i><.0001</i>)	0.90633 (<i><.0001</i>)	0.89806 (<i><.0001</i>)	0.90462 (<i><.0001</i>)
Kesimpulan	Signifikan	Signifikan	Tidak Signifikan	Signifikan

Sumber: Hasil olah SAS

4.3.3. Uji *White Noise* dan Kenormalan Residual Model

Dari 3 model dugaan yang sudah lolos uji parameter selanjutnya dilakukan uji *white noise* dan uji kenormalan dari residual yang dihasilkan model. Tabel 4.6 memperlihatkan bahwa dua model dugaan yaitu ARIMA(0,1,1)(0,1,1)¹² dan ARIMA(2,1,1)(0,1,1)¹² lolos uji *white noise* tetapi untuk uji kenormalan residual model tidak ada satupun model dugaan yang lolos uji. Oleh sebab itu semua model dugaan dianggap tidak layak untuk digunakan.

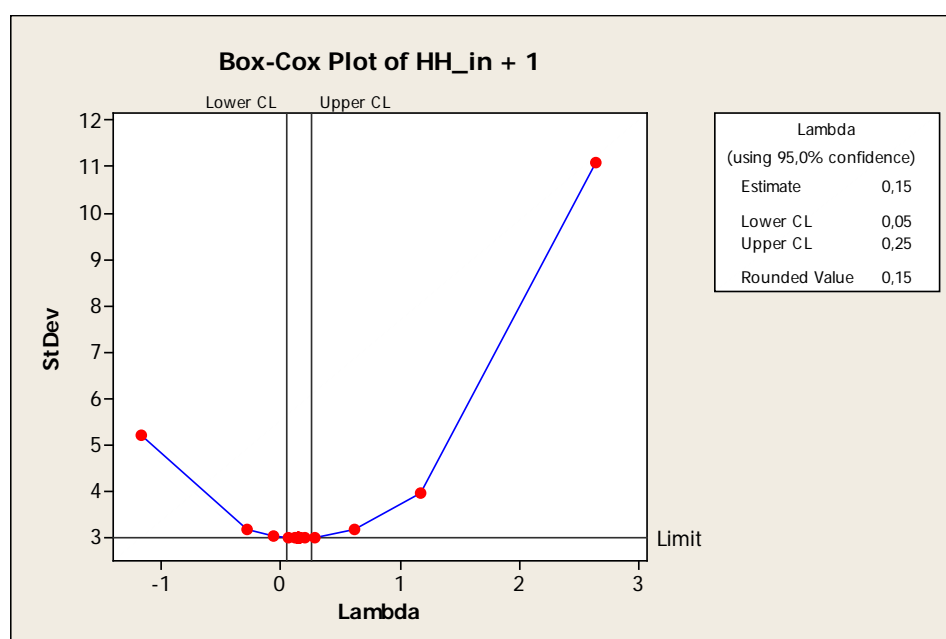
Ketidaknormalan residual model ini sangat mungkin disebabkan oleh tidak stasionernya data terhadap varian dimana terlihat pada deskriptif data variannya cukup besar. Pada tahapan awal model ARIMA sebelumnya tidak dilakukan proses stasioner data terhadap varian dikarenakan metode *Box-Cox Transformation* mensyaratkan setiap data harus bernilai lebih dari nol.

Tabel 4.6. Uji *White Noise* dan Kenormalan Residual Model Dugaan pada Data *in-sample*

Uji <i>White Noise</i> dan Kenormalan Residual	Model Dugaan		
	ARIMA (5,1,0)(0,1,1) ¹²	ARIMA (0,1,1)(0,1,1) ¹²	ARIMA (2,1,1)(0,1,1) ¹²
Lag			
6		0.7585	0.1511
12	0.0115	0.1892	0.1755
18	0.0390	0.2279	0.1924
24	0.1421	0.3714	0.3782
30	0.2316	0.4745	0.4929
36	0.2154	0.2967	0.2624
42	0.2394	0.3080	0.2468
48	0.3485	0.4354	0.3660
Test Normalitas			
Shapiro-Wilk	<0.0001	<0.0001	<0.0001
Kolmogorov-Smirnov	<0.0100	<0.0100	<0.0100
Kesimpulan	Tidak <i>White Noise</i>/Normal	<i>White Noise</i>/Normal	<i>White Noise</i>/Normal

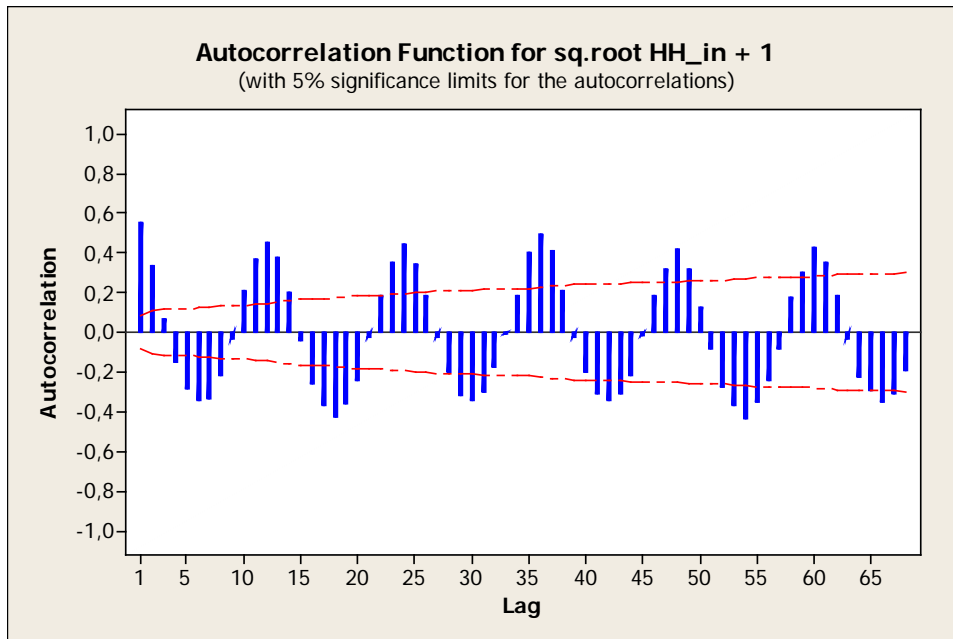
Sumber: Hasil olah SAS

Untuk mengatasi permasalahan ini maka dilakukan proses identifikasi ulang terhadap data *series* dengan cara menambahkan nilai 1 pada setiap data sehingga dimungkinkan untuk melakukan proses stasionerisasi data terhadap varian dengan menggunakan metode *Box-Cox Transformation*.



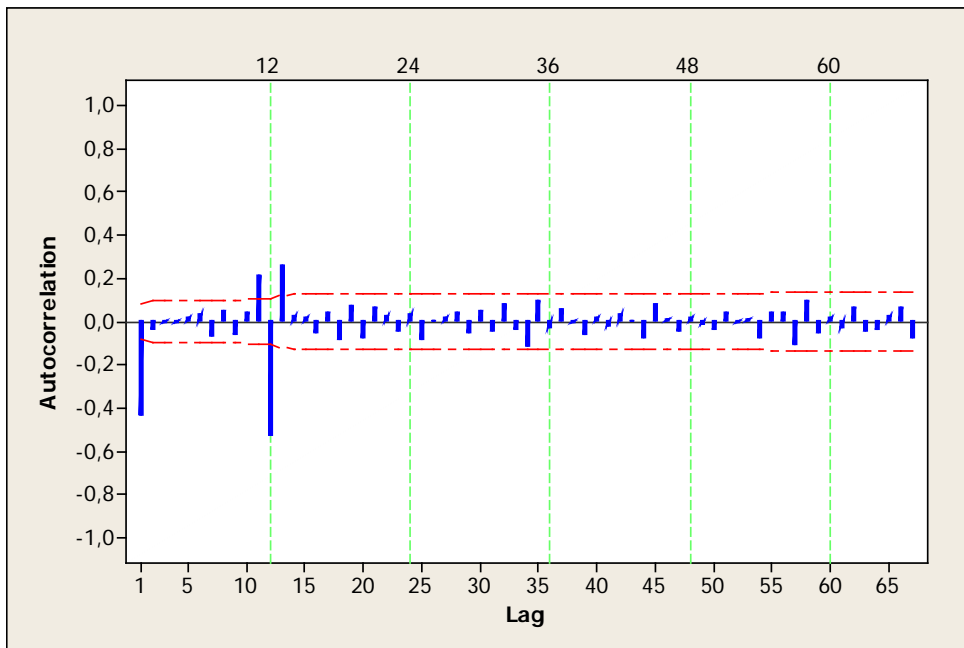
Gambar 4.9. Box-Cox Plot Data *in-sample*+ 1 (Hasil olah Minitab)

Dari plot data Box-Cox pada Gambar 4.9 terlihat bahwa nilai *rounded values* dari Lamba sebesar 0,15 yang berarti data belum stasioner terhadap varian dikarenakan tidak bernilai 1 atau interval tidak melalui nilai 1 sehingga diperlukan transformasi data. Untuk memudahkan penghitungan maka data (Y_t) akan ditransformasi menjadi ($\dot{Y}_t = \sqrt{Y_t} + 1$).

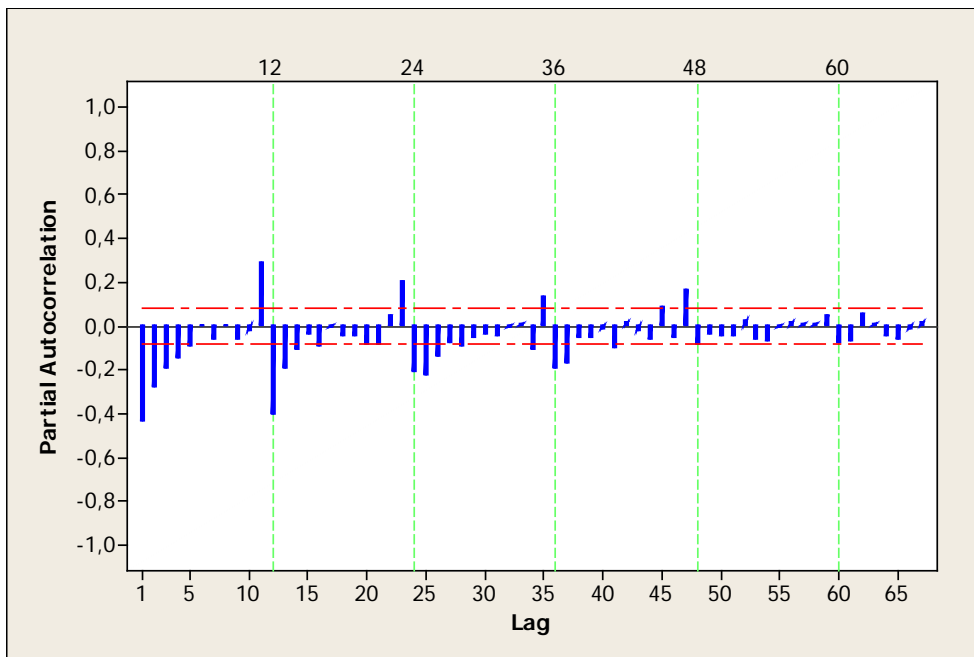


Gambar 4.10. Plot ACF Data *in-sample* transformasi (Hasil olah Minitab)

Gambar 4.10 menunjukkan plot ACF dari data yang sudah ditransformasi dimana seperti sebelumnya terlihat pola data belum stasioner terhadap mean dan terlihat adanya pola musiman sehingga diperlukan *differencing* reguler 1 dan *differencing* musiman 12.



Gambar 4.11. Plot ACF Data *in-sample* transformasi setelah *differencing* 1reguler dan musiman 12 (Hasil olah Minitab)



Gambar 4.12. Plot PACF Data *in-sample* transformasi setelah *differencing* 1 reguler dan musiman 12 (Hasil olah Minitab)

Berdasarkan pada plot ACF dan PACF pada Gambar 4.11 dan 4.12 maka beberapa dugaan model ARIMA yang mungkin adalah $ARIMA(5,1,0)(0,1,1)^{12}$, $ARIMA(0,1,1)(0,1,1)^{12}$, dan $ARIMA(5,1,1)(0,1,1)^{12}$.

Tabel 4.7. Estimasi dan Signifikansi Parameter Model Dugaan pada Data *in-sample* transformasi

Parameter	Estimasi/ <i>p-value</i> Model Dugaan			
	ARIMA (5,1,0)(0,1,1) ¹²	ARIMA (0,1,1)(0,1,1) ¹²	ARIMA (5,1,1)(0,1,1) ¹²	ARIMA (2,1,1)(0,1,1) ¹²
ϕ_1	-0.65091 (<i><.0001</i>)		0.25203 (<i><.0001</i>)	0.25418 (<i><.0001</i>)
ϕ_2	-0.43764 (<i><.0001</i>)		0.13819 (<i>0.0034</i>)	0.15300 (<i>0.0011</i>)
ϕ_3	-0.32036 (<i><.0001</i>)		0.05374 (<i>0.2520</i>)	
ϕ_4	-0.22115 (<i><.0001</i>)		0.03738 (<i>0.4215</i>)	
ϕ_5	-0.10843 (<i>0.0132</i>)		0.04432 (<i>0.3335</i>)	
θ_1		0.73699 (<i><.0001</i>)	0.96450 (<i><.0001</i>)	0.95176 (<i><.0001</i>)
Θ_1	0.89277 (<i><.0001</i>)	0.89671 (<i><.0001</i>)	0.88991 (<i><.0001</i>)	0.89448 (<i><.0001</i>)
Kesimpulan	Signifikan	Signifikan	Tidak Signifikan	Signifikan

Sumber: Hasil olah SAS

Dari tabel 4.7 terlihat bahwa dari 3 model dugaan awal yang diestimasi, model ARIMA(5,1,0)(0,1,1)¹² dan ARIMA(0,1,1)(0,1,1)¹² signifikan untuk semua parameter pada $\alpha = 0,05$, sedangkan untuk model ARIMA(5,1,1)(0,1,1)¹² ada 3 parameter yaitu ϕ_3, ϕ_4, ϕ_5 tidak signifikan sehingga dirasa perlu membuat model dugaan tambahan dengan mengeluarkan parameter yang tidak signifikan dimana hasilnya model ARIMA(2,1,1)(0,1,1)¹² memberikan hasil yang signifikan untuk semua parameter.

Dari 3 model dugaan yang sudah lolos uji parameter selanjutnya dilakukan uji *white noise* dan uji kenormalan dari residual yang dihasilkan model. Tabel 4.8 memperlihatkan bahwa hanya ada satu model dugaan yaitu model ARIMA(2,1,1)(0,1,1)¹² yang lolos uji *white noise* sekaligus uji kenormalan residual sehingga model ini merupakan satu-satunya model yang layak digunakan sekaligus sebagai model terbaik yang dapat digunakan dalam pemodelan ARIMA ini.

Tabel 4.8. Uji *White Noise* dan Kenormalan Residual Model Dugaan pada Data *in-sample* transformasi

Uji <i>White Noise</i> dan Kenormalan Residual	Model Dugaan		
	ARIMA (5,1,0)(0,1,1) ¹²	ARIMA (0,1,1)(0,1,1) ¹²	ARIMA (2,1,1)(0,1,1) ¹²
Lag			
6		0.4194	0.1095
12	0.0124	0.0431	0.2736
18	0.0421	0.0740	0.3318
24	0.1484	0.1922	0.5938
30	0.2060	0.2531	0.6896
36	0.1239	0.1335	0.3693
42	0.1540	0.1318	0.3544
48	0.2879	0.2599	0.5283
Test Normalitas			
Shapiro-Wilk	0.3928	0.4997	0.0752
Kolmogorov-Smirnov	0.0119	0.1051	0.0535
Kesimpulan	Tidak <i>White Noise</i>/Normal	Tidak <i>White Noise</i>/ Normal	<i>White Noise</i>/ Normal

Sumber: Hasil olah SAS

Adapun model ARIMA yang terbentuk yaitu:

$$\begin{aligned} \dot{Y}_t = & \dot{Y}_{t-1} + \dot{Y}_{t-12} - \dot{Y}_{t-13} + \phi_1 \dot{Y}_{t-1} - \phi_1 \dot{Y}_{t-2} + \phi_1 \dot{Y}_{t-13} + \phi_2 \dot{Y}_{t-2} - \phi_2 \dot{Y}_{t-3} \\ & - \phi_2 \dot{Y}_{t-14} + \phi_2 \dot{Y}_{t-15} - \theta_1 a_{t-1} - \theta_1 a_{t-12} + \theta_1 \theta_1 a_{t-13} + a_t \end{aligned}$$

dimana $\dot{Y}_t = \sqrt{\bar{Y}_t} + 1$.

4.4. Perbandingan Model Terbaik Antara Model ZIM dan Model ARIMA

Untuk melihat perbandingan kebaikan model antara model ZIM dengan model ARIMA adalah dengan melihat nilai MSE dari hasil ramalan kedua model dibandingkan dengan data *out-sample* tahun 2010. Tabel 4.9 memperlihatkan perbandingan hasil ramalan dari metode ARIMA dengan metode ZIM untuk data *out-sample*.

Tabel 4.9. Nilai Ramalan Data *out-sample* Metode ARIMA dan ZIM

Observasi	Jumlah Hari Hujan	\hat{Y}_t ($\hat{Y}_t = (\hat{Y}_t)^2 - 1$)	$\hat{\lambda}_t$
541	17	15,018	9,016
542	10	12,074	9,021
543	8	7,015	9,026
544	6	8,403	9,031
545	12	5,765	9,036
546	13	4,693	9,041
547	8	1,461	9,047
548	8	1,081	9,052
549	9	0,564	9,057
550	13	2,798	9,062
551	13	7,880	9,067
552	13	10,547	9,072

Sumber: Hasil olah SAS

Dari hasil penghitungan kedua model yang terlihat pada Tabel 4.10 terlihat bahwa hasil ramalan dari model NB *autoregression* memberikan nilai MSE dan yang jauh lebih kecil dibandingkan dengan nilai ramalan yang dihasilkan oleh metode ARIMA. Dengan demikian dapat disimpulkan bahwa bahwa model NB *autoregression* $\hat{\lambda}_t = e^{(0,4558+1,4404X_tI_{(Y_t=0)}+0,5597Z_t)}$ yang dihasilkan dengan metode ZIM lebih baik digunakan untuk pemodelan pada data diskrit yang mempunyai *excess zeros* dibandingkan dengan pemodelan dengan menggunakan model ARIMA.

Tabel 4.10. Perbandingan Nilai MSE dari Model ZIM dan ARIMA

Kriteria Keباikan Model	ARIMA(2,1,1)(0,1,1) ¹²	NB <i>autoregression</i>
MSE	35,0800	12,3341
MAPE	49,9424	24,4059

Sumber: Hasil olah SAS

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Pemodelan jumlah hari hujan per bulan di Kabupaten Indramayu dengan menggunakan pendekatan ZIM menghasilkan model terbaik yaitu model NB *autoregression* dengan

$$\log(\hat{\lambda}_t) = 0,4558 + 1,4404X_tI_{(Y_t=0)} + 0,5597Z_t$$

dimana

$$\hat{\lambda}_t = e^{(0,4558+1,4404X_tI_{(Y_t=0)}+0,5597Z_t)}.$$

2. Pemodelan dengan pendekatan metode ARIMA menghasilkan model terbaik ARIMA(2,1,1)(0,1,1)¹² dengan persamaan model

$$\begin{aligned} \hat{Y}_t = & \dot{Y}_{t-1} + \dot{Y}_{t-12} - \dot{Y}_{t-13} + 0,254\dot{Y}_{t-1} - 0,254\dot{Y}_{t-2} + 0,254\dot{Y}_{t-13} \\ & + 0,153\dot{Y}_{t-2} - 0,153\dot{Y}_{t-3} - 0,153\dot{Y}_{t-14} + 0,153\dot{Y}_{t-15} - 0,952a_{t-1} \\ & - 0,894a_{t-12} + 0,851a_{t-13} \end{aligned}$$

dimana $\dot{Y}_t = \sqrt{\bar{Y}_t} + 1$

3. Berdasarkan hasil penghitungan, model NB *autoregression* mempunyai nilai MSE yang lebih kecil dibandingkan dengan model ARIMA(2,1,1)(0,1,1)¹² sehingga disimpulkan bahwa pemodelan dengan pendekatan ZIM pada data count dengan *excess zero* lebih baik dibandingkan dengan pemodelan dengan pendekatan ARIMA.

Model Final yang dihasilkan untuk NB *autoregression* adalah:

$$\log(\hat{\lambda}_t) = 0,4409 + 1,4467X_tI_{(Y_t=0)} + 0,6104Z_t$$

dimana

$$\hat{\lambda}_t = e^{(0,4409+1,4467X_tI_{(Y_t=0)}+0,6104Z_t)}.$$

5.2. Saran

Beberapa saran yang dapat diambil berdasarkan hasil yang telah diperoleh adalah sebagai berikut:

1. Pada penelitian ini parameter *zero inflation* (ω_t) tidak signifikan diduga karena persentase *excess zero* masih di bawah 20 persen (terlalu sedikit) sehingga untuk peneliti selanjutnya dapat menggunakan data yang mempunyai persentase *excess zero* yang lebih besar.
2. Penelitian jumlah hari hujan pada penelitian ini menggunakan periode bulanan, untuk peneliti selanjutnya dapat menggunakan jumlah hari hujan dasaharian sehingga analisis dapat menjadi lebih terperinci/detail.

DAFTAR PUSTAKA

- Ariawan, B., Suparti, & Sudarno. (2012). Pemodelan *Zero-Inflated Negatif Binomial* Untuk Data Respon Diskrit dengan *Excess Zeros*. *Jurnal Gaussian, Volume 1, Nomor 1, Tahun 2012*, Halaman 55-64.
- Box, G. E. P., Jenkins, G. M. (1970). *Time Series Analysis, Forecasting, and Control*. Oakland, CA: Holden-Day.
- Byers, A. L., Allore, H., Gill, T. M., Pedduzi, P. N. (2003). Application of negative binomial modeling for discrete outcomes: A case study in aging research. *Journal of Clinical Epidemiology, Volume 56*, Pages 559–564.
- Davis, R. A., & Wu, R. (2009). A negative binomial model for time series of count. *Biometrika, 96*, pp.735-749.
- Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression *Department of Economics Discussion Paper University of Copenhagen*, No. 08-35.
- Lambert, D. (1992). ZIP with an application to defects in manufacturing. *Journal of Tekhnometrics, 34*, 1-14.
- Linden, A., U., Mantyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology Society of Amerika, 92(7)*, pp. 1414-1421. Diunduh melalui alamat <http://www.esajournals.org/doi/pdf/10.1890/10-1831.1>, pada tanggal 18 Desember 2013.
- Narayanan, P., Basistha, A., Sarkar, S., Sachdeva, K. (2013). Trend Analysis and ARIMA modelling of pre-monsoon rainfall data for western India. *C. R. Geoscience 345*, 22-27.
- Suhartono. (2009). *Analisis Data Statistik dengan R*. Graha Ilmu
- Surmaini, E., Boer, R., Siregar, H. (2006). Pemanfaatan Informasi Iklim untuk Menunjang Usahatani Tanaman Pangan. *Jurnal Tanah dan Iklim, No. 26*, 2006.
- Wei, W. W. S. (2006). *Time Series Analysis Univariate and Multivariate Methods*. New York: Addison-Wesley Company Inc.
- Weesakul, U., Lowanichchai, S. (2005). Rainfall Forecasting for Agricultural Water Allocation Planning in Thailand. *Thammasat Int. J. Sc., Vol. 10, No. 3, Juli-September 2005*. Diunduh melalui alamat

http://tijst.net/issues/2005/no3/2005_V10_No3_3.PDF, pada tanggal 18 Desember 2013.

- Wulandari, S. P., Ulama, B. S. S., & Rahmawati, I. (2010). Pemodelan Resiko Penyakit Kaki Gajah (Filariasis) di Provinsi Papua dengan Regresi *Zero-Inflated Poisson*. *Forum Statistik dan Komputasi*, April 2010, p.8-16.
- Yang, M. (2012). *Statistical models for count time series with excess zeros*. Theses and Dissertations. University of Iowa.
- Yau, K. K. W., Lee, A. H., & Carrivick, P. J. W. (2010). Modeling zero-inflated count series with application to occupational health. *Computer Methods and Programs in Biomedicine*, 74, 47-52.
- Zafakali, N. S., Ahmad, W. M. A. W. (2013). Modeling and Handling Overdispersion Health Science Data with Zero-Inflated Poisson Model. *Journal of Modern Applied Statistical Methods*, Vol. 12, No. 1, 255-260. Diunduh melalui alamat <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1027&context=jmasm>, pada tanggal 18 Desember 2013.

BIODATA PENULIS



Penulis lahir di Banjarmasin, 10 Nopember 1981 dengan nama lengkap Frits Fahridws Damanik. Penulis merupakan anak kedua dari pasangan Mahmud Damanik (Alm) dan Aminah Sumiati Saragih. Penulis telah menempuh pendidikan formal di SDN Mawar 2 Banjarmasin (1987-1993), SMP Marsudi Wiyata Banjarmasin (1993-1996), SMAN 7 Banjarmasin (1996-1998), SMAN 13 Banjarmasin (1998-1999), Jurusan Statistika Ekonomi di Sekolah Tinggi Ilmu Statistik (STIS) Jakarta (1999-2003), dan saat ini penulis telah menyelesaikan pendidikan program Magister di Jurusan Statistika ITS Surabaya. Bagi pembaca yang ingin berdiskusi ataupun memberikan saran dan kritik mengenai tesis ini dapat disampaikan melalui email frits.damanik@bps.go.id.

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Jumlah Hari Hujan per Bulan di Kabupaten Indramayu Tahun 1965 – 2010	49
Lampiran 2. Statistika Deskriptif Data Jumlah Hari Hujan per Bulan	50
Lampiran 3. Hasil Output Model ZIP <i>autoregression</i> pada Data <i>in-sample</i>	51
Lampiran 4. Hasil Output Model Poisson <i>autoregression</i> pada Data <i>in-sample</i>	52
Lampiran 5. Hasil Output Model ZINB <i>autoregression</i> pada Data <i>in-sample</i>	53
Lampiran 6. Hasil Output Model ARIMA (5,1,0)(0,1,1) ¹² pada Data <i>in-sample</i>	54
Lampiran 7. Hasil Output Model ARIMA (0,1,1)(0,1,1) ¹² pada Data <i>in-sample</i>	55
Lampiran 8. Hasil Output Model ARIMA (5,1,1)(0,1,1) ¹² pada Data <i>in-sample</i>	56
Lampiran 9. Hasil Output Model ARIMA (2,1,1)(0,1,1) ¹² pada Data <i>in-sample</i>	57
Lampiran 10. Hasil Output Model ARIMA (5,1,0)(0,1,1) ¹² pada Data <i>in-sample</i> transformasi	58
Lampiran 11. Hasil Output Model ARIMA (0,1,1)(0,1,1) ¹² pada Data <i>in-sample</i> transformasi	59
Lampiran 12. Hasil Output Model ARIMA (5,1,1)(0,1,1) ¹² pada Data <i>in-sample</i> transformasi	60
Lampiran 13. Hasil Output Model ARIMA (2,1,1)(0,1,1) ¹² pada Data <i>in-sample</i> transformasi	61
Lampiran 14. Nilai Ramalan Model Dugaan ARIMA (2,1,1)(0,1,1) ¹² pada Data <i>in-sample</i> transformasi	62

“Halaman ini sengaja dikosongkan”

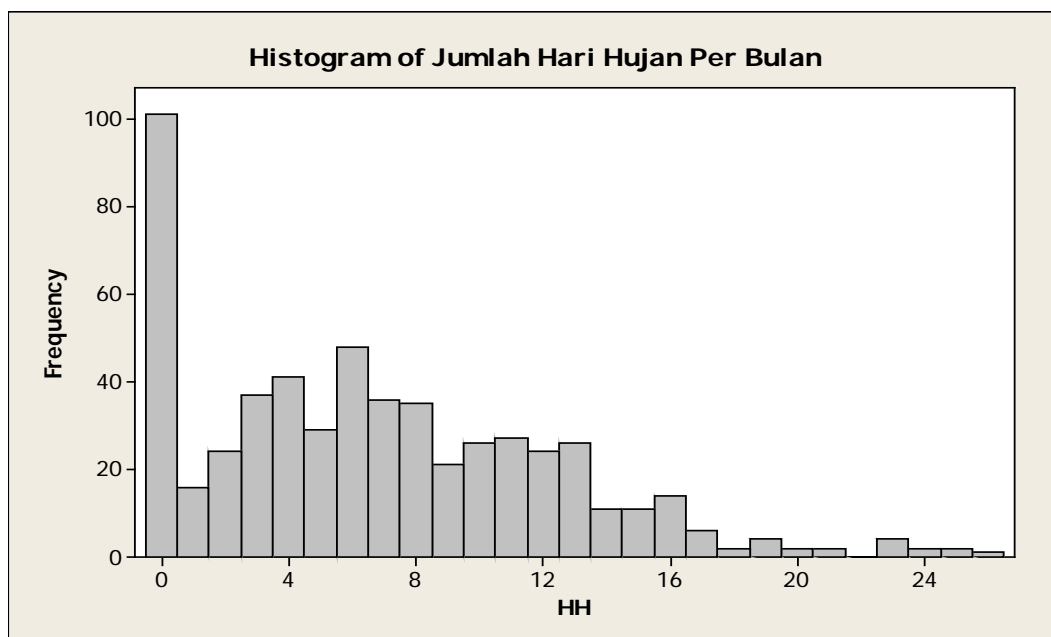
**Lampiran 1. Data Jumlah Hari Hujan per Bulan di Kabupaten Indramayu
Tahun 1965 – 2010**

No.	Tahun	Bulan	Jumlah Hari Hujan per Bulan
1	1965	Januari	13
2	1965	Februari	3
3	1965	Maret	7
4	1965	April	6
5	1965	Mei	8
6	1965	Juni	4
7	1965	Juli	2
8	1965	Agustus	2
9	1965	September	0
10	1965	Oktober	0
11	1965	November	2
12	1965	Desember	7
13	1966	Januari	15
14	1966	Februari	6
15	1966	Maret	11
16	1966	April	9
17	1966	Mei	4
18	1966	Juni	2
19	1966	Juli	0
20	1966	Agustus	0
21	1966	September	2
22	1966	Oktober	5
23	1966	November	10
24	1966	Desember	13
.			
.			
.			
541	2010	Januari	17
542	2010	Februari	10
543	2010	Maret	8
544	2010	April	6
545	2010	Mei	12
546	2010	Juni	13
547	2010	Juli	8
548	2010	Agustus	8
549	2010	September	9
550	2010	Oktober	13
551	2010	November	13
552	2010	Desember	13

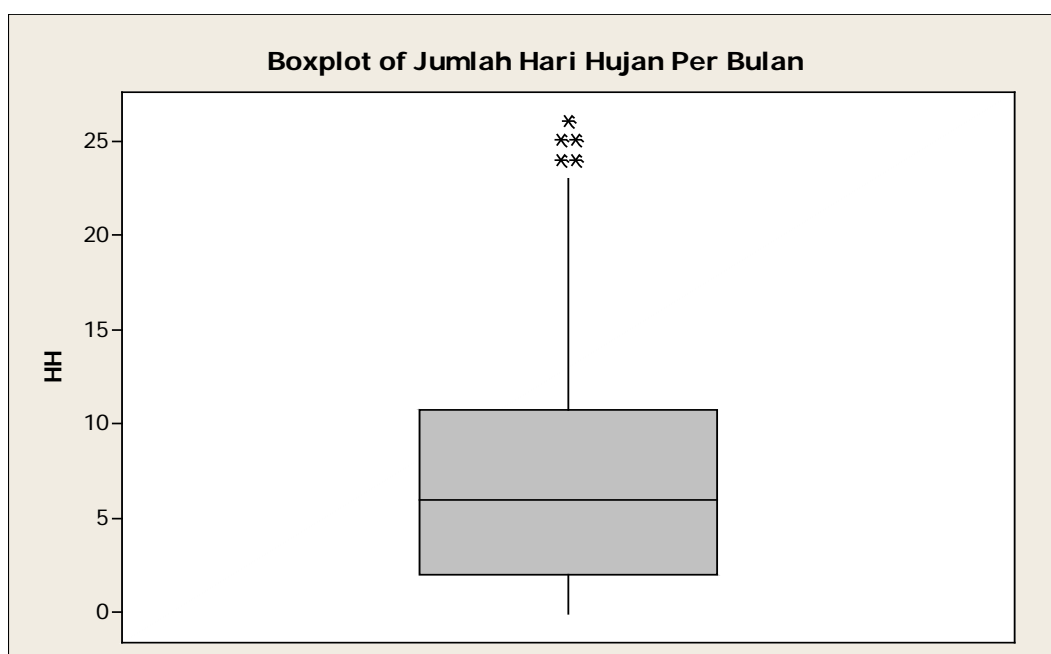
Lampiran 2. Statistika Deskriptif Data Jumlah Hari Hujan per Bulan

Descriptive Statistics: HH

Variable	Total Count	Mean	Variance	Minimum	Maximum
HH	552	6,759	30,263	0,000	26,000



Gambar Histogram Jumlah Hari Hujan Per Bulan



Gambar Boxplot Jumlah Hari Hujan Per Bulan

Lampiran 3. Hasil Output Model ZIP *autoregression* pada Data *in-sample*

zim(formula = count ~ ar1 + trend | trend)

Coefficients (log-linear):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.124413	0.103080	10.9082	< 2.2e-16 ***
ar1	0.854396	0.097321	8.7792	< 2.2e-16 ***
trend	0.646262	0.111726	5.7843	7.28e-09 ***

Coefficients (logistic):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.80051	0.25943	-6.9402	3.914e-12 ***
trend	0.89965	0.77764	1.1569	0.2473

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

Test for overdispersion (H0: ZIP vs. H1: ZINB)

score.test: 26.6767

p.value: < 2.22e-16

Criteria for assessing goodness of fit

loglik: -1652.818

aic: 3315.637

bic: 3337.086

tic: 3328.387

Number of EM-NR iterations: 12

Maximum absolute gradient: 1.646705e-11

Lampiran 4. Hasil Output Model Poisson *autoregression* pada Data *in-sample*

glm(formula = count ~ ar1 + trend, family = poisson)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.2167	-1.8304	-0.3912	1.1686	5.3944

Coefficients :

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.47005	0.07982	5.889	3.88e-09 ***
ar1	1.43673	0.07495	19.170	< 2e-16 ***
trend	0.51995	0.10879	4.779	1.76e-06 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2778.2 on 538 degrees of freedom
Residual deviance: 2189.5 on 536 degrees of freedom
(1 observation deleted due to missingness)
AIC: 3843.9

Number of Fisher Scoring iterations: 6

Lampiran 5. Hasil Output Model ZINB *autoregression* pada Data *in-sample*

zim(formula = count ~ ar1 + trend | trend, dist = "zinb")

Coefficients (log-linear):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.76249	0.13849	5.5056	3.679e-08 ***
ar1	1.18893	0.12762	9.3165	< 2.2e-16 ***
trend	0.71883	0.20038	3.5874	0.000334 ***

Coefficients (logistic):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.34956	0.37457	-6.2727	3.547e-10 ***
trend	1.59527	1.04528	1.5262	0.127

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

(Dispersion parameter for negative binomial taken to be 3.8406)

Criteria for assessing goodness of fit

loglik: -1490.828

aic: 2993.656

bic: 3019.394

tic: 2997.536

Lampiran 6. Hasil Output Model ARIMA (5,1,0)(0,1,1)¹² pada Data in-sample

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.90242	0.01975	45.69	<.0001	12
AR1,1	-0.74586	0.04355	-17.13	<.0001	1
AR1,2	-0.50139	0.05363	-9.35	<.0001	2
AR1,3	-0.34357	0.05597	-6.14	<.0001	3
AR1,4	-0.21925	0.05366	-4.09	<.0001	4
AR1,5	-0.10961	0.04360	-2.51	0.0122	5

Variance Estimate 17.58029

Std Error Estimate 4.192886

AIC 3012.319

SBC 3037.922

Number of Residuals 527

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	.	0	.	-0.005	-0.017	-0.025	-0.044	-0.066	-0.088
12	16.46	6	0.0115	-0.060	-0.016	-0.069	0.020	-0.018	-0.080
18	21.87	12	0.0390	0.051	0.049	-0.027	-0.050	0.011	-0.039
24	24.41	18	0.1421	0.001	-0.014	0.026	-0.031	-0.038	-0.036
30	28.70	24	0.2316	-0.045	-0.018	0.022	0.062	-0.024	0.023
36	35.78	30	0.2154	0.006	0.056	0.042	-0.051	0.063	0.031
42	41.62	36	0.2394	0.066	0.006	-0.067	-0.027	-0.005	0.025
48	44.97	42	0.3485	0.024	0.004	0.037	0.044	-0.037	0.024

Tests for Normality

Test	--Statistic---		-----p Value-----	
Shapiro-Wilk	W	0.985462	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.051361	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.37546	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.075214	Pr > A-Sq	<0.0050

Lampiran 7. Hasil Output Model ARIMA (0,1,1)(0,1,1)¹² pada Data *in-sample*

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.79052	0.02669	29.62	<.0001	1
MA2,1	0.90633	0.01935	46.84	<.0001	12

Variance Estimate 17.04934

Std Error Estimate 4.129085

AIC 2992.188

SBC 3000.723

Number of Residuals 527

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	1.88	4	0.7585	0.015	0.051	0.008	-0.004	-0.016	-0.018
12	13.66	10	0.1892	-0.069	-0.018	-0.082	-0.001	-0.035	-0.094
18	19.83	16	0.2279	0.026	0.023	-0.043	-0.064	-0.011	-0.063
24	23.55	22	0.3714	-0.013	-0.032	0.007	-0.041	-0.046	-0.040
30	27.81	28	0.4745	-0.046	-0.015	0.024	0.057	-0.018	0.034
36	37.88	34	0.2967	0.021	0.071	0.058	-0.03	0.076	0.047
42	43.95	40	0.3080	0.081	0.024	-0.047	-0.011	0.007	0.033
48	46.90	46	0.4354	0.031	0.003	0.028	0.037	-0.041	0.016

Tests for Normality

Test	--Statistic--		-----p Value-----	
Shapiro-Wilk	W	0.981491	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.060658	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.474719	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.664277	Pr > A-Sq	<0.0050

Lampiran 8. Hasil Output Model ARIMA (5,1,1)(0,1,1)¹² pada Data in-sample

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.94527	0.02497	37.85	<.0001	1
MA2,1	0.89806	0.02137	42.03	<.0001	12
AR1,1	0.14853	0.05056	2.94	0.0035	1
AR1,2	0.15506	0.04884	3.17	0.0016	2
AR1,3	0.08340	0.04797	1.74	0.0827	3
AR1,4	0.05331	0.04713	1.13	0.2585	4
AR1,5	0.03111	0.04691	0.66	0.5075	5

Variance Estimate 16.83783
 Std Error Estimate 4.103393
 AIC 2990.566
 SBC 3020.437
 Number of Residuals 527

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	.0	.	-0.001	-0.000	-0.000	0.001	0.003	0.025	
12	6.16	5	0.2914	-0.025	0.020	-0.045	0.033	-0.007	-0.081
18	11.83	11	0.3763	0.051	0.049	-0.028	-0.050	0.005	-0.046
24	13.97	17	0.6691	0.002	-0.014	0.026	-0.027	-0.034	-0.034
30	18.35	23	0.7380	-0.037	-0.007	0.035	0.063	-0.017	0.032
36	27.84	29	0.5264	0.020	0.070	0.053	-0.041	0.072	0.045
42	34.63	35	0.4859	0.084	0.025	-0.048	-0.012	0.012	0.040
48	37.75	41	0.6158	0.034	0.002	0.031	0.037	-0.040	0.017

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.977079	Pr < W <0.0001
Kolmogorov-Smirnov	D 0.079634	Pr > D <0.0100
Cramer-von Mises	W-Sq 0.592261	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq 3.136123	Pr > A-Sq <0.0050

Lampiran 9. Hasil Output Model ARIMA (2,1,1)(0,1,1)¹² pada Data *in-sample*

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.91224	0.02425	37.62	<.0001	1
MA2,1	0.90462	0.01978	45.73	<.0001	12
AR1,1	0.13293	0.05092	2.61	0.0093	1
AR1,2	0.14739	0.04928	2.99	0.0029	2

Variance Estimate 16.90679
 Std Error Estimate 4.111787
 AIC 2989.752
 SBC 3006.821
 Number of Residuals 527

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	Pr > DF	ChiSq	-----Autocorrelations-----						
6	3.78	2	0.1511	-0.010	-0.016	0.060	0.040	0.032	0.022	
12	11.49	8	0.1755	-0.028	0.013	-0.054	0.027	-0.018	-0.097	
18	18.32	14	0.1924	0.035	0.034	-0.051	-0.064	-0.006	-0.059	
24	21.32	20	0.3782	-0.018	-0.031	0.008	-0.038	-0.041	-0.031	
30	25.46	26	0.4929	-0.040	-0.015	0.031	0.058	-0.020	0.030	
36	36.63	32	0.2624	0.022	0.078	0.059	-0.037	0.075	0.052	
42	43.56	38	0.2468	0.086	0.033	-0.041	0.000	0.019	0.040	
48	46.60	44	0.3660	0.030	-0.006	0.027	0.034	-0.047	0.015	

Tests for Normality

Test	--Statistic--		-----p Value-----	
Shapiro-Wilk	W	0.975993	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.070115	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.585734	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.16644	Pr > A-Sq	<0.0050

Lampiran 10. Hasil Output Model ARIMA (5,1,0)(0,1,1)¹² pada Data *in-sample transformasi*

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.89277	0.02039	43.78	<.0001	12
AR1,1	-0.65091	0.04355	-14.94	<.0001	1
AR1,2	-0.43764	0.05118	-8.55	<.0001	2
AR1,3	-0.32036	0.05288	-6.06	<.0001	3
AR1,4	-0.22115	0.05124	-4.32	<.0001	4
AR1,5	-0.10843	0.04362	-2.49	0.0132	5

Variance Estimate 0.589222

Std Error Estimate 0.767608

AIC 1222.769

SBC 1248.373

Number of Residuals 527

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	.0	.	-0.004	-0.013	-0.021	-0.036	-0.054	-0.078	
12	16.27	6	0.0124	-0.072	-0.044	-0.072	-0.008	-0.031	-0.078
18	21.61	12	0.0421	0.029	0.067	0.008	-0.046	-0.013	-0.046
24	24.21	18	0.1484	0.011	-0.024	0.039	-0.010	-0.016	-0.046
30	29.38	24	0.2060	-0.066	-0.004	0.002	0.048	-0.009	0.050
36	39.08	30	0.1239	0.018	0.047	0.001	-0.091	0.068	0.039
42	44.60	36	0.1540	0.074	0.013	-0.054	-0.001	-0.024	0.025
48	46.63	42	0.2879	0.010	-0.008	0.048	0.018	-0.025	0.007

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.996826	Pr < W 0.3928
Kolmogorov-Smirnov	D 0.044601	Pr > D 0.0119
Cramer-von Mises	W-Sq 0.138866	Pr > W-Sq 0.0352
Anderson-Darling	A-Sq 0.693981	Pr > A-Sq 0.0736

Lampiran 11. Hasil Output Model ARIMA (0,1,1)(0,1,1)¹² pada Data in-sample transformasi

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.73699	0.02954	24.95	<.0001	1
MA2,1	0.89671	0.01998	44.87	<.0001	12

Variance Estimate 0.577639
 Std Error Estimate 0.760026
 AIC 1208.336
 SBC 1216.871
 Number of Residuals 527
 * AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	3.90	4	0.4194	0.064	0.015	-0.034	-0.035	-0.016	-0.018
12	18.78	10	0.0431	-0.079	-0.055	-0.085	-0.031	-0.050	-0.088
18	24.77	16	0.0740	0.013	0.051	-0.000	-0.054	-0.033	-0.065
24	27.52	22	0.1922	-0.001	-0.025	0.031	-0.015	-0.025	-0.050
30	32.54	28	0.2531	-0.066	-0.003	0.006	0.041	-0.004	0.054
36	43.22	34	0.1335	0.031	0.055	0.003	-0.080	0.070	0.060
42	50.08	40	0.1318	0.088	0.022	-0.050	-0.005	-0.016	0.031
48	51.73	46	0.2599	0.020	-0.005	0.039	0.012	-0.028	0.001

Tests for Normality

Test	--Statistic--		-----p Value-----	
Shapiro-Wilk	W	0.997164	Pr < W	0.4997
Kolmogorov-Smirnov	D	0.035439	Pr > D	0.1051
Cramer-von Mises	W-Sq	0.09064	Pr > W-Sq	0.1516
Anderson-Darling	A-Sq	0.502135	Pr > A-Sq	0.2139

Lampiran 12. Hasil Output Model ARIMA (5,1,1)(0,1,1)¹² pada Data *in-sample transformasi*

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.96450	0.01694	56.93	<.0001	1
MA2,1	0.88991	0.02185	40.73	<.0001	12
AR1,1	0.25203	0.04712	5.35	<.0001	1
AR1,2	0.13819	0.04700	2.94	0.0034	2
AR1,3	0.05374	0.04686	1.15	0.2520	3
AR1,4	0.03738	0.04646	0.80	0.4215	4
AR1,5	0.04432	0.04579	0.97	0.3335	5

Variance Estimate 0.55808
 Std Error Estimate 0.747048
 AIC 1195.14
 SBC 1225.01
 Number of Residuals 527
 * AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----						
6	0		-0.001	0.001	0.002	0.003	0.005	0.036		
12	5.64	5	0.3429	-0.029	-0.005	-0.045	0.008	-0.018	-0.076	
18	11.37	11	0.4128	0.029	0.063	0.007	-0.047	-0.018	-0.055	
24	13.66	17	0.6908	0.008	-0.021	0.039	-0.010	-0.014	-0.042	
30	18.42	23	0.7345	-0.056	0.006	0.011	0.045	-0.003	0.056	
36	29.32	29	0.4485	0.028	0.059	0.009	-0.080	0.076	0.053	
42	35.84	35	0.4288	0.087	0.025	-0.042	0.004	-0.011	0.037	
48	37.76	41	0.6153	0.021	-0.008	0.040	0.011	-0.032	-0.002	

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.995852	Pr < W 0.1795
Kolmogorov-Smirnov	D 0.034933	Pr > D 0.1183
Cramer-von Mises	W-Sq 0.130752	Pr > W-Sq 0.0445
Anderson-Darling	A-Sq 0.688423	Pr > A-Sq 0.0759

Lampiran 13. Hasil Output Model ARIMA (2,1,1)(0,1,1)¹² pada Data *in-sample transformasi*

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Pr > t	Approx Lag
MA1,1	0.95176	0.01664	57.20	<.0001	1
MA2,1	0.89448	0.02083	42.94	<.0001	12
AR1,1	0.25418	0.04711	5.40	<.0001	1
AR1,2	0.15300	0.04649	3.29	0.0011	2

Variance Estimate 0.559327

Std Error Estimate 0.747882

AIC 1193.348

SBC 1210.417

Number of Residuals 527

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	4.42	2	0.1095	-0.009	-0.024	0.038	0.035	0.054	0.045
12	9.88	8	0.2736	-0.025	-0.003	-0.041	0.012	-0.021	-0.085
18	15.70	14	0.3318	0.016	0.053	-0.008	-0.057	-0.027	-0.060
24	17.90	20	0.5938	-0.001	-0.033	0.025	-0.018	-0.018	-0.040
30	21.98	26	0.6896	-0.055	0.006	0.012	0.039	-0.013	0.049
36	34.05	32	0.3693	0.029	0.067	0.014	-0.079	0.078	0.058
42	40.65	38	0.3544	0.086	0.026	-0.042	0.013	-0.005	0.038
48	42.68	44	0.5283	0.017	-0.015	0.036	0.006	-0.040	-0.005

Tests for Normality

Test	--Statistic--		-----p Value-----	
Shapiro-Wilk	W	0.994851	Pr < W	0.0752
Kolmogorov-Smirnov	D	0.038711	Pr > D	0.0535
Cramer-von Mises	W-Sq	0.145206	Pr > W-Sq	0.0280
Anderson-Darling	A-Sq	0.774775	Pr > A-Sq	0.0451

**Lampiran 14. Nilai Ramalan Model Dugaan ARIMA (2,1,1)(0,1,1)¹² pada
Data *in-sample transformasi***

Model for variable HH

Period(s) of Differencing 1,12

No mean term in this model.

Autoregressive Factors

Factor 1: $1 - 0.13293 B^{**}(1) - 0.14739 B^{**}(2)$

Moving Average Factors

Factor 1: $1 - 0.91224 B^{**}(1)$

Factor 2: $1 - 0.90462 B^{**}(12)$

Forecasts for variable HH

Obs	Forecast	Std Error	95% Confidence Limits	
541	4.0023	0.7479	2.5365	5.4682
542	3.6158	0.7813	2.0844	5.1472
543	2.8311	0.8085	1.2464	4.4158
544	3.0665	0.8179	1.4634	4.6696
545	2.6009	0.8239	0.9860	4.2158
546	2.3861	0.8278	0.7636	4.0086
547	1.5689	0.8309	-0.0597	3.1975
548	1.4426	0.8336	-0.1912	3.0764
549	1.2504	0.8360	-0.3882	2.8890
550	1.9489	0.8384	0.3057	3.5921
551	2.9799	0.8406	1.3323	4.6275
552	3.3981	0.8429	1.7461	5.0501