



TUGAS AKHIR - SS141501

**PERAMALAN KEJADIAN DEMAM BERDARAH DENGUE
DI JAWA TIMUR MENGGUNAKAN METODE ARIMA DAN
INGARCH**

**LINA ZAKIYAH
NRP 1310 100 084**

**Dosen Pembimbing
Dr. Suhartono
Diaz Fitra Aksioma, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018**



FINAL PROJECT - SS141501

FORECASTING ANALYSIS OF DENGUE HEMORRHAGIC FAVER IN EAST JAVA WITH ARIMA AND INGARCH

**LINA ZAKIYAH
NRP 1310 100 084**

**Supervisor
Dr. Suhartono
Diaz Fitra Aksioma, M.Si**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2018**

LEMBAR PENGESAHAN

PERAMALAN KEJADIAN DEMAM BERDARAH DENGUE DI JAWA TIMUR MENGGUNAKAN METODE ARIMA DAN INGARCH

TUGAS AKHIR

Ditujukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada

Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

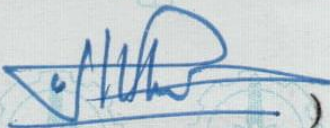
Oleh :

Lina Zakiyah
NRP. 1310 100 084

Disetujui oleh Pembimbing:

Dr. Suhartono

NIP. 19710929 199512 1 001

()



Mengetahui,
Kepala Departemen

Dr. Suhartono

NIP. 19710929 199512 1 001

SURABAYA, AGUSTUS 2018

PERAMALAN KEJADIAN DEMAM BERDARAH DENGUE DI JAWA TIMUR MENGGUNAKAN METODE ARIMA DAN INGARCH

Nama : Lina Zakiyah
NRP : 1310100084
Departemen : Statistika FMKSD-ITS
Dosen Pembimbing : Dr. Suhartono, S.Si., M.Sc.

Abstrak

Fluktuasi dan tingginya jumlah kasus Demam Berdarah Dengue (DBD) di Jawa Timur membutuhkan penanggulangan serius. Penelitian ini bertujuan untuk mendapatkan model terbaik serta nilai peramalan untuk mengetahui jumlah kasus DBD di Jawa Timur pada periode 1 tahun ke depan. Dalam penelitian ini, dilakukan analisis terhadap jumlah kejadian DBD di Jawa Timur per bulan mulai bulan Januari 2007 sampai dengan Desember 2016. Data pada bulan Januari 2007 sampai dengan Desember 2015 digunakan sebagai data in-sample, sedangkan data pada bulan Januari 2016 sampai dengan Desember 2016 sebagai out-sample. Metode yang digunakan untuk pemodelan adalah ARIMA dan INGARCH dengan differencing reguler pertama dan musiman pertama. Model terbaik untuk menggambarkan jumlah kejadian DBD di Jawa Timur adalah model ARIMA dengan nilai MAPE sebesar 14,74% dan RMSE sebesar 338,89. Model INGARCH yang diperoleh adalah model INGARCH-d yaitu model INGARCH dengan differencing pertama dan memuat efek intervensi berdasarkan hasil ARIMA terbaik. Pengembangan dalam penelitian ini adalah melakukan analisis model INGARCH pada data yang tidak stasioner.*

Kata kunci: *Aedes aegypty*, ARIMA, DBD, INGARCH-d*, Intervensi, MAPE, RMSE

FORECASTING ANALYSIS OF DENGUE HEMORRHAGIC FEVER IN EAST JAVA WITH ARIMA AND INGARCH

Name : Lina Zakiyah
Student Number : 1310100084
Department : Statistics
Supervisors : Dr. Suhartono, S.Si., M.Sc.

Abstract

Fluctuations and the high number of cases of Dengue Hemorrhagic Fever (DHF) in East Java require serious mitigation. This study aims to get the best model and forecasting value to find out the number of dengue cases in East Java in the next 1 year period. In this study, the number of dengue cases in East Java per month, from January 2007 to December 2016 was analyzed. Data from January 2007 to December 2015 were used as in-sample data, while data from January 2016 to December 2016 as out-sample. The methods used for modeling are ARIMA and INGARCH with first and seasonal differencing. The best model to describe the number of dengue fever cases in East Java is the ARIMA model with MAPE value is equal to 14.74% and RMSE value is equal to 338.89. The INGARCH model obtained is the INGARCH-d model, which is the INGARCH model with the first differencing and contains the intervention effect based on the best ARIMA results. The development in this study is analyzing the INGARCH model on non-stationary data.*

Keyword: *Aedes aegypti, ARIMA, DBD, INGARCH-d*, Intervention, MAPE, RMSE*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat dan karunia, kesehatan, dan kesempatan yang diberikan kepada penulis. Shalawat dan salam selalu tercurah kepada Nabi Muhammad saw., sehingga penulis dapat menyelesaikan penyusunan laporan Tugas Akhir (TA) yang berjudul

“PERAMALAN KEJADIAN DEMAM BERDARAH DENGUE DI JAWA TIMUR MENGGUNAKAN METODE ARIMA DAN INGARCH”

Dalam proses pelaksanaan tugas akhir hingga tersusunnya laporan ini, penulis telah didukung oleh banyak pihak. Untuk itu, penulis ingin menyampaikan apresiasi dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta, Ibu dan Bapak yang tak pernah berhenti memberikan do'a dan dukungan untuk penulis. Serta seluruh keluarga besar penulis yang selalu memberikan dukungan yang luar biasa. Sungguh kalian adalah harta terindah yang penulis miliki.
2. Bapak Dr. Suhartono selaku Kepala Departemen Statistika sekaligus sebagai dosen pembimbing tugas akhir, atas kesabaran, ketelatenan, waktu, semangat, dan ilmu yang telah diberikan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir dengan baik.
3. Bapak Dr. Sutikno selaku Kepala Program Studi Sarjana Departemen Statistika, atas kesabaran dan ketelatenan dalam membina penulis sehingga menjadi pribadi yang lebih baik.
4. Ibu Diaz Fitra Aksioma, S.Si., M.Si. selaku dosen pembimbing tugas akhir, atas kesabaran, ketelatenan, waktu, semangat, dan ilmu yang telah diberikan dalam diskusi-diskusi yang telah dilalui, sehingga penulis dapat menyelesaikan tugas akhir dengan baik.
5. Bapak Imam Safawi Ahmad, M.Si. dan Ibu Dr. Santi Puteri Rahayu selaku dosen penguji atas masukan dan kritikan yang telah diberikan dalam penyusunan laporan tugas akhir ini.

6. Ibu Santi Wulan Purnami, Ph.D selaku dosen wali penulis yang telah memberikan arahan dan motivasi kepada penulis selama masa perkuliahan.
7. Segenap dosen dan karyawan departemen Statistika ITS, atas ilmu, pengalaman dan dukungan serta kebersamaan yang telah dilalui selama masa kuliah.
8. Sahabat-sahabat masa kuliah dan semua kawan-kawan $\Sigma 21$ yang *Amazing*. Kebersamaan, semangat, canda tawa, dan kekeluargaan yang telah diberikan.
9. Teman-teman LDJ Statistika ITS, UKM Pena 1-0, JMMI ITS, Kogase ITS, dan sahabat-sahabat mentoring atas kepercayaan, kebersamaan dan kekeluargaan yang diberikan.
10. Pihak-pihak lain yang telah mendukung dan membantu terselesaikannya tugas akhir ini yang tidak bisa penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penyusunan laporan tugas akhir ini masih ditemui banyak ketidaksempurnaan. Untuk itu saran dan kritik yang membangun sangatlah diharapkan guna perbaikan, sehingga hasil tugas akhir ini dapat memberikan manfaat yang besar bagi semua pihak.

Surabaya, Agustus 2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xvii
DAFTAR TABEL	xix
DAFTAR LAMPIRAN	xxi

BAB I PENDAHULUAN

1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan.....	5
1.4 Manfaat.....	6
1.5 Batasan Masalah	6

BAB II TINJAUAN PUSATAKA

2.1 Definisi Peramalan (<i>Forecasting</i>).....	7
2.2 Analisis <i>Time Series</i>	8
2.2.1 Stasioneritas Data.....	8
2.2.2 <i>Autocorrelation Function</i> (ACF).....	10
2.2.3 <i>Partial Autocorrelation Function</i> (PACF)	11
2.3 <i>Auto Regressive Integrated Moving Average</i> (ARIMA) .	12
2.3.1 Identifikasi Model	13
2.3.2 Estimasi dan Pengujian Signifikansi Parameter	14
2.3.3 Uji Kesesuaian Model	16
2.3.4 Uji Asumsi Distribusi Residual Berdistribusi Normal.....	16
2.3.5 Deteksi <i>Outlier</i>	17
2.3.6 Kriteria Model Terbaik.....	20

2.4	<i>Generalized Linear Model (GLM) untuk Data Time Series</i>	21
2.4.1	Model Poisson untuk Data Diskrit.....	22
2.4.2	Regresi Binomial Negatif untuk Data Diskrit	23
2.4.3	Model <i>Integer-Valued Generalized Autoregressive Conditional Heteroscedasticity</i> (INGARCH)	24
2.4.4	Estimasi Parameter	25
2.4.5	Pengujian Kecocokan Model.....	28
4.4	Analisis Intervensi	30
4.5	Demam Berdarah Dengue (DBD).....	31
4.5.1	Epidemi Dengue	32
4.5.2	Virus Penyebab Dengue	33
4.5.3	Dengue di Indonesia	33

BAB III METODELOGI PENELITIAN

3.1	Sumber Data dan Variabel Penelitian	35
3.2	Metode Pengolahan Data	35

BAB IV ANALISIS DATA DAN PEMBAHASAN

4.1	Eksplorasi Data Jumlah Penderita Demam Berdarah Dengue di Jawa Timur	39
4.2	Identifikasi Model ARIMA untuk Meramalkan Jumlah Kejadian Demam Berdarah Dengue di Jawa Timur	43
4.2.1	Identifikasi Model	43
4.2.2	Estimasi dan Pengujian Signifikansi Parameter	47
4.2.3	Uji Kesesuaian Model	49
4.2.4	Uji Asumsi Distribusi Residual Berdistribusi Normal.....	50
4.2.5	Pemilihan Model ARIMA Terbaik.....	51
4.3	Identifikasi Model INGARCH untuk Meramalkan Kejadian Demam Berdarah Dengue di Jawa Timur	53
4.3.1	Identifikasi Model	54
4.3.2	Estimasi dan Pengujian Signifikansi Parameter Model.....	54

4.3.3 Pemilihan Model INGARCH Terbaik.....	57
4.4 Analisis Intervensi	61
4.5 Perbandingan Pemodelan ARIMA dan INGARCH	64

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan.....	67
5.2 Saran.....	68

DAFTAR PUSATAKA	69
------------------------------	-----------

LAMPIRAN	72
-----------------------	-----------

BIODATA PENULIS	83
------------------------------	-----------

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

	Halaman
Gambar 3.1 Diagram Alir Model Peramalan ARIMA dan INGARCH	38
Gambar 4.1 <i>Boxplot</i> Data Jumlah Kejadian DBD Berdasarkan Tahun	41
Gambar 4.2 <i>Boxplot</i> Data Jumlah Kejadian DBD Berdasarkan Bulan.....	43
Gambar 4.3 Plot <i>Time Series</i> Jumlah Kejadian DBD di Jawa Timur	44
Gambar 4.4 Plot <i>Box-Cox</i> Jumlah Kejadian DBD di Jawa Timur	45
Gambar 4.5 Plot <i>Time Series</i> Jumlah Kejadian DBD di Jawa Timur Setelah Transformasi	45
Gambar 4.6 Plot ACF dan PACF Jumlah Kejadian DBD di Jawa Timur Sebelum <i>Differencing</i>	46
Gambar 4.7 Plot ACF dan PACF Jumlah Kejadian DBD di Jawa Timur Setelah <i>Differencing</i> pada Lag 1	47
Gambar 4.8 Plot ACF dan PACF Jumlah Kejadian DBD di Jawa Timur Setelah <i>Differencing</i> pada Lag 12	47
Gambar 4.9 Perbandingan Data <i>Out-Sample</i> dengan Hasil Peramalan Ketiga Model ARIMA.....	52
Gambar 4.10 Perbandingan Data <i>Out-Sample</i> dengan Hasil Peramalan Model INGARCH	59
Gambar 4.11 Perbandingan Data <i>Out-Sample</i> dengan Hasil Peramalan Model INGARCH-d dan Model INGARCH-d*	63
Gambar 4.12 Plot <i>Time Series</i> Perbandingan Hasil Peramalan ARIMA, INGARCH, dan Data Asli	64

Halaman ini sengaja dikosongkan

DAFTAR TABEL

	Halaman
Tabel 2.1	Nilai λ dan <i>Power Transformation</i> 10
Tabel 2.2	Struktur ACF dan PACF yang Stasioner 14
Tabel 4.1	Statistika Deskriptif Data Jumlah Kejadian DBD Berdasarkan Tahun..... 39
Tabel 4.2	Statistika Deskriptif Data Jumlah Kejadian DBD Berdasarkan Bulan 41
Tabel 4.3	Uji Signifikansi Model ARIMA 48
Tabel 4.4	MSE Model ARIMA 49
Tabel 4.5	Pemeriksaan Diagnostik <i>White Noise</i> Terhadap Residual pada Model ARIMA 49
Tabel 4.6	Uji Kolmogorov Smirnov 51
Tabel 4.7	Pemilihan Model ARIMA Terbaik Berdasarkan Kriteria Keباikan Model 53
Tabel 4.8	Data <i>Differencing</i> Pertama dan Musiman Pertama 55
Tabel 4.9	Uji Signifikansi Model INGARCH 56
Tabel 4.10	Pemilihan Model INGARCH Terbaik Berdasarkan Kriteria Keباikan Model 60
Tabel 4.11	Deteksi <i>Outlier</i> Residual Model ARIMA Terbaik..... 61
Tabel 4.12	Hasil Estimasi Paramater Model INGARCH-d* 61
Tabel 4.13	Perbandingan Kriteria Keباikan Model untuk <i>Out-sample</i> pada Model ARIMA Terbaik dan INGARCH Terbaik 65
Tabel 4.14	Peramalan <i>Out-sample</i> oleh Model ARIMA dan INGARCH Tahun 2016 66

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 <i>Output Minitab Model ARIMA</i> $(1,1,0)(1,1,0)^{12}$	72
Lampiran 2 <i>Output Minitab Model ARIMA</i> $(1,1,0)(0,1,1)^{12}$	73
Lampiran 3 <i>Output Minitab Model ARIMA</i> $(0,1,1)(0,1,1)^{12}$	74
Lampiran 4 <i>Output Minitab Model ARIMA</i> $(1,1,0)(1,1,0)^{12}$	75
Lampiran 5 <i>Output R Model INGARCH Tanpa</i> <i>Differencing</i>	76
Lampiran 6 <i>Output R Hasil Peramalan Model</i> INGARCH (1,0)	77
Lampiran 7 <i>Output R Hasil Peramalan Model</i> INGARCH ([1,13],0).....	78
Lampiran 8 <i>Output R Hasil Estimasi Parameter</i> <i>dan Peramalan Model INGARCH-d</i> $([1,12,13],0)$ /Model INGARCH <i>dengan Differencing</i>	79
Lampiran 9 <i>Output R Hasil Estimasi Paramater</i> <i>dan Peramalan Model INGARCH-d</i> $([1,12,13],0)^*$ /Model INGARCH <i>dengan Intervensi</i>	80

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

1.6 Latar Belakang

Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan oleh vektor nyamuk *Aedes aegypti*. Penyakit DBD akut sering terjadi di wilayah yang beriklim tropis (Winahju dkk., 2012). Di dalam buku yang dipublikasikan pada tahun 2001 oleh badan International dunia, *World Health Organization* (WHO) menyebutkan bahwa tujuh dari sepuluh negara di Asia Tenggara menghadapi masalah DBD yang serius, bahkan menjadi penyebab utama rawat inap dan kematian kalangan anak-anak pada negara tersebut. Beberapa kasus DBD menelan korban jiwa dalam waktu yang relatif singkat, yaitu tidak lebih dari 12 jam (Satari & Meiliasari, 2004). WHO juga menyatakan bahwa stratifikasi DBD di wilayah Asia Tenggara menunjukkan Indonesia, Myanmar, dan Thailand berada pada kategori A, artinya kasus DBD di 3 negara ini sangat tinggi.

Kasus DBD di Indonesia sebagai salah satu negara beriklim tropis masih menjadi ancaman yang sangat serius dan merupakan masalah kesehatan utama masyarakat. Seiring dengan meningkatnya mobilitas dan kepadatan penduduk, jumlah penderita dan luas daerah penyebarannya semakin bertambah (Soedarto, 2012). Di Indonesia, kasus DBD pertama kali ditemukan di Surabaya pada tahun 1968, dimana sebanyak 58 orang terinfeksi dan 24 orang diantaranya meninggal dunia, dengan Angka Kematian (AK) mencapai 41,3%. Sejak saat itu penyakit ini menyebar luas ke seluruh Indonesia (Kemenkes RI, 2010). Pada tahun 1988 jumlah penderita meningkat drastis mencapai 47 ribu orang dengan angka kematian sebesar 3,2% atau 1527 orang. Sejak pemerintah memberlakukan program DBD mulai tahun 1989, insiden DBD menurun drastis dari 27,1 per 100.000 penduduk pada tahun 1988 menjadi 6,1 per 100.000 penduduk pada tahun 1989. Kemudian antara tahun 1898-1995 terjadi fluktuasi insiden DBD di Indonesia. Insiden DBD kembali

meningkat pada tahun 1996 yaitu sebesar 22,96 per 100.000 penduduk (Soedarto, 2010). Kementerian Kesehatan Republik Indonesia (Kemenkes RI) menyatakan bahwa Indonesia adalah daerah endemis DBD dan mengalami epidemi sekali dalam 4-5 tahun. Berdasarkan publikasi oleh Kemenkes RI (2010) pada tahun 2016, di tahun 2015 tercatat ada sebanyak 126.675 penderita DBD di 34 provinsi di Indonesia, dan 1.229 orang diantaranya meninggal dunia.

Jawa Timur sebagai salah satu provinsi di Indonesia dengan endemi DBD yang tinggi sering menyandang status Kejadian Luar Biasa (KLB) untuk kasus ini. Soedarto (2010) menyatakan bahwa Jawa Timur adalah salah satu provinsi dengan insiden lebih dari 10 per 100.000 penduduk setiap tahunnya. Namun demikian, tren Demam Berdarah di Jawa Timur cenderung fluktuatif. Sebagai contoh, pada tahun 2013 menuju tahun 2014 jumlah kasus DBD menurun dari 14.936 kasus menjadi 8.906 kasus. Sementara pada tahun 2015 jumlah kasus DBD mengalami kenaikan sebanyak dua kali lipat dari tahun sebelumnya. Berdasarkan data Dinas Kesehatan Jawa Timur, tingginya kasus DBD sering terjadi pada musim hujan, terutama Januari. Pada 2014, misalnya dari total 14.936 kasus selama setahun, 3.264 kasus terjadi pada bulan Januari. Kasus terbesar pada Januari 2010, yaitu 5.599 kasus dari total 26.059 kasus sepanjang tahun tersebut. Jumlah kematian penderita DBD pun meningkat pada musim penghujan sekitar Januari hingga Mei (Kompas, 2016). Fluktuasi jumlah kasus DBD di Jawa Timur ini terjadi seiring dengan diberlakukannya program DBD yang dicanangkan pemerintah yaitu program pemberantasan sarang nyamuk (PSN) di berbagai daerah endemis DBD. Hal ini diarahkan dengan melakukan pemberantasan vektor (*Aedes aegypti*) dan kewaspadaan dini terhadap penyakit DBD. Sejak Juni 2015 Kemenkes juga mencanangkan program 1 rumah 1 Jumantik (juru pemantau jentik) yang bertujuan untuk menurunkan angka kematian dan kesakitan akibat Demam Berdarah Dengue.

Gerakan ini merupakan salah satu upaya preventif mencegah DBD dari mulai pintu masuk negara sampai ke pintu rumah.

Berdasarkan paparan di atas, maka kejadian Demam Berdarah tidak bisa dipandang remeh. Harus ada upaya penanggulangan yang serius sehingga jumlah kejadian bisa ditekan. Data kejadian DBD merupakan contoh data *time series*, sebab variabel data berurutan menurut waktu. Data DBD juga memuat variasi musiman, yakni kejadian akan naik atau turun pada periode waktu tertentu. Maka sangat penting untuk meneliti kejadian DBD menggunakan metode yang tepat sesuai dengan karakteristik data yang ada.

Penelitian sebelumnya tentang peramalan kasus DBD di Jawa Timur pernah dilakukan oleh Asrirawan dan Suhartono (2015) dengan metode *Generalized Seasonal Autoregressive Integrated Moving Average* (GSARIMA) dan *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Model GSARIMA menghubungkan komponen antara ARIMA dan variabel prediktor dengan transformasi parameter rata-rata dari distribusi data menggunakan fungsi *link*, dan model ini juga mempertimbangkan stasioneritas dan efek musiman. Hasil penelitian menunjukkan bahwa model GSARIMA lebih baik dari model SARIMA berdasarkan kriteria nilai *Mean Absolute Relative Error* (MARE) yang terkecil. Namun, model GSARIMA ini masih belum mampu mempertimbangkan adanya efek intervensi baik intervensi internal maupun eksternal terhadap perubahan pola data jumlah kasus demam berdarah pada periode waktu tertentu, selain itu model ini juga belum mempertimbangkan jenis data yang akan diteliti apakah termasuk data yang bernilai integer atau tidak.

Data jumlah penderita DBD merupakan data yang bernilai integer atau data diskrit. Model peramalan data *time series* untuk jenis data diskrit telah banyak dikenalkan oleh para ahli statistik, salah satunya adalah model *Integer-Valued Generalized Autoregressive Conditional Heteroscedasticity* (INGARCH) oleh Ferland, Latour, dan Oraichi (2006). Model INGARCH adalah

model yang dikembangkan dari model regresi linear tergeneralisir atau *Generalized Linear Model* (GLM) dengan mengikuti sebaran keluarga eksponensial. Data *time series* untuk jenis data diskrit sangat banyak ditemukan di dalam kehidupan salah satunya dalam dunia kesehatan, khususnya untuk penyakit epidemiologi. Kondisi simpangan baku atau deviasi yang homogen sangat tidak memungkinkan terjadi khususnya dalam situasi-situasi yang penting atau tidak terprediksi. Ferland, dkk. (2006) melakukan peramalan jumlah penderita infeksi *Campylobacter* atau infeksi saluran pencernaan yang dilaporkan sebanyak 13 kali dalam setahun yaitu setiap 28 hari sejak Januari 1990 hingga Oktober 2000 di sebelah utara provinsi Quebec, Kanada. Di dalam penelitian tersebut, Ferland, dkk. mengenalkan model INGARCH dengan mengikuti sebaran Poisson. Estimasi parameter menggunakan metode *Maximum Likelihood Estimator* (MLE), dan hasilnya menunjukkan bahwa model telah sesuai.

Sementara penelitian lain yang menggunakan metode INGARCH pada data *time series* juga dilakukan oleh Liboschik, Fokianos, & Fried (2016) untuk data yang sama yaitu jumlah penderita infeksi *Campylobacter* di sebelah utara provinsi Quebec, Kanada namun dengan menyertakan efek intervensi. Di dalam penelitian tersebut, peneliti memasukkan efek intervensi yang terdeteksi berdasarkan penelitian sebelumnya yang dilakukan oleh Fokianos dan Fried (2010). Mereka juga melakukan perbandingan antara model dengan menggunakan distribusi Poisson dan distribusi Binomial Negatif. Hasilnya adalah model dengan mengikuti sebaran Binomial Negatif lebih sesuai dengan nilai standar deviasi mendekati 0 yaitu sebesar 0,0297.

Berdasarkan paparan yang telah dikemukakan, bahwa jumlah kasus DBD di Jawa Timur merupakan data *time series* untuk jenis data diskrit, maka dalam penelitian ini akan dilakukan peramalan jumlah kejadian DBD di Jawa Timur menggunakan metode INGARCH. Selain itu, berdasarkan penelitian yang dilakukan oleh Liboschik, Fokianos, & Fried (2016), maka di dalam penelitian ini akan disertakan efek intervensi eksternal yaitu data

outlier yang terdeteksi. Peneliti juga akan membandingkan hasil metode INGARCH dengan metode *Autoregressive Integrated Moving Average* (ARIMA) dengan pertimbangan model ini merupakan model umum yang digunakan dalam peramalan data *time series* dengan menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. Namun, model ARIMA belum mampu mendeskripsikan kovariat dan korelasi negatif pada data secara langsung. Sementara, kelebihan model berbasis GLM, selain mampu mempertimbangkan kestasioneritasan dan efek musiman pada data, GLM juga mampu untuk mendeskripsikan kovariat dan korelasi negatif secara langsung (Liboschik, Fokianos, & Fried, 2016).

1.7 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah mengetahui karakteristik data DBD di Jawa Timur selama periode sepuluh tahun mulai Januari 2007 sampai dengan Desember 2016, kemudian akan dilakukan pemodelan dan peramalan pada data dengan membandingkan dua metode yaitu metode ARIMA dan INGARCH untuk data yang tidak stasioner dalam *mean* dan *variance* serta terdapat *outlier* pada data.

1.8 Tujuan

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Mengetahui karakteristik data kejadian Demam Berdarah Dengue di Jawa Timur secara deskriptif.
2. Mengetahui bentuk model ARIMA pada jumlah kejadian Demam Berdarah Dengue di Jawa Timur.
3. Mengetahui bentuk model INGARCH pada jumlah kejadian Demam Berdarah Dengue di Jawa Timur.
4. Mendeteksi *outlier* pada model ARIMA dan menerapkannya sebagai fungsi intervensi pada model INGARCH.

5. Mengetahui model yang terbaik untuk memprediksi kejadian Demam Berdarah Dengue di Jawa Timur.

1.9 Manfaat

1. Mampu memahami karakteristik data kejadian Demam Berdarah Dengue di Jawa Timur.
2. Mampu memahami aplikasi ilmu statistika dalam kehidupan khususnya metode ARIMA dan INGARCH untuk data yang tidak stasioner dalam *mean* dan *variance* serta terdapat *outlier* guna memprediksi jumlah kejadian Demam Berdarah Dengue di Jawa Timur.
3. Hasil penelitian ini diharapkan mampu memberikan manfaat bagi Dinas Kesehatan Jawa Timur, sehingga mereka dapat membuat kebijakan untuk melakukan tindakan preventif guna mengurangi jumlah kasus DBD di Jawa Timur. Selain itu juga dapat membantu pihak kesehatan dalam mensosialisasikan kebijakan tersebut kepada masyarakat.

1.10 Batasan Masalah

Penelitian ini hanya menyertakan analisis intervensi jika terdapat *outlier* pada data.

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka yang ada dalam penelitian ini berisi tentang penjelasan dari metode yang digunakan diantaranya analisis *time series*, model ARIMA, INGARCH, serta uraian tentang penyakit Demam Berdarah Dengue (DBD).

2.5 Definisi Peramalan (*Forecasting*)

Forecasting adalah peramalan (perkiraan) mengenai sesuatu yang belum terjadi (Pangestu, 1986). *Forecasting* atau peramalan adalah memperkirakan sesuatu pada waktu-waktu yang akan datang berdasarkan data masa lampau yang dianalisa secara ilmiah, khususnya menggunakan metode statistika (Supranto, 1984). Ada banyak peristiwa yang dapat diramalkan dalam kehidupan sehari-hari, diantaranya perencanaan penjualan, perencanaan produksi, penganggaran modal, curah hujan, jumlah penderita penyakit malaria, dan lain sebagainya. Untuk melakukan suatu peramalan tentunya dibutuhkan metode yang sesuai dengan data dan informasi yang akan diramalkan sehingga dapat mencapai tujuan yang diinginkan.

Peramalan berdasarkan jangka waktu dibagi menjadi tiga jenis, yaitu peramalan jangka pendek (kurang dari satu tahun, umunya kurang dari tiga bulan), peramalan jangka menengah (tiga bulan hingga satu tahun), dan peramalan jangka panjang (tiga tahun atau lebih). Sedangkan peramalan berdasarkan metode/pendekatan dibagi menjadi dua, yaitu metode peramalan kuantitatif dan metode peramalan kualitatif. Metode kuantitatif adalah metode peramalan menggunakan metode statistika dan matematika yang membutuhkan data masa lalu dalam bentuk numerik. Metode ini dapat dikelompokkan menjadi dua jenis, yaitu model seri waktu/metode deret berskala (*time series*) dan model kausal (*causal/explanatory model*). Sedangkan metode kualitatif umumnya bersifat subjektif, dipengaruhi oleh intuisi,

emosi, dan pengalaman seseorang di masa lalu. Oleh karena itu hasil peramalan satu orang dengan orang lain dapat berbeda.

2.6 Analisis *Time Series*

Data *time series* merupakan serangkaian nilai dari suatu variabel tertentu yang berurutan tiap periodenya. Adanya data *time series* ini dapat dijadikan sebagai dasar dalam melakukan perencanaan kegiatan di masa depan (peramalan). Menurut Wei (2006), *time series* merupakan suatu rangkaian kegiatan dalam melakukan pengamatan terhadap variabel yang akan diamati secara berurutan berdasarkan urutan waktu kejadian dalam interval waktu tertentu secara konstan. Setiap pengamatan yang dilakukan dapat dinyatakan dalam bentuk variabel random Z_t yang didapatkan berdasarkan indeks waktu tertentu t_i dengan $i=1,2,...,n$ sehingga penulisan dari data *time series* adalah $Z_{t_1}, Z_{t_2}, ..., Z_{t_n}$. Ada beberapa hal yang harus diperhatikan dalam melakukan analisis data *time series*, diantaranya stasioneritas data, fungsi autokorelasi (*Autocorrelation Function/ACF*), dan fungsi autokorelasi parsial (*Partial Autocorrelation function/PACF*).

2.6.1 Stasioneritas Data

Stasioneritas data dalam *time series* ditunjukkan apabila rata-rata dan variansnya berfluktuasi konstan setiap waktu. Menurut Cryer & Chan (2008), suatu proses $\{Z_t\}$ dapat dinyatakan *strictly stationary* jika distribusi bersama dari $Z_{t_1}, Z_{t_2}, ..., Z_{t_n}$ dengan $Z_{t_1-k}, Z_{t_2-k}, ..., Z_{t_n-k}$ adalah sama pada setiap waktu $t_1, t_2, ..., t_n$ dan semua waktu pada lag k . Hal ini menyebabkan nilai dari $E(Z_t) = E(Z_{t-k})$ dan $Var(Z_t) = Var(Z_{t-k})$ pada setiap waktu t dan lag k .

Misalkan pada suatu proses *strictly stationary* dengan fungsi distribusi bersama $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ untuk setiap nilai t_1, t_2 dan k mempunyai nilai $\mu_1 = \mu$ dan $\sigma_t^2 = \sigma^2$ pada semua titik waktu t dengan $t_1 = t - k$ dan $t_2 = t$, sehingga didapatkan,

$$\begin{aligned}\gamma(t_1, t_2) &= \gamma(t_1 + k, t_2 + k) = \gamma((t - k) + k, t + k) \\ &= \gamma(t, t + k) = \gamma_k\end{aligned}\quad (2.1)$$

dan

$$\rho(t_1, t_2) = \rho((t - k) + k, t + k) = \rho(t, t + k) = \rho_k \quad (2.2)$$

dari persamaan (2.1) dan (2.2) menunjukkan bahwa suatu proses *strictly stationary* dengan varians yang berhingga memiliki kovarians dan korelasi antara Z_t dan Z_{t+k} hanya bergantung pada *time lag* (Wei, 2006).

Apabila suatu data *series* dalam pengujian stasioneritas varians dinyatakan tidak stasioner, maka dapat diatasi dengan melakukan transformasi. Salah satu transformasi yang dapat digunakan dalam menstasionerkan varians adalah *Power Transformation*. Rumus umum dalam melakukan transformasi dengan *Power Transformation* adalah sebagai berikut (Wei, 2006),

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda}, \quad \text{untuk } \lambda \neq 0 \quad (2.3)$$

dengan λ menunjukkan parameter *Power Transformation*. Jika $\lambda = 0$, maka dapat dilakukan pendekatan sebagai berikut:

$$\lim_{\lambda \rightarrow 0} (Z_t) = \lim_{\lambda \rightarrow 0} \frac{Z_t^\lambda - 1}{\lambda} = \ln(Z_t) \quad (2.4)$$

Pada Tabel 2.1 berikut merupakan beberapa nilai λ yang biasanya digunakan dan transformasi yang sesuai. Data *time series* bersifat stasioner dalam *mean* jika plot *time series* berfluktuasi di sekitar nilai rata-rata yang konstan. Apabila data *time series* dinyatakan tidak stasioner dalam *mean*, maka langkah

selanjutnya adalah melakukan *differencing* dengan rumus pada persamaan (2.5) (Wei, 2006).

Tabel 2.1 Nilai λ dan *Power Transformation*

Nilai Estimasi λ	Transformasi
-1	$1/Z_t$
-0,5	$1/\sqrt{Z_t}$
0	$\ln Z_t$
0,5	$\sqrt{Z_t}$
1	Z_t (tidak ada transformasi)

$$\Delta^d Z_t = (1 - B)^d Z_t. \quad (2.5)$$

2.6.2 Autocorrelation Function (ACF)

Fungsi autokorelasi (ACF) digunakan dalam melakukan identifikasi model data *time series* untuk melihat kestasioneran dan menunjukkan hubungan linear antara Z_t dengan Z_{t+k} . Secara umum, fungsi autokorelasi dirumuskan sebagai berikut (Wei, 2006),

$$\rho_k = \frac{\text{cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (2.6)$$

dan kovarians antara Z_t dengan Z_{t+k} adalah

$$\gamma_k = \text{cov}(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \quad (2.7)$$

Syarat yang harus dipenuhi oleh fungsi autokorelasi dan autokovarians pada proses stasioner diantaranya (Wei, 2006):

1. $\gamma_0 = \text{var}(Z_t)$; $\rho_0 = 1$
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$
3. $\gamma_k = \gamma_{-k}$; $\rho_k = \rho_{-k}$

Menurut Wei (2006), dalam melakukan pengambilan sampel data *time series* untuk fungsi autokorelasi dapat ditulis sebagai berikut,

$$\rho_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2}, k = 0, 1, 2, \dots \quad (2.8)$$

dimana

$$\bar{Z} = \frac{\sum_{t=1}^n Z_t}{n}$$

Untuk n yang besar, ρ_k mendekati distribusi normal dengan *mean* ρ_k (2.8) dan *varians*,

$$\text{var}(\rho_k) = \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k} \rho_{i-k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_k^2 \rho_i^2). \quad (2.9)$$

2.6.3 Partial Autocorrelation Function (PACF)

Fungsi autokorelasi parsial merupakan suatu alat yang digunakan untuk mengukur tingkat keeratan antara Z_t dengan Z_{t+k} yang dirumuskan sebagai berikut (Wei, 2006),

$$P_k = \frac{\text{cov}[(Z_t - \bar{Z}_t), (Z_{t+k} - \bar{Z}_{t+k})]}{\sqrt{\text{var}(Z_t - \bar{Z}_t)} \sqrt{\text{var}(Z_{t+k} - \bar{Z}_{t+k})}} \quad (2.10)$$

Keterangan:

P_k : PACF teoritis

Z_t : nilai variabel *time series* pada waktu ke- t

Z_{t+k} : data yang dipisahkan oleh waktu ke- k , dengan $k = 0, 1, 2, \dots$

$\bar{Z}_t$: dugaan dengan model linear.

Menurut Tsay (2010), PACF pada suatu *time series* digunakan untuk menentukan orde p model AR (*Autoregressive*) dengan syarat data telah stasioner. Perhitungan nilai sampel PACF (ϕ_{kk}) dapat dilakukan dengan mendistribusikan nilai ρ_j ke dalam persamaan berikut ini dengan diawali nilai $\phi_{11} = \rho_1$.

$$\phi_{k+1} = \frac{\rho_{k+1} - \sum_{j=1}^k \phi_{kj} \rho_{k+1-j}}{1 - \sum_{j=1}^k \phi_{kj} \rho_j}. \quad (2.11)$$

2.7 *Autoregressive Integrated Moving Average (ARIMA)*

ARIMA merupakan model peramalan yang biasanya sangat baik digunakan untuk melakukan peramalan jangka pendek. Metode ini sering disebut sebagai metode runtun waktu Box-Jenkins. Selain itu, model ARIMA juga merupakan salah satu model yang digunakan dalam peramalan data *time series* yang bersifat tidak stasioner. Model ini dapat menghasilkan peramalan jangka pendek yang akurat karena menggunakan data masa lalu dan sekarang dari variabel dependen. Secara umum model *ARIMA*(p, d, q) atau yang disebut sebagai *ARIMA non seasonal* ditulis sebagai berikut (Wei, 2006),

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B) a_t \quad (2.12)$$

Adapun model *multiplicative seasonal ARIMA* secara matematis ditulis sebagai berikut (Wei, 2006),

$$\Phi_p(B^s) \phi_p(B)(1-B)^d (1-B^s)^D \dot{Z}_t = \theta_q(B) \Theta_Q(B^s) a_t \quad (2.13)$$

dimana $\dot{Z}_t = Z_t - \mu$

Keterangan:

$\phi_p(B)$: koefisien komponen AR orde p

- $\theta_q(B)$: koefisien komponen MA orde q
 $\Phi_p(B^s)$: koefisien komponen AR periode musiman s orde P
 $\Theta_Q(B^s)$: koefisien komponen MA periode musiman s orde Q
 $(1-B)^d$: *differencing* orde d
 $(1-B^s)^D$: *differencing* musiman s dengan orde D
 B : operator *backshift*
 a_t : barisan *white noise* ($a_t \sim WN(0, \sigma^2)$).

2.7.1 Identifikasi Model

Hal yang harus dilakukan sebelum melakukan peramalan yaitu identifikasi model ARIMA untuk menentukan kestasioneran data dalam varians atau mean dan identifikasi pola data yang digunakan sebagai dasar pemilihan model yang sesuai. Petunjuk pemilihan model ARIMA ditunjukkan dalam Tabel 2.2 (Wei, 2006).

Tabel 2.2 Struktur ACF dan PACF yang Stasioner

Model	ACF	PACF
$AR(p)$	<i>Dies down</i>	<i>Cuts off</i> setelah lag ke- p
$MA(q)$	<i>Cuts off</i> setelah lag ke- q	<i>Dies down</i>
$ARMA(p, q)$	<i>Dies down</i>	<i>Dies down</i>

2.7.2 Estimasi dan Pengujian Signifikansi Parameter

Menurut Cryer & Chan (2008), salah satu metode yang dapat digunakan untuk menaksir parameter model adalah *Least squares (Conditional Least Squares)*. Metode *Least Squares* ini dilakukan dengan cara meminimumkan jumlah kuadrat *error*. Misalkan jika dilakukan estimasi parameter pada model $AR(1)$, maka langkah-langkahnya adalah sebagai berikut:

1. Menentukan persamaan model *error* dari $AR(1)$ yang dirumuskan dalam persamaan berikut,

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + a_t \quad (2.14)$$

sehingga dari persamaan (2.14) didapatkan persamaan,

$$a_t = (Z_t - \mu) - \phi(Z_{t-1} - \mu) \quad (2.15)$$

2. Menentukan persamaan jumlah kuadrat *error* yang dirumuskan dalam persamaan berikut,

$$S_c(\phi, \mu) = \sum_{t=2}^n a_t^2 = \sum_{t=2}^n ((Z_t - \mu) - \phi(Z_{t-1} - \mu))^2 \quad (2.16)$$

3. Meminimumkan jumlah kuadrat *error* dengan cara menurunkan $S_c(\phi, \mu)$ terhadap μ dan ϕ , kemudian masing-masing ekuivalen dengan nol seperti persamaan berikut,

$$\frac{\partial S_c}{\partial \mu} = \sum_{t=2}^n 2[(Z_t - \mu) - \phi(Z_{t-1} - \mu)](-1 + \phi) = 0 \quad (2.17)$$

$$\frac{\partial S_c}{\partial \phi} = \sum_{t=2}^n 2[(Z_t - \mu) - \phi(Z_{t-1} - \mu)](Z_{t-1} - \mu) = 0 \quad (2.18)$$

Berdasarkan persamaan (2.17) dan (2.18) didapatkan hasil untuk taksiran μ dan ϕ , masing-masing adalah sebagai berikut,

$$\mu = \bar{Z} \quad (2.19)$$

$$\phi = \frac{\sum_{t=2}^n (Z_t - \bar{Z})(Z_{t-1} - \bar{Z})}{\sum_{t=2}^n (Z_{t-1} - \bar{Z})^2} \quad (2.20)$$

Setelah melakukan estimasi parameter, langkah selanjutnya adalah melakukan uji signifikansi parameter dengan hipotesis sebagai berikut:

$H_0 : \beta_i = 0$ (parameter model tidak sesuai)

$H_1 : \beta_i \neq 0$ (parameter model sesuai)

Statistik uji :

$$t_{hitung} = \frac{\beta_i}{S.E(\beta_i)} \quad (2.21)$$

Daerah kritis:

Tolak H_0 apabila $|t_{hitung}| > t_{\alpha/2, n-n_p}$, artinya parameter telah signifikan dan model dapat digunakan untuk peramalan. Nilai n pada t tabel menunjukkan jumlah data yang efektif dan n_p adalah banyaknya parameter yang ditaksir.

2.7.3 Uji Kesesuaian Model

Uji kesesuaian model ini bertujuan untuk menguji asumsi *white noise* pada residual. Uji yang digunakan adalah uji Ljung-Box-Pierce (LBQ) dengan hipotesis sebagai berikut (Cryer & Chan, 2008):

$H_0 : \rho_1 = \rho_2 = \dots = \rho_K = 0$ (residual bersifat *white noise*)

H_1 : minimal terdapat satu $\rho_k \neq 0$ untuk $k=1, 2, \dots, K$ (residual tidak bersifat *white noise*)

Statistik uji:

$$Q = n(n+2) \sum_{k=1}^K (n-k)^{-1} \rho_k^2 \quad (2.22)$$

Daerah kritis: tolak H_0 jika $Q > \chi_{\alpha, K-m}^2$

Keterangan:

ρ_k = ACF residual pada lag ke- k

K = maksimum lag

m = $p + q$

2.7.4 Uji Asumsi Distribusi Residual Berdistribusi Normal

Setelah uji asumsi *white noise* dilakukan, asumsi lain yang harus dipenuhi adalah residual harus berdistribusi normal. Salah satu uji yang digunakan dalam menentukan kenormalan data adalah uji *Kolmogorov Smirnov* dengan hipotesis sebagai berikut (Chakravart, Laha, dan Roy, 1967),

$$H_0 : F(x) = F_0(x) \text{ (residual berdistribusi normal)}$$

$$H_1 : F(x) \neq F_0(x) \text{ (residual tidak berdistribusi normal)}$$

Statistik uji:

$$D_{hitung} = \sup |F_n(x) - F_0(x)| \quad (2.23)$$

Daerah kritis: tolak H_0 jika $D_{hitung} > D_{(n,1-\alpha)}$

Keterangan:

$F_n(x)$ = fungsi peluang kumulatif yang dihitung berdasarkan data sampel

$F_0(x)$ = fungsi peluang kumulatif dari distribusi normal

sup = nilai maksimum dari $|F_n(x) - F_0(x)|$.

2.7.5 Deteksi Outlier

Pada pengamatan *time series* biasanya dipengaruhi oleh suatu peristiwa tertentu, misalnya adanya kebijakan tertentu, adanya perang, krisis ekonomi, perubahan keadaan sosial dan politik, perubahan suhu lingkungan yang sangat drastis, kesalahan penulisan pada saat memasukkan data, dan lain-lain. Akibat adanya peristiwa tersebut menyebabkan pengamatan pada saat peristiwa terjadi jauh dari nilai rata-ratanya. Hal ini biasanya disebut sebagai *outlier*. Ada empat jenis *outlier* yaitu *Additive Outlier* (AO), *Innovative Outlier* (IO), *Left Shift* (LS), dan *Temporary Change* (TC). Jenis AO merupakan suatu *outlier* yang hanya dipengaruhi pada saat pengamatan ke- T , sedangkan untuk model IO, LS, dan TC berpengaruh pada pengamatan ke- $T, T+1, T+2$, dan seterusnya. Sedangkan dalam penelitian ini hanya menggunakan deteksi *outlier* dengan model AO dan IO.

Hal ini dipertimbangkan karena metode AO adalah jenis *outlier* yang hanya dipengaruhi oleh pengamatan ke- T saja, sementara untuk metode IO adalah jenis *outlier* yang didasarkan pada asumsi bahwa *outlier* pada data telah diketahui atau diidentifikasi (Wei, 2006).

Misalkan suatu *time series* $X_1, X_2, \dots, X_t$, dengan model $ARMA(p, q)$ atau dapat ditulis sebagai berikut (Wei, 2006),

$$\phi(B)X_t = \theta(B)a_t \quad (2.24)$$

dimana model telah stasioner dan memenuhi asumsi residual *white noise* dan berdistribusi normal, sehingga masing-masing model *outlier* dapat dirumuskan sebagai berikut :

1. Model ARIMA dengan *outlier* AO

$$\begin{aligned} Z_t &= \begin{cases} X_t, & t \neq T \\ X_t + \omega_{AO}I_t^{(T)}, & t = T \end{cases} \\ &= X_t + \omega_{AO}I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)}a_t + \omega_{AO}I_t^{(T)} \end{aligned} \quad (2.25)$$

$$\text{dengan } I_t^{(T)} = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases}$$

2. Model ARIMA dengan *outlier* LS

$$Z_t = X_t + \frac{1}{(1-B)}\omega_{LS}I_t^{(T)} \quad (2.26)$$

3. Model ARIMA dengan *outlier* IO

$$\begin{aligned} Z_t &= X_t + \frac{\theta(B)}{\phi(B)}\omega_{IO}I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)}(a_t + \omega_{IO}I_t^{(T)}) \end{aligned} \quad (2.27)$$

4. Model ARIMA dengan *outlier* TC

$$Z_t = X_t + \frac{1}{(1-\delta B)} \omega_{TC} I_t^{(T)} \quad (2.28)$$

Sedangkan secara umum model ARIMA dengan *outlier* sebanyak k dapat ditulis sebagai berikut (Wei, 2006),

$$Z_t = \sum_{j=1}^k \omega_j v_j(B) I_j^{(T_j)} + \frac{\theta(B)}{\phi(B)} a_t \quad (2.29)$$

Keterangan:

$I_j^{(T_j)}$ = variabel yang menunjukkan adanya *outlier* pada waktu ke- T_j

$v_j(B)$ = 1 untuk AO

$v_j(B) = \frac{\theta(B)}{\phi(B)}$ untuk IO pada saat $t_j = T_j$

$v_j(B) = \frac{1}{(1-B)}$ untuk LS

$v_j(B) = \frac{1}{(1-\delta B)}$; $0 < \delta < 1$ untuk TC

Langkah-langkah dalam melakukan deteksi *outlier* dengan prosedur iteratif menurut Wei (2006) adalah sebagai berikut (misalkan ada dua jenis *outlier* yang dideteksi yaitu AO dan IO) :

1. Melakukan pemodelan data *time series* dengan asumsi tidak ada *outlier*, sehingga didapatkan model residual dan estimasi dari σ_a^2 seperti dalam persamaan 2.30 dan 2.31

$$\hat{e}_t = \pi(B) Z_t = \frac{\phi(B)}{\theta(B)} Z_t \quad (2.30)$$

$$\sigma_a^2 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2 \quad (2.31)$$

Keterangan:

$$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

2. Langkah kedua yaitu menghitung nilai dari $\lambda_{1,t}$ dan $\lambda_{2,t}$ dengan $t=1,2,\dots,n$ menggunakan estimasi model dengan persamaan yang didefinisikan sebagai berikut,

$$\lambda_T = \max_t \max_i \left\{ \left| \lambda_{i,t} \right| \right\} \quad (2.32)$$

dimana T merupakan waktu pada saat nilai maksimum. Jika $\lambda_T = \left| \lambda_{i,t} \right| > C$ dimana C merupakan konstanta positif yang biasanya bernilai antara 3 dan 4 dan AO pada waktu T , maka persamaan model baru, dapat ditulis sebagai berikut,

$$Z_t = Z_t - \omega_{AT}(B) I_t^{(T)} \quad (2.33)$$

dan persamaan residual yang baru adalah

$$\hat{e}_t = e_t - \omega_{AT} \pi(B) I_t^{(T)} \quad (2.34)$$

Jika $\lambda_T = \left| \lambda_{i,t} \right| > C$ dan IO pada waktu T , maka persamaan model baru dapat ditulis sebagai berikut,

$$Z_t = Z_t - \frac{\phi(B)}{\theta(B)} \omega_{IT} I_t^{(T)} \quad (2.35)$$

dan persamaan residual yang baru adalah

$$\hat{e}_t = e_t - \omega_{IT} I_t^{(T)} \quad (2.36)$$

3. Langkah ketiga yaitu melakukan perhitungan pada $\lambda_{1,t}$ dan $\lambda_{2,t}$ berdasarkan hasil residual dan σ_a^2 yang baru dan mengulang langkah kedua secara terus menerus hingga semua *outlier* teridentifikasi.
4. Setelah *outlier* sebanyak k telah teridentifikasi pada waktu $T_1, T_2, \dots, T_k$ dengan estimasi parameter *outlier* adalah

$\omega_1, \omega_2, \dots, \omega_k$, maka langkah berikutnya adalah membentuk model seperti pada persamaan (2.29).

2.7.6 Kriteria Model Terbaik

Untuk melakukan pemilihan model terbaik pada kriteria data *in sample* dan *out sample* dapat menggunakan perhitungan *Root Mean Square Error* (RMSE) yang dirumuskan dengan persamaan sebagai berikut,

$$RMSE = \sqrt{\frac{1}{L} \sum_{l=1}^L \left(Z_{n+l} - Z_n(l) \right)^2} \quad (2.37)$$

Perhitungan RMSE ini merupakan salah satu ukuran yang paling sering digunakan karena ukuran ini memiliki relevansi secara teoritis dalam pemodelan statistik (Hyndman & Koehler, 2006). Selain RMSE, perhitungan kriteria pemilihan model terbaik dari data *in sample* dan *out sample* juga dapat menggunakan *Symmetric Mean Absolute Percentage Error* (SMAPE) dan *Mean Absolute Percentage Error* (MAPE). MAPE merupakan suatu persentase kesalahan rata-rata absolut. Rumus umum dari SMAPE dan MAPE adalah sebagai berikut,

$$SMAPE = \left(\frac{1}{L} \sum_{l=1}^L \frac{|Z_{n+l} - Z_n(l)|}{|Z_{n+l} + Z_n(l)|} \right) \times 100 \quad (2.38)$$

$$MAPE = \left(\frac{1}{L} \sum_{l=1}^L \frac{|Z_{n+l} - Z_n(l)|}{Z_{n+l}} \right) \times 100 \quad (2.39)$$

Keterangan:

Z_{n+l} : data aktual *out sample* ke- l , $l = 1, 2, \dots, L$

$Z_n(l)$: data hasil ramalan *out sample* ke- l .

2.8 Generalized Linear Model (GLM) untuk Data Time Series

Model linear khususnya model regresi sudah mulai digunakan sejak awal abad 19, ditandai dengan kajian-kajian yang

dilakukan oleh Francis Galton (1822-1911) tentang hubungan tinggi badan ayah dan anaknya. Dalam perkembangannya, model regresi linear dengan asumsi peubah respon $Y_i \sim (\mu_i, \sigma^2)$ tidaklah mampu menjawab masalah-masalah yang dihadapi dalam pemodelan statistik. *Generalized Linear Model (GLM)* merupakan pengembangan dari model linear "klasik" khususnya dalam mengatasi kendala peubah respon yang tidak normal. Namun demikian, peubah respon dalam GLM diasumsikan memiliki sebaran yang termasuk dalam keluarga sebaran eksponensial.

Ada tiga komponen utama dalam GLM (McCullagh dan Nelder, 1989), meliputi:

1. Komponen acak, yaitu peubah respon $Y_1, Y_2, \dots, Y_n$ yang merupakan contoh acak dimana $Y_i \sim (\mu_i, \sigma^2)$, dan termasuk dalam keluarga sebaran eksponensial.
2. Komponen sistematis yang merupakan fungsi dari peubah penjelas : $\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_p x_{pi}$.
3. Fungsi hubung (*link function*) yang menghubungkan suatu fungsi dari nilai tengah komponen acak dengan komponen sistematis $g(\mu)_i = \eta_i$.

2.8.1 Model Poisson untuk Data Diskrit

Menurut Cameron dan Trivedi (1998) model regresi Poisson digunakan sebagai pendekatan untuk analisis data cacah. Model regresi Poisson merupakan model regresi *non-linier* yang berasal dari sebaran Poisson. Misalkan nilai y_i dengan $i = 1, 2, \dots, n$, melambangkan jumlah kejadian yang terjadi dalam satu periode dengan nilai parameter dari sebaran Poisson λ_i . Peubah y merupakan peubah acak yang menyebar secara Poisson dengan fungsi massa peluang sebagai berikut,

$$f(y; \lambda_t) = \frac{\exp(-\lambda_t) \lambda_t^y}{y!}, \quad y = 0, 1, 2, \dots$$

dan $E(Y_t) = \text{Var}(Y_t) = \lambda_t$. Dalam proses Poisson, nilai rata-rata bersyarat identik dengan nilai varians dari proses yang diamati. Namun, dalam beberapa kasus hal ini sulit terjadi dimana kondisi varians lebih besar daripada nilai *mean* λ_t , yang mana ini disebut dengan kondisi overdispersi. Model Binomial Negatif memungkinkan untuk mengatasi kondisi overdispersi pada data (Liboschik, Fokianos, dan Fried, 2016).

2.8.2 Regresi Binomial Negatif untuk Data Diskrit

Distribusi binomial negatif memungkinkan untuk menganalisis data dengan nilai varians lebih besar dari *mean* λ_t , yang disebut dengan keadaan overdispersi. Kejadian overdispersi dapat terjadi karena banyaknya jumlah pengamatan yang bernilai nol pada variabel respon. Maka untuk menangani kejadian overdispersi pada regresi Poisson digunakan regresi Binomial Negatif. Misalkan diasumsikan $P(Y_t = y | \mathcal{F}_{t-1}) = \text{NegBin}(\lambda_t, \phi)$ dimana distribusi Binomial Negatif diukur pada nilai *mean* dengan paramater nilai overdispersi tambahan yaitu $\phi \in (0, \infty)$. Maka fungsi kepadatan probabilitas dari distribusi binomial negatif adalah sebagai berikut (Liboschik, Fokianos, dan Fried, 2016),

$$P(Y_t = y | \mathcal{F}_{t-1}) = \frac{\Gamma(\phi + y)}{\Gamma(y+1)\Gamma(\phi)} \left(\frac{\phi}{\phi + \lambda_t} \right)^\phi \left(\frac{\lambda_t}{\phi + \lambda_t} \right)^\lambda, \quad y = 0, 1, \dots \quad (2.44)$$

dimana $\text{Var}(Y_{t-1} = \lambda_t + \lambda_t^2 / \phi)$ adalah varians yang meningkat secara kuadratik dengan λ_t . Distribusi Poisson adalah kasus khusus dari Binomial Negatif ketika $\phi \rightarrow \infty$.

Jika diperhatikan, maka distribusi Binomial Negatif termasuk dalam kelas proses Poisson campuran (*mixed Poisson*).

Proses Poisson campuran ditentukan dengan menetapkan $Y_t = N_t(0, Z_t, \lambda_t)$, dimana $\{N_t\}$ bersifat i.i.d. (*independent and identically distributed*) atau identik dan independen antar pengamatan. Sementara $\{Z_t\}$ adalah variabel acak yang bersifat i.i.d. dengan *mean* 1 dan varians σ^2 . Ketika $\{Z_t\}$ merupakan variabel acak dari proses Gamma, maka kita bisa memperoleh proses Binomial Negatif dengan $\sigma^2 = 1/\phi$ dengan mengacu pada σ^2 sebagai koefisien overdispersi. Dimana $\sigma^2 = 0$ adalah kasus yang sangat jarang terjadi dalam proses Poisson (Liboschik, Fokianos, dan Fried, 2016).

2.8.3 Model *Integer-Valued Generalized Autoregressive Conditional Heteroscedasticity* (INGARCH)

Misalkan sebuah data *time series* ditunjukkan dengan $\{Y_t : t \in \mathbb{N}\}$ dan $\{X_t : t \in \mathbb{N}\}$ adalah vektor kovariat dengan r -dimensi, dan $X_t = (X_{t,1}, X_{t,2}, \dots, X_{t,r})^T$. $E(Y_t | \mathcal{F}_{t-1}) = \lambda_t$ adalah rata-rata bersyarat dari data *time series*, dimana $\{\lambda_t : t \in \mathbb{N}\}$. Maka model umum persamaan yang dibentuk adalah sebagai berikut (Liboschik, Fokianos, & Fried, 2016),

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^p \beta_k g(Y_{t-k}) + \sum_{l=1}^q \alpha_l g(\lambda_{t-j_l}) + \boldsymbol{\eta}^T \mathbf{X}_t \quad (2.40)$$

dimana

$g : \mathbb{R}^+ \rightarrow \mathbb{R}$ adalah fungsi link (*link function*)

$g : \mathbb{N}_0 \rightarrow \mathbb{R}$ adalah fungsi transformasi

Parameter vektor $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_r)^T$ sesuai dengan pengaruh dari kovariat. Di dalam terminologi GLM, $v_t = g(\lambda_t)$ disebut sebagai prediktor linear (Liboschik, Fokianos, & Fried, 2016).

Misalkan g dan g identik sehingga $g(x) = g(x) = x$, selanjutnya $P = \{1, 2, \dots, p\}$, $Q = \{1, 2, \dots, q\}$, dan $\eta = \mathbf{0}$. Maka model pada persamaan (2.40) menjadi sebagai berikut,

$$\lambda_t = \beta_0 + \sum_{k=1}^p \beta_k (Y_{t-k}) + \sum_{l=1}^q \alpha_l \lambda_{t-l} \quad (2.41)$$

Jika diasumsikan bahwa Y_t berdistribusi Poisson, maka diperoleh model *integer-valued* GARCH dengan orde p dan q atau yang disingkat dengan model INGARCH(p, q). Model ini juga dikenal dengan model *Autoregressive Conditional Poisson* (ACP) (Heinen, 2003). Ketika $\eta \neq 0$ dan *mean* bernilai positif, maka model INGARCH dengan kovariat *non-negative* bisa memuat efek intervensi (Liboschik, Fokianos, & Fried, 2016).

Misalkan ditentukan fungsi *link* $g(x) = \log(x)$, $g(x) = \log(x+1)$ dengan $P = \{1, 2, \dots, p\}$ dan $Q = \{1, 2, \dots, q\}$, maka diperoleh model *log-linear* dengan orde p dan q dengan persamaan sebagai berikut (Liboschik, Fokianos, & Fried, 2016),

$$\lambda_t = \beta_0 + \sum_{k=1}^p \beta_k \log(Y_{t-k} + 1) + \sum_{l=1}^q \alpha_l \lambda_{t-l} \quad (2.42)$$

Pada model (2.40) efek kovariat yang ada disebut efek kovariat internal (*internal covariate*) sebab kovariat menyebar ke pengamatan di masa depan baik dengan proses regresi pada pengamatan di masa lalu maupun proses regresi pada nilai rata-rata bersyarat di masa lalu.

2.8.4 Estimasi Parameter

Setelah asumsi Poisson terpenuhi, selanjutnya dilakukan estimasi parameter dengan *Maximum Likelihood Estimator* (MLE) biasa. Sementara untuk proses Poisson campuran (*mixed Poisson*) dilakukan estimasi parameter dengan estimasi *quasi likelihood* (*quasi-MLE*).

Misal $\theta = (\beta_0, \beta_1, \dots, \beta_p, \alpha_1, \dots, \alpha_q, \eta_1, \dots, \eta_r)^T$ adalah vektor parameter regresi. Terlepas dari asumsi distribusi, maka parameter untuk model INGARCH dengan kovariat dinyatakan sebagai berikut,

$$\Theta = \left\{ \theta \in \mathbb{R}^{p+q+r+1} : \beta_0, \beta_1, \dots, \beta_p, \alpha_1, \dots, \alpha_q, \eta_1, \dots, \eta_r \geq 0, \sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j < 1 \right\} \quad (2.45)$$

dimana, $\theta \in \mathbb{R}^+ \rightarrow \mathbb{R}^{p+q+r+1}$ adalah fungsi *link* sesuai dengan jumlah paramater model.

Intersep β_0 harus bernilai positif, sedangkan semua parameter harus bernilai *nonnegatif* untuk memastikan positifitas nilai *mean* λ_t . Selain itu juga untuk memastikan bahwa model fit sudah memenuhi asumsi stasioner dan ergodik pada momen dengan orde berapapun. (Ferland dkk., 2006; Doukhan, Fokianos, dan Tjostheim 2012).

Untuk model log-linear dengan kovariat, maka parameter untuk model ini dinyatakan sebagai berikut,

$$\Theta = \left\{ \theta \in \mathbb{R}^{p+q+r+1} : |\beta_1|, \dots, |\beta_p|, |\alpha_1|, \dots, |\alpha_q| < 1, \left| \sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j \right| < 1 \right\} \quad (2.46)$$

Persamaan log-likelihood untuk Binomial Negatif adalah sebagai berikut,

$$\begin{aligned} \ln L(\phi; \beta) = & \sum_{i=1}^n \ln \left[\frac{\Gamma(y_i + \phi_i)}{\Gamma(\phi_i)} \right] - \sum_{i=1}^n \ln(y_i!) + \sum_{i=1}^n y_i \ln(\mu_i \phi_i) \\ & - \sum_{i=1}^n (y_i + \phi_i) \ln(1 + \mu_i \phi_i) \end{aligned} \quad (2.47)$$

Pendugaan parameter (β) diperoleh dengan menurunkan persamaan (2.47) terhadap β dan ϕ dengan menggunakan metode *Iterative Weighted Least Square* (IWLS). Alternatif lain yang mampu mengatasi masalah overdispersi adalah pendekatan *quasi-likelihood*. Pendekatan *quasi-likelihood* dianggap robust terhadap sebaran (ragam) karena memanfaatkan hubungan fungsi

ragam yaitu menghubungkan antara ragam dengan nilai tengah model sehingga diperoleh koreksi terhadap dugaan besarnya ragam.

Dalam masalah pemodelan, McCullagh dan Nelder (1989) menjelaskan mengenai konsep *quasi-likelihood* sebagai berikut,

$$\sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta} V(Y_i)^{-1} (Y_i - \mu_i) = 0 \quad (2.48)$$

dengan $E[Y_i] = \mu_i(\beta)$ dan $Var[Y_i] = \phi \mu_i(\beta)$. Jika $\phi > 1$ menunjukkan overdispersi pada model Poisson, *quasi-likelihood* dianggap mampu mengatasi overdispersi, jika fungsi ragam yang digunakan mampu menggambarkan ragam datanya.

Algoritma untuk pendugaan parameter pendekatan *quasi-likelihood* dapat dinyatakan sebagai kuadrat terkecil terboboti iteratif (*Iterative Weighted Least Square*). Pendugaan ini dapat diturunkan sebagai algoritma Gauss-Newton untuk menyelesaikan persamaan pendugaan. Metode ini merupakan algoritma umum untuk menyelesaikan persamaan *nonlinier*. Misalnya akan diselesaikan persamaan berikut,

$$\frac{\partial \mu_i}{\partial \beta} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} = \frac{\partial \mu_i}{\partial \eta_i} \mathbf{X}_i \quad (2.49)$$

Sehingga

$$\begin{aligned} \mu_i &\approx \mu_i^{(0)} + \frac{\partial \mu_i}{\partial \beta} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) \\ &= \mu_i^{(0)} + \frac{\partial \mu_i}{\partial \eta_i} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) \end{aligned} \quad (2.50)$$

dan

$$y_i - \mu_i = y_i - \mu_i^{(0)} - \frac{\partial \mu_i}{\partial \eta_i} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) \quad (2.51)$$

Dengan menempatkan persamaan (2.51) ke dalam persamaan (2.48), akan diperoleh,

$$\sum_{i=1}^n \frac{\partial \mu_i}{\partial \eta_i} V(Y_i)^{-1} \mathbf{X}_i \left\{ Y_i - \mu_i^{(0)} - \frac{\partial \mu_i}{\partial \eta_i} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) \right\} = 0 \quad (2.52)$$

yang dapat digunakan untuk menyelesaikan β pada iterasi berikutnya yang diberikan oleh persamaan sebagai berikut,

$$\beta^{(i)} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \mathbf{z} \quad (2.53)$$

dengan,

$\mathbf{X}$ = Matriks model peubah penjelas

$\mathbf{W}$ = Matriks diagonal dengan unsur-unsur sebagai berikut:

$$W_{ii} = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 V_i \text{ dan } V_i = \phi V(\mu_i^{(0)})$$

$\mathbf{z}$ = Peubah tak bebas yang disesuaikan,

$$\mathbf{z}_i = \mathbf{X}_i^T \beta^{(0)} + \frac{\partial \mu_i}{\partial \eta_i} (Y_i - \mu_i^{(0)}).$$

2.8.5 Pengujian Kesesuaian Model

Untuk melihat apakah model yang didapatkan dari hasil estimasi menggunakan metode quasi-MLE maka dilakukan penilaian terhadap model dengan menggunakan Uji Pearson *Chi-Square*. Pengujian ini dilakukan dengan hipotesis pada persamaan (2.54), yaitu jika rasio ϕ menghasilkan nilai lebih dari satu, maka model tersebut mengalami overdispersi. Hipotesis pada uji Pearson *Chi-Square* adalah sebagai berikut (Cameron dan Trivedi 1998),

$$\begin{aligned} H_0 : \phi &= 1 \\ H_1 : \phi &> 1 \end{aligned} \quad (2.54)$$

Nilai statistik uji Pearson *Chi-Square* dapat diperoleh dengan persamaan sebagai berikut,

$$\chi^2_{hitung} = \sum_{i=1}^n \frac{(y_i - E(Y))^2}{E(Y)} \quad (2.55)$$

dengan rasio dispersi

$$\phi = \frac{\chi^2}{n - k} \quad (2.56)$$

dengan y_i adalah nilai aktual pada amatan ke- i dari peubah respon, $E(Y)$ dan $Var(Y)$ adalah nilai rata-rata dan ragam dari regresi Poisson, regresi Binomial Negatif, dan pendekatan *quasi-likelihood*. n adalah banyaknya amatan pada peubah respon, dan k adalah banyaknya parameter. Di bawah kondisi H_0 benar, statistik uji χ^2 pada persamaan (2.55) akan mendekati sebaran χ^2 dengan derajat bebas $(n-k)$, sehingga keputusannya adalah menolak H_0 pada α , jika $\chi^2_{hitung} > \chi^2_{\alpha, (n-k)}$. Jika ϕ pada persamaan (2.56) lebih besar dari 1, maka dalam model terdapat masalah overdispersi, sebaliknya jika kurang dari 1 menunjukkan underdispersi. Kondisi over/underdispersi tersebut menunjukkan model tidak cocok.

Salah satu *tools* statistik untuk menilai kalibrasi probabilitas dari distribusi yang diprediksi adalah histogram PIT (*Probability Integral Transform*), yang akan mengikuti sebaran probabilitas jika distribusi yang diprediksi benar. Misalkan y_t adalah nilai dari pengamatan dan distribusi yang diprediksi adalah $P_t(y)$ dengan persamaan sebagai berikut,

$$F_t(u, y) = \begin{cases} 0, & u \leq P_t(y-1) \\ \frac{u - P_t(y-1)}{P_t(y) - P_t(y-1)}, & P_t(y-1) < u < P_t(y) \\ 1, & u \geq P_t(y) \end{cases} \quad (2.57)$$

Nilai rata-rata PIT ditentukan dengan persamaan sebagai berikut,

$$\bar{F}(u) = \frac{1}{n} \sum_{t=1}^n F_t(u|y_t), \quad 0 \leq u \leq 1 \quad (2.58)$$

Bentuk U pada histogram PIT menunjukkan overdispersi pada distribusi, sebaliknya bentuk U terbalik menunjukkan adanya underdispersi pada distribusi (Gneiting dkk., 2007).

Alat yang lain yang digunakan untuk menilai kecocokan model adalah plot kalibrasi marginal (*Marginal Calibration*), yaitu selisih dari rata-rata fungsi kepadatan probabilitas yang diprediksi dengan fungsi kepadatan probabilitas empirik dari pengamatan. Persamaan fungsi kepadatan probabilitas yang diprediksi adalah sebagai berikut (Gneiting dkk., 2007),

$$\bar{F}_T(x) = \frac{1}{T} \sum_{t=1}^T F_t(x), \quad x \in \mathbb{R} \quad (2.59)$$

dan fungsi kepadatan probabilitas empirik dari pengamatan adalah sebagai berikut (Gneiting dkk., 2007),

$$G_T(x) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}(x_t \leq x) \quad x \in \mathbb{R} \quad (2.59)$$

Jika prediksi dari model sesuai, distribusi marginal data yang diprediksi akan menyerupai distribusi marginal dari pengamatan, dan plot yang terbentuk akan mendekati nol (Christou dan Fokianos, 2015b).

2.8.6 Analisis Intervensi

Dalam banyak aplikasi, kadangkala terjadi perubahan mendadak serta kejadian yang luar biasa. Box dan Tiao (1975) menjelaskan kejadian khusus semacam ini sebagai intervensi. Sebagai contoh, wabah epidemi dalam deret waktu tertentu dari jumlah pasien mingguan yang terinfeksi penyakit tertentu, atau kebijakan yang berakibat pada menurunnya wabah epidemi, dan lain sebagainya. Maka, penting untuk memeriksa pengaruh intervensi yang diketahui ini, misalnya untuk menilai apakah adanya perubahan kebijakan memiliki dampak yang diharapkan, atau untuk memeriksa pengaruh intervensi yang tidak diketahui dan menemukan penjelasan untuk keadaan ini.

Menurut Fokianos dan Fried (2010, 2012), prediktor linear untuk model dengan sejumlah s intervensi dengan parameter $\delta_1, \dots, \delta_s$ yang terjadi pada waktu $\tau_1, \dots, \tau_s$ ditentukan dengan persamaan sebagai berikut,

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^p \beta_k g(Y_{t-i_k}) + \sum_{l=1}^q \alpha_l g(\lambda_{t-j_l}) + \eta^T X_t + \sum_{m=1}^s \omega_m \delta_m^{t-\tau_m} \mathbf{1}(t \geq \tau_m) \quad (2.60)$$

dimana $\omega_m, m=1, \dots, s$ adalah ukuran dari intervensi dan η^T adalah vektor paramater kovariat.

Pada saat terjadinya, intervensi dapat mengubah tingkat *time series* dengan menambahkan besaran ω_m untuk model linear seperti pada persamaan (2.41), atau dengan mengalikan faktor $\exp(\omega_m)$ untuk model log-linear seperti pada persamaan (2.42) (Liboschik, Fokianos, & Fried, 2016).

2.9 Demam Berdarah Dengue (DBD)

Demam dengue (DB) dan Demam Berdarah Dengue (DBD) adalah penyakit demam akut yang dapat menyebabkan kematian. Penyakit ini disebabkan oleh empat serotipe virus dari genus *Flavivirus*, virus RNA dari keluarga *Flaviviridae*. Infeksi oleh satu serotipe virus dengue menyebabkan terjadinya kekebalan yang lama terhadap serotipe virus tersebut, dan kekebalan sementara dalam waktu pendek terhadap serotipe virus dengue lainnya. Pada waktu terjadi epidemi di dalam darah seorang penderita dapat beredar lebih dari satu serotipe virus dengue. Dengue dikeluarkan oleh virus *Aedes*, nyamuk yang tersebar luas di daerah tropis dan subtropis di seluruh dunia. Demam dengue juga disebut *breakbone fever* dan merupakan penyakit virus yang ditularkan oleh nyamuk yang terpenting pada manusia (Soedarto, 2012).

2.9.1 Epidemi Dengue

Sekitar 2,5 miliar manusia yang merupakan duaperlima dari penduduk dunia mempunyai risiko tinggi tertular demam dengue. Berdasarkan laporan dari *World Health Organization* (WHO), setiap tahunnya ada sekitar 50-100 juta penderita dengue

dan 500.000 penderita Demam Berdarah Dengue di seluruh dunia, dengan jumlah kematian sekitar 22.000 jiwa, terutama anak-anak. Sekitar 2,5-3 miliar manusia yang hidup di 112 negara tropis dan subtropis berada dalam keadaan terancam infeksi dengue. Hanya benua Eropa dan Antartika yang secara alami bebas dari infeksi dengue.

Pada 50 tahun terakhir, insiden dengue di dunia telah meningkat 30 kali. Di benua Amerika saja jumlah penderita dengue meningkat dari 250.000 pada tahun 1995 menjadi lebih dari 890.000 pada tahun 2007, sedangkan penderita Demam Berdarah Dengue meningkat dari 7.000 orang pada tahun 1995 menjadi 26.000 orang (hampir 4 kali lipat) pada tahun 2007.

Daerah sebaran dengue meliputi Argentina, Australia, Bangladesh, Barbados, Bolivia, Belize, Brazil, Cambodia, Cina, Costa Rica, Dominika, El Salvador, Guatemala, Guyana, Honduras, India, Indonesia, Jamaika, Laos, Malaysia, Mexico, Micronesia, Panama, Paraguay, Philipina, Puerto Rico, Samoa, Singapura, Suriname, Taiwan, Thailand, Venezuela, Vietnam.

Berbeda dengan malaria yang sebenarnya terutama di daerah rural dan jarang dilaporkan dari daerah urban (perkotaan), sebaran dengue terutama dilaporkan dari daerah perkotaan dan lebih jarang dijumpai di daerah pedesaan (rural). Hal ini disebabkan karena nyamuk *Aedes aegypti* berkembang biak di genangan air yang terdapat di wadah (*container*) buatan seperti kaleng bekas yang diisi air, bak mandi, dan wadah penyimpanan air lainnya yang ada di dalam rumah dan di sekitar rumah (untuk nyamuk *Aedes aegypti*) atau di wadah alami seperti lubang di tanah, kolam ikan dan lainnya yang terdapat di pemukiman yang banyak tanamannya (untuk nyamuk *Aedes albopictus*).

Masing-masing serotipe virus dengue berbeda sifatnya dan tidak terdapat proteksi silang sesudah terinfeksi serotipe virus dengue yang berbeda. Pada epidemi dengue di suatu daerah hiperendemis dapat disebabkan oleh lebih dari satu jenis serotipe virus dengue (Soedarto, 2012).

2.9.2 Virus Penyebab Dengue

Demam dengue disebabkan oleh virus dengue (DEN), yang termasuk virus flavivirus. Virus yang ditularkan oleh nyamuk ini tergolong ss RNA *positive-strand virus* dari keluarga *Flaviviridae*. Terdapat empat serotipe virus DEN yang sifat antigeniknya berbeda, yaitu virus dengue-1 (DEN1), virus dengue-2 (DEN2), virus dengue-3 (DEN3) dan virus dengue-4 (DEN4). Spesifikasi virus dengue yang dilakukan oleh Albert Sabin pada tahun 1944 menunjukkan bahwa masing-masing serotipe virus dengue memiliki genotip yang berbeda antara serotipe-serotipe tersebut (Soedarto, 2012).

2.9.3 Dengue di Indonesia

Indonesia termasuk daerah endemis Demam Berdarah Dengue (DBD) dan epidemi terjadi setiap 4-5 tahun sekali. DBD pertama kali di Indonesia dilaporkan di Surabaya pada tahun 1968 dengan 58 orang penderita yang menyebabkan kematian 24 orang atau 41,3%. Puncak DBD di Indonesia terjadi pada tahun 1988 dengan angka insiden sebesar 13,45 per 100.000. Saat ini, seluruh propinsi di Indonesia sudah melaporkan terjadinya penyakit dengue/DBD baik di daerah perkotaan (urban) maupun di daerah pedesaan (rural). Menurut Departemen Kesehatan Republik Indonesia (Depkes RI) pada tahun 1989 angka kejadian DBD sebesar 6,9% dan meningkat menjadi 9,2% lima tahun kemudian, dengan angka kematian sebesar 4,5%. Selama tahun 1968-1988 insiden DBD selalu meningkat. Sebelum tahun 2000, sebagian besar penderita DBD adalah anak-anak berumur di bawah 15 tahun, tetapi sejak tahun 2001 sebagian besar penderita adalah orang dewasa (Soedarto, 2012).

Di Indonesia pada tahun 1968, DBD dilaporkan terjadi di 2 Dati II dan pada tahun 1988 meningkat menjadi 201 Dati II, dengan penderita berjumlah 47.573 orang atau 27,1 per 100.000 penduduk, dengan 1.527 kematian atau 3,2%. Sesudah dilaksanakan program DBD, pada tahun 1989 insiden penyakit ini menurun menjadi 6,1 per 100.000 penduduk. Kemudian pada

tahun 1990 dan 1991 insiden DBD meningkat menjadi 12,73 dan 11,56 per 100.000 penduduk. Sesudah insiden menurun pada tahun 1993 menjadi 9,2 per 100.000 penduduk, mulai tahun 1994 sampai tahun 1996 insiden DBD selalu meningkat dengan insien pada tahun 1996 sebesar 22,96 per 100.000 penduduk dan 211 dati II. Propinsi dengan insiden lebih dari 10 per 100.000 penduduk antara lain adalah propinsi Sumatera Barat, Jambi, Riau, Sumatera Selatan, DKI, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan dan Nusa Tenggara Timur. Semua serotipe virus dengue endemis di Indonesia, terutama serotipe DEN-2. Kota-kota besar di Jawa misalnya Jakarta, Surabaya, dan Yogyakarta umumnya merupakan daerah endemis semua serotipe virus dengue. Di luar Jawa sebagian besar endemis DEN-2 dan DEN-3, sedangkan DEN-4 dilaporkan dari Merauke (Soedarto, 2012).

Sejak tahun 1968 angka kematian akibat penyakit DBD (41,3%) selalu menurun. penurunan angka kematian rata-rata adalah 2,5% per tahun. Pada tahun 1996 angka kematian (*Case Fatality Rate*-CFR) adalah sebesar 2,7%. Menurunnya CFR disebabkan antara lain oleh penatalaksanaan kasus DBD baik di rumah sakit dan puskesmas, dan meningkatnya pemahaman masyarakat tentang DBD sehingga segera membawa penderita berobat ke rumah sakit atau puskesmas. Musim penularan DBD di Indonesia umumnya terjadi pada bulan September sampai Februari dan mencapai puncaknya pada bulan Desember-Januari, pada waktu musim penghujan. Di kota-kota besar, misalnya Jakarta, Bandung, Yogyakarta, dan Surabaya, musim penularan terjadi pada bulan Maret sampai dengan Agustus dengan puncak insiden pada bulan Juni hingga Juli.

Halaman ini sengaja dikosongkan

BAB III

METODELOGI PENELITIAN

3.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Dinas Kesehatan Provinsi Jawa Timur, yaitu jumlah penderita penyakit Demam Berdarah Dengue (DBD) setiap bulannya, yakni sebanyak 120 data dari Tahun 2007 hingga 2016.

3.2 Metode Pengolahan Data

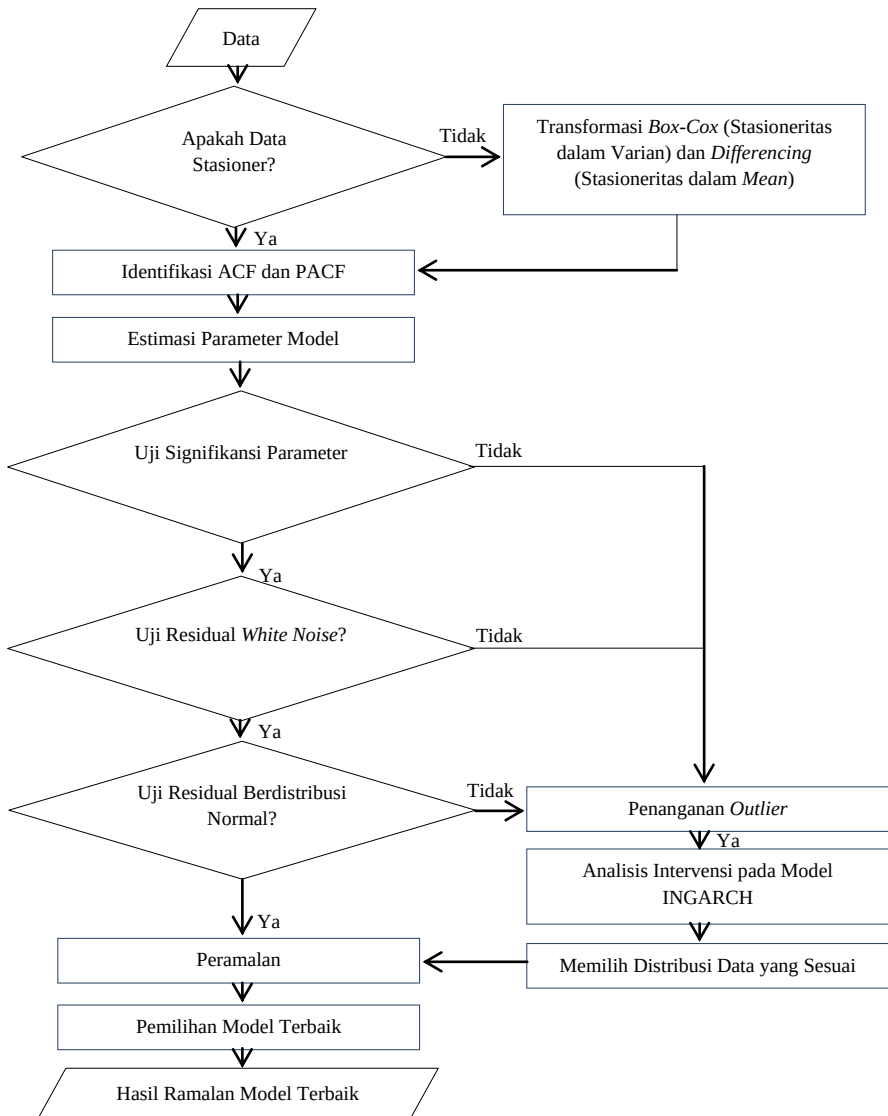
Langkah awal yang dilakukan sebelum menganalisis adalah membagi data menjadi dua yaitu data *training* dan data *testing*. Data *training* yang digunakan adalah data dari tahun 2007 hingga 2015 sebanyak 108 data, sedangkan data *testing* adalah data tahun 2016 sebanyak 12 data. Model yang digunakan pada penelitian ini adalah model *Autoregressive Integrated Moving Average* (ARIMA) dan *Integer-Valued Autoregressive Generalized Conditional Heteroscedasticity* (INGARCH). Pada model INGARCH akan disertakan pengaruh intervensi eksternal pada data, yaitu data *outlier* yang terdeteksi pada model ARIMA. Estimasi parameter yang akan digunakan adalah estimasi *quasi-Maximum Likelihood*. Hasil dari kedua model yaitu model ARIMA terbaik dan model INGARCH akan dibandingkan dengan melihat tingkat akurasi prediksi berdasarkan nilai RMSE dan MAPE yang terkecil sebagai model yang paling sesuai. Program yang digunakan untuk melakukan estimasi parameter dari kedua model adalah R, minitab, SAS, dan *Microsoft Excel*.

Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan identifikasi terhadap karakteristik kejadian Demam Berdarah di Jawa Timur dengan metode statistika deskriptif. Yaitu menganalisis nilai maksimum, nilai minimum, rata-rata dan variansi jumlah kejadian DBD berdasarkan tahun dan bulan.

2. Meramalkan kejadian Demam Berdarah di Jawa Timur menggunakan metode ARIMA berdasarkan prosedur Box-Jenkin. Langkahnya adalah sebagai berikut:
 - a. Melakukan identifikasi model pada data *training*, yaitu apakah sudah stasioner dalam *mean* dan *varians*. Jika belum stasioner dalam *mean* maka dilakukan *differencing*, sedangkan jika belum stasioner dalam *varians* maka dilakukan transformasi *Box-Cox*. Rumus untuk melakukan langkah ini terdapat pada persamaan (2.3) dan persamaan (2.5).
 - b. Melihat plot ACF dan PACF berdasarkan data yang sudah stasioner untuk menentukan orde p , d , q dan P , D , Q , S sesuai ketentuan pada Tabel 2.2.
 - c. Melakukan uji signifikansi parameter dan estimasi parameter dengan menggunakan statistik uji t sesuai dengan persamaan (2.21).
 - d. Melakukan uji kesesuaian model dengan uji *white noise* dan uji normalitas pada residual. Rumus untuk melakukan langkah ini terdapat pada persamaan (2.22) dan persamaan (2.23). Apabila residual belum memenuhi asumsi berdistribusi normal maka dilakukan deteksi *outlier* dan penanganan *outlier* dengan menyertakan *outlier* di dalam model ARIMA. Rumus untuk melakukan langkah ini terdapat pada persamaan (2.29)
 - e. Melakukan prediksi pada data *testing*, kemudian membandingkan hasil prediksi tersebut berdasarkan nilai RMSE dan MAPE menggunakan rumus pada persamaan (2.37) dan persamaan (2.39).
 - f. Melakukan pemilihan model ARIMA terbaik berdasarkan nilai RMSE dan MAPE terkecil.
3. Meramalkan kejadian Demam Berdarah di Jawa Timur menggunakan metode INGARCH dengan langkah-langkah sebagai berikut:

- a. Membuat plot ACF dan PACF data *training* untuk menentukan orde p , q dan P , Q , S sesuai ketentuan pada Tabel 2.2.
 - b. Melakukan estimasi parameter model GLM dan uji signifikansi parameter dengan menggunakan statistik uji t . Rumus untuk melakukan langkah ini terdapat pada persamaan (2.53) dan persamaan (2.55).
 - c. Jika terdapat outlier pada model ARIMA, maka disertakan fungsi intervensi pada model sesuai dengan model pada persamaan (2.60).
 - d. Melakukan estimasi parameter dan uji signifikansi parameter model GLM yang telah memuat fungsi intervensi dengan menggunakan statistik uji t . Rumus untuk melakukan langkah ini terdapat pada persamaan (2.53) dan persamaan (2.55).
 - e. Melakukan prediksi pada data *testing* secara manual.
4. Melakukan pemilihan model terbaik dengan membandingkan hasil peramalan dari metode ARIMA dan INGARCH dengan melihat tingkat akurasi yang paling tinggi berdasarkan nilai RMSE dan MAPE yang terkecil.



Gambar 3.1 Diagram Alir Model Peramalan ARIMA dan INGARCH

BAB IV

ANALISIS DATA DAN PEMBAHASAN

Analisis dan pembahasan dalam penelitian ini berisi tentang analisis deskriptif jumlah penderita DBD di Jawa Timur periode Januari 2007 sampai dengan Desember 2016, pemodelan menggunakan metode ARIMA, pemodelan menggunakan metode INGARCH pada data yang tidak stasioner dalam *mean* dan *variance*, deteksi *outlier* model ARIMA, analisis intervensi pada model INGARCH, dan perbandingan model ARIMA dan INGARCH yang terbaik berdasarkan kriteria RMSE dan MAPE.

4.1 Eksplorasi Data Jumlah Penderita Demam Berdarah Dengue di Jawa Timur

Penyakit Demam Berdarah Dengue (DBD) masih menjadi persoalan yang serius di Indonesia khususnya di Jawa Timur. Jumlah kejadian DBD terus mengalami peningkatan setiap tahunnya dan merupakan penyakit yang menjadi penyebab kematian tertinggi bagi anak-anak. Jumlah kejadian DBD setiap tahunnya berbeda-beda, maka penting untuk mengetahui karakteristik jumlah kejadian DBD dilihat berdasarkan tahun. Berikut ini ditampilkan eksplorasi data jumlah kejadian DBD di Jawa Timur selama kurun waktu sepuluh tahun sejak tahun 2007 hingga 2016.

Tabel 4.1 Statistika Deskriptif Data Jumlah Kejadian DBD Berdasarkan Tahun

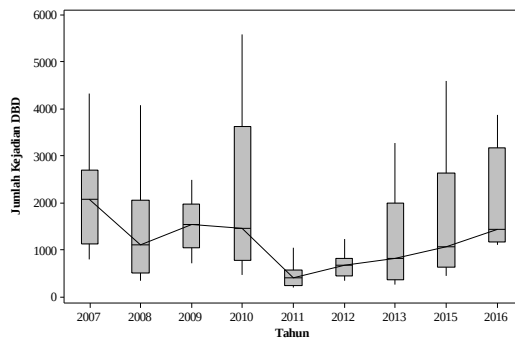
Tahun	Jumlah Bulan (n)	Rata-rata	Variansi	Min	Median	Maks
2007	12	2163	1299805	794	2088	4326
2008	12	1444	1408122	338	1102	4082
2009	12	1553	300847	725	1537	2486
2010	12	2168	3319845	476	1458	5599
2011	12	448	59663,2	192	397	1037

Tabel 4.1 Statistika Deskriptif Data Jumlah Kejadian DBD Berdasarkan Tahun (Lanjutan)

2012	12	688	66632,6	348	679,5	1225
2013	12	1211	1000206	256	828	3272
2014	12	805	28338,7	574	826	1097
2015	12	1726	2268422	451	1061	4597
2016	12	2038	1127213	1115	1435	3884

Sepanjang tahun 2007 hingga 2016 terdapat masing-masing 12 bulan, Januari sampai dengan Desember. Berdasarkan statistika deskriptif, rata-rata jumlah kejadian DBD paling sedikit terjadi pada tahun 2011 yaitu sebesar 448 kasus. Sementara, rata-rata jumlah kejadian DBD paling banyak terjadi pada tahun 2010 yaitu sebesar 2168 kasus. Nilai variansi pada tahun 2011 juga cukup kecil sehingga jumlah kejadian DBD pada tahun ini memiliki keragaman yang kecil. Sementara, nilai variansi pada tahun 2010 paling besar, artinya jumlah kejadian DBD pada tahun ini memiliki keragaman yang besar. Diketahui pula jumlah kejadian DBD paling banyak terjadi pada tahun 2010 yaitu sebesar 5599 kasus, dan jumlah kejadian DBD paling sedikit terjadi pada tahun 2011 yaitu sebesar 192 kasus. Berdasarkan publikasi dari Kementerian Kesehatan Provinsi Jawa Timur, untuk kasus DBD dikenal adanya siklus 5 tahunan. Selama periode 5 tahun yakni sejak tahun 2011 hingga tahun 2015, biasanya kasus DBD akan mengalami peningkatan pada tahun 2015 dibandingkan periode tahun sebelumnya. Berdasarkan Tabel 4.1 terlihat bahwa sepanjang tahun 2011 hingga tahun 2014, jumlah kasus DBD kecil dibandingkan tahun-tahun sebelumnya. Kemudian meningkat di tahun 2015. Demikian juga kasus DBD tertinggi terjadi pada tahun 2010. Berdasarkan keterangan dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Jawa Timur, pada tahun 2010 dan 2015 di awal tahun terjadi curah hujan yang sangat tinggi. Curah hujan yang tinggi di setiap awal tahun menciptakan banyak genangan air yang merupakan habitat perkembangbiakan nyamuk penular DBD, yakni *Aedes aegypti*.

Tingginya jumlah kejadian DBD pada waktu tertentu bisa menjadi pengamatan yang tidak biasa atau *outlier*. Berdasarkan hasil *boxplot* pada Gambar 4.1, secara visual tidak ada *outlier* yang terdeteksi. Pada Gambar 4.1 juga terlihat bahwa jumlah kejadian DBD di Jawa Timur sepanjang tahun 2007 hingga 2016 cenderung fluktuatif atau naik turun jika dilihat dari median data.



Gambar 4.1 Boxplot Data Jumlah Kejadian DBD Berdasarkan Tahun

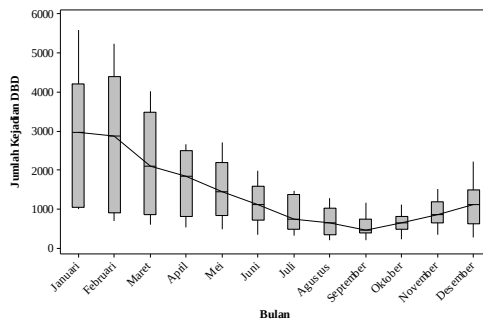
Selain melihat karakteristik kejadian DBD berdasarkan tahun, perlu juga untuk mengetahui karakteristik kejadian DBD berdasarkan bulan. Sebab, kejadian DBD biasanya akan sangat tinggi pada masa peralihan musim, yaitu sekitar bulan Januari hingga Maret. Berikut ini ditampilkan eksplorasi data jumlah kejadian DBD di Jawa Timur selama kurun waktu sepuluh tahun sejak tahun 2007 hingga 2016 berdasarkan bulan.

Data jumlah hari pada setiap bulannya selama 10 tahun sejak tahun 2007 hingga tahun 2016 berbeda-beda. Adapun rata-rata jumlah kejadian DBD tertinggi terjadi pada bulan Januari yaitu sebesar 2933 kasus, sementara jumlah kejadian paling sedikit terjadi pada bulan September yaitu sebesar 552. Hal ini dipengaruhi oleh tinggi rendahnya curah hujan di setiap bulannya.

Tabel 4.2 Statistika Deskriptif Data Jumlah Kejadian DBD Berdasarkan Bulan

Variable	Mean	Variance	Minimum	Median	Maximum
Januari	2933	2557223	1004	2960	5599
Februari	2810	2854846	682	2867	5251
Maret	2237	1602152	593	2107	4018
April	1730	628434	534	1836	2654
Mei	1505	509213	485	1445	2714
Juni	1131	294082	332	1121	1971
Juli	872	198068	308	728	1469
Agustus	677	121708	197	634	1264
September	552	75448,4	192	463,5	1147
Oktober	640	66146,7	220	640	1116
November	879	133120	342	862	1516
Desember	1129	412255	256	1106	2205

Pada bulan September hingga Oktober baru memasuki musim penghujan, dimana pada peralihan musim kemarau menuju musim hujan, curah hujan sangat rendah sehingga vektor nyamuk *Aedes aegypti* masih belum berkembang menjadi nyamuk dewasa.



Gambar 4.2 Boxplot Data Jumlah Kejadian DBD Berdasarkan Bulan

Pada saat curah hujan tinggi, yakni setiap awal tahun pada bulan Januari hingga Maret berkembangbiakan vektor nyamuk *Aedes aegypti* akan semakin pesat. Hal ini memberikan pengaruh kepada tingginya jumlah kejadian DBD pada bulan-bulan tersebut. Terlihat pula adanya variansi yang tinggi pada bulan Januari, artinya jumlah kejadian DBD pada bulan ini memiliki keragaman yang besar. Sementara nilai variansi pada bulan September cukup kecil dibandingkan bulan yang lain, artinya jumlah kejadian DBD di bulan ini memiliki keragaman yang kecil.

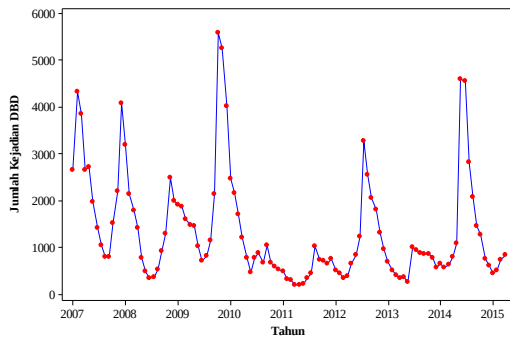
Berdasarkan *boxplot* pada Gambar 4.2 tidak terlihat adanya *outlier*. Namun tampak bahwa kejadian DBD pada bulan Januari hingga Desember cenderung membentuk pola menurun.

4.2 Identifikasi Model ARIMA untuk Meramalkan Jumlah Kejadian Demam Berdarah Dengue di Jawa Timur

Identifikasi model ARIMA dalam penelitian ini digunakan untuk mengetahui model yang tepat dalam meramalkan kejadian DBD di Jawa Timur untuk masa yang akan datang dalam jangka pendek.

4.2.1 Identifikasi Model

Identifikasi model dilakukan pada data *in-sample* yang sudah stasioner, yaitu data kejadian DBD di Jawa Timur mulai tahun 2007 sampai dengan 2015, yakni sebanyak 108 data. Untuk menstasionerkan data, harus diketahui kebutuhan transformasi yang tepat untuk data jumlah kejadian DBD seperti transformasi untuk menstabilkan variansi dan transformasi untuk *differencing*. Identifikasi model juga dilakukan untuk menentukan karakteristik data *out sample*, serta menentukan orde p dan q yang tepat untuk model. Langkah pertama adalah membuat plot *time series* sebagai visualisasi data jumlah kejadian DBD bulanan sebanyak 108 untuk data *in sample* mulai dari tahun 2007 hingga tahun 2015. Plot *time series* yang memberikan visualisasi jumlah kejadian DBD ditunjukkan oleh Gambar 4.3.

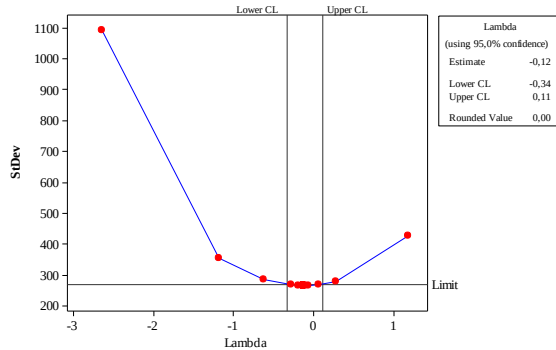


Gambar 4.3 Plot *Time Series* Jumlah Kejadian DBD di Jawa Timur

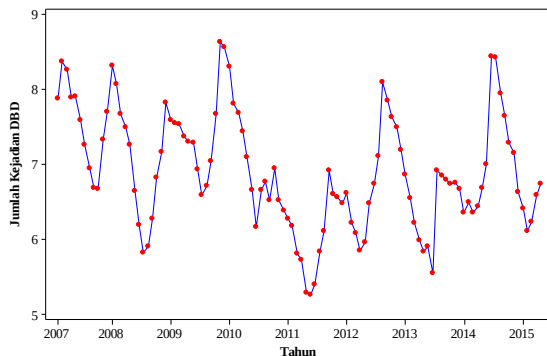
Visualisasi plot *time series* dari data jumlah kejadian DBD di Jawa Timur terlihat fluktuatif dan menunjukkan adanya efek musiman. Jumlah kejadian berubah setiap periode waktu 12 bulan, maka bisa dikatakan data jumlah kejadian DBD di Jawa Timur memiliki pola musiman setiap 12 bulan atau pola tahunan. Terlihat pula jumlah kejadian DBD di sepanjang tahun 2007, 2008, 2010, 2013 dan 2015 cenderung tinggi, sementara di tahun 2009, 2011, 2012, dan 2014 jumlah kejadian cenderung rendah. Berdasarkan visualisasi tersebut, diindikasikan bahwa data jumlah kejadian DBD di Jawa Timur tidak stasioner dalam variansi dan rata-rata yang terlihat pada penyebaran jumlah kejadian yang tidak berada di sekitar variansi dan rata-rata. Data jumlah kejadian DBD yang tidak stasioner dalam variansi dapat diatasi dengan transformasi *Box-Cox*.

Plot *Box-Cox* pada Gambar 4.4 menunjukkan bahwa batas kendali lamda sebelum transformasi sebesar $-0,12$ dan p -value sebesar $0,000$ sehingga bisa dilakukan transformasi dengan rumus $\ln(Z_t)$. Setelah dilakukan transformasi dan pemeriksaan stasioneritas terlihat pada Gambar 4.5 bahwa plot *time series* data telah stasioner dalam variansi. Namun, terlihat masih adanya inkonsistensi jumlah kejadian DBD, yang mana pada tahun 2011,

2012, dan 2014 jumlah kejadian cenderung rendah, sementara pada tahun-tahun yang lain cenderung tinggi.



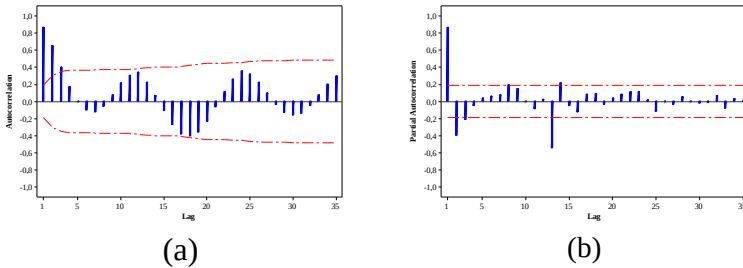
Gambar 4.4 Plot *Box-Cox* Jumlah Kejadian DBD di Jawa Timur



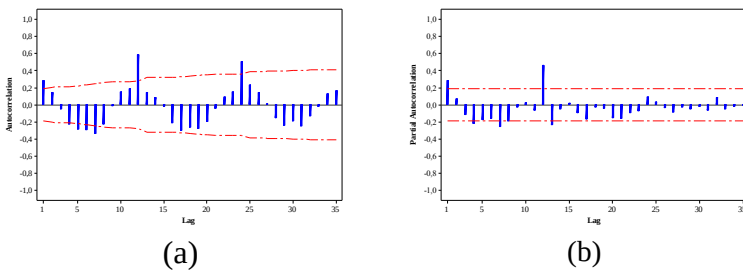
Gambar 4.5 Plot *Time Series* Jumlah Kejadian DBD di Jawa Timur Setelah Transformasi

Indikasi data jumlah kejadian DBD yang tidak stasioner dalam rata-rata dapat divisualisasikan dengan plot ACF dan PACF pada data *in-sampel* yang ditunjukkan pada Gambar 4.6. Plot ACF menunjukkan bahwa terdapat pola sinusoidal teredam sedangkan PACF menunjukkan pola *cut off*. Selanjutnya untuk menstasionerkan rata-rata pada data dilakukan *differencing* pada

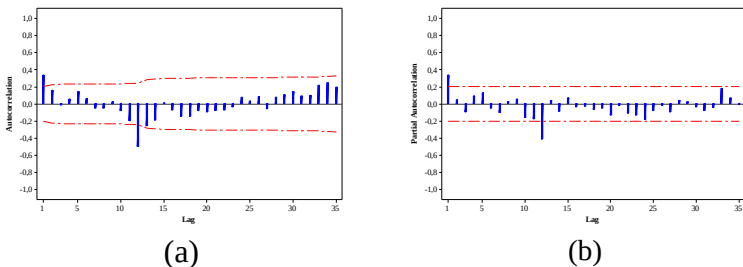
lag 1. Hasil dari proses *differencing* secara umum menunjukkan bahwa kondisi data sudah stasioner dalam *mean*. Plot ACF dan PACF dapat dilihat pada Gambar 4.7.



Gambar 4.6 Plot (a) ACF dan (b) PACF Jumlah Kejadian DBD di Jawa Timur Sebelum *Differencing*



Gambar 4.7 Plot (a) ACF dan (b) PACF Jumlah Kejadian DBD di Jawa Timur Setelah *Differencing* pada Lag 1



Gambar 4.8 Plot (a) ACF dan (b) PACF Jumlah Kejadian DBD di Jawa Timur Setelah *Differencing* pada Lag 12

Plot ACF menunjukkan pola musiman, sehingga dilakukan *differencing* pada lag 12. Maka dilakukan dua kali *differencing*

yaitu *differencing* reguler pertama dan *differencing* musiman pertama. Plot ACF dan PACF hasil *differencing* pada lag 12 dapat dilihat pada Gambar 4.8. Plot ACF pada Gambar 4.7 (a) menunjukkan bahwa data bergerak turun cepat dan *cut off* pada lag 1, 12, dan 24. Sedangkan plot PACF pada Gambar 4.7 (b) menunjukkan bahwa data bergerak turun cepat dan *cut off* pada lag 1 dan 12.

Plot ACF hasil *differencing* musiman pertama pada Gambar 4.8 menunjukkan bahwa data bergerak turun cepat dan *cut off* pada lag 1 dan 12, sedangkan pola PACF menunjukkan bahwa data bergerak turun cepat dan *cut off* pada lag 1 dan 12. Dugaan model sementara berdasarkan plot ACF dan PACF pada data jumlah kejadian DBD di Jawa Timur adalah ARIMA (0,1,1)(0,1,1)¹², ARIMA (0,1,1)(1,1,0)¹², ARIMA (1,1,0)(1,1,0)¹², dan ARIMA (1,1,0)(0,1,1)¹².

4.2.2 Estimasi dan Pengujian Signifikansi Parameter

Setelah dilakukan identifikasi model untuk mendapatkan model dugaan sementara peramalan jumlah kejadian DBD di Jawa Timur, selanjutnya dilakukan pengujian signifikansi parameter untuk menunjukkan apakah parameter sudah signifikan terhadap model atau layak untuk dimasukkan ke dalam model. Hasil estimasi parameter ditunjukkan dalam Tabel 4.3.

Berdasarkan hasil estimasi parameter yang ditunjukkan oleh Tabel 4.3 terlihat bahwa semua model yang telah diduga memiliki parameter yang signifikan. Hal ini terlihat dari parameter masing-masing model tersebut yang memiliki nilai *p-value* kurang dari taraf signifikansi *alpha* 5%. Bisa juga dilihat dari nilai $|t_{hitung}| > t_{\alpha/2, n-n_p}$. Nilai $t_{0,025;93}$ sebesar 1,986 yang artinya tolak H_0 , sehingga parameter dari model telah sesuai.

Dari keempat model yang telah diduga, terlihat bahwa model ARIMA (1,1,0)(1,1,0)¹², ARIMA (1,1,0)(0,1,1)¹², dan ARIMA (0,1,1)(1,1,0)¹² telah memiliki parameter yang signifikan, sementara model ARIMA (0,1,1)(0,1,1)¹² hanya memiliki satu parameter yang signifikan, sehingga model yang dipilih adalah

model ARIMA dengan semua paramaternya yang telah signifikan.

Tabel 4.3 Uji Signifikansi Model ARIMA

Model ARIMA	Parameter	Estimasi	t-Value	p-Value
(1,1,0)(1,1,0) ¹²	ϕ_1	0,2710	2,70	0,008
	Φ_1	-0,5828	-6,66	0,000
(1,1,0)(0,1,1) ¹²	ϕ_1	0,2089	2,11	0,038
	Θ_1	0,8456	10,17	0,000
(0,1,1)(0,1,1) ¹²	θ_1	-0,1925	-1,88	0,063
	Θ_1	0,8057	8,60	0,000
(0,1,1)(1,1,0) ¹²	θ_1	-0,5953	-6,85	0,000
	Φ_1	-0,2372	-2,36	0,021

Selain itu, berdasarkan nilai *Mean Absolute Error* (MSE), terlihat bahwa masing-masing model memiliki nilai MSE yang kecil. Untuk memperkuat keputusan pemilihan model terbaik, maka dilakukan uji kesesuaian model pada semua model ARIMA, yaitu apakah residual telah memenuhi asumsi *white noise* dan berdistribusi normal. Nilai kriteria kebaikan model ARIMA ditampilkan pada Tabel 4.4.

Tabel 4.4 MSE Model ARIMA

Model ARIMA	MSE
(1,1,0)(1,1,0) ¹²	0,05773
(1,1,0)(0,1,1) ¹²	0,04688
(0,1,1)(0,1,1) ¹²	0,04838
(0,1,1)(1,1,0) ¹²	0,05824

4.2.3 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui apakah residual yang satu dengan yang lainnya pada ketiga model terpilih, tidak memiliki korelasi atau independen dan memiliki pola yang menyebar atau acak (identik). Residual harus memenuhi asumsi *white noise*. Pengujian asumsi *white noise* pada masing-masing model dapat dilihat pada Tabel 4.5.

Tabel 4.5 Pemeriksaan Diagnostik *White Noise* Terhadap Residual pada Model ARIMA

Model ARIMA	Sampai Lag ke-	χ^2_{hitung}	db	p-Value
(1,1,0)(1,1,0) ¹²	12	4,4	10	0,927
	24	14,2	22	0,893
	36	24,2	34	0,894
	48	33,6	46	0,914
(1,1,0)(0,1,1) ¹²	12	6,8	10	0,745
	24	17,1	22	0,756
	36	27,3	34	0,787
	48	38,7	46	0,769
(0,1,1)(1,1,0) ¹²	12	7,2	10	0,710
	24	18,1	22	0,702
	36	28,4	34	0,739
	48	41,1	46	0,678

Berdasarkan uji diagnostik *white noise* atau uji asumsi residual berdistribusi normal yang ditunjukkan oleh Tabel 4.5, maka terlihat bahwa semua model ARIMA yang diduga telah memenuhi asumsi residual berdistribusi normal. Hal ini dilihat dari nilai *p-value* yang lebih besar dari nilai signifikansi *alpha* sebesar 5%, yang artinya gagal menolak H_0 , sehingga pengamatan sekarang tidak memiliki autokorelasi dengan pengamatan sebelumnya dan cenderung memiliki pola yang random/acak. Tahap selanjutnya adalah melakukan pemeriksaan diagnostik residual berdistribusi normal.

4.2.4 Uji Asumsi Distribusi Residual Berdistribusi Normal

Selanjutnya dilakukan pemeriksaan diagnostik residual berdistribusi normal. Hasil pengujian diagnostik residual berdistribusi normal ditunjukkan oleh Tabel 4.6. Residual yang diperoleh dari ketiga model ini adalah sebanyak data *in-sample* dikurangi dengan tiga belas ($108 - 13 = 95$), sehingga kuantil dari statistik Kolmogorov smirnov dalam taraf signifikansi α 5% adalah sebesar $D_{(1-0,05;n=95)} = 1,22/\sqrt{95} = 0,125$. Pada pengujian ini, hipotesis yang diuji adalah residual berdistribusi normal. Berdasarkan pengujian yang ditampilkan pada Tabel 4.6, nilai $D < D_{(1-0,05;n=95)}$, yang artinya gagal menolak H_0 sehingga residual dari ketiga model telah berdistribusi normal. Tahap selanjutnya adalah melakukan peramalan pada data untuk kriteria *in-sample* dan *out-sample*, kemudian dilakukan pemilihan model terbaik berdasarkan nilai RMSE dan MAPE terkecil.

Tabel 4.6 Uji Kolmogorov Smirnov

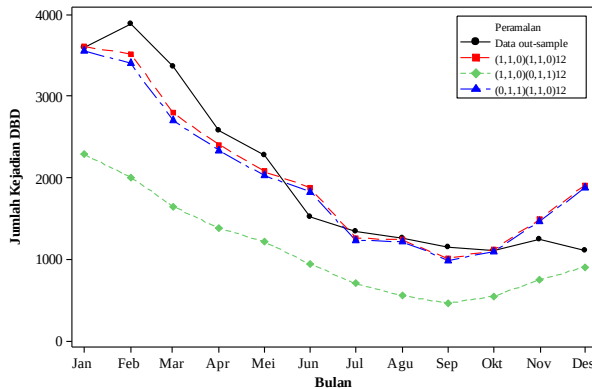
Model ARIMA	D	$D_{(1-0,05;n>40)}$	p-value
$(1,1,0)(1,1,0)^{12}$	0,048	0,125	$> 0,150$
$(1,1,0)(0,1,1)^{12}$	0,044	0,125	$> 0,150$
$(0,1,1)(1,1,0)^{12}$	0,049	0,125	$> 0,150$

Tahap selanjutnya adalah melakukan peramalan pada data untuk memilih model terbaik berdasarkan hasil ramalan terbaik.

4.2.5 Pemilihan Model ARIMA Terbaik

Pemilihan model ARIMA menghasilkan tiga model dari pemodelan data jumlah kejadian DBD di Jawa Timur yang diterapkan pembeda reguler dan musiman yang telah memenuhi signifikansi parameter, asumsi residual *white noise*, dan asumsi residual berdistribusi normal. Selanjutnya akan dibandingkan ketiga model tersebut untuk memilih model terbaik yang akan

digunakan untuk melakukan peramalan di masa mendatang. Suatu model yang digunakan dalam peramalan, dipilih berdasarkan kemampuannya dalam meramalkan nilai-nilai di masa mendatang. Hasil peramalan dari masing-masing model dapat ditunjukkan secara visual menggunakan plot *time series*. Perbandingan hasil peramalan dari ketiga model dengan data *out-sample* jumlah kejadian DBD di Jatim sebanyak setahun atau 12 bulan ditunjukkan oleh plot *time series* pada Gambar 4.9.



Gambar 4.9 Perbandingan Data *Out-Sample* dengan Hasil Peramalan Ketiga Model ARIMA

Berdasarkan plot *time series*, diketahui bahwa peramalan ketiga model ARIMA memiliki kecenderungan pola yang sama dan telah mengikuti pola data *out-sample*. Namun, jika dibandingkan dengan dua model ARIMA yang lainnya, hasil peramalan untuk model ARIMA $(1,1,0)(0,1,1)^{12}$ berada jauh di bawah data *out-sampl*. Adapun akurasi model dapat ditunjukkan berdasarkan kriteria kebaikan model. Dalam hal ini, kriteria yang digunakan adalah RMSE dan MAPE. Kriteria *in-sample* dan *out-sample* diterapkan pada ketiga model sebagai berikut.

Ketiga model hasil dari pemodelan dengan data yang telah mengalami pembedaan reguler pertama dan musiman pertama cenderung menghasilkan nilai RMSE *in-sample* yang lebih besar.

Adapun nilai RMSE *in-sample* yang terkecil dihasilkan oleh model ARIMA (1,1,0)(0,1,1)¹². Tujuan dari pemodelan adalah peramalan nilai di masa mendatang, sehingga kriteria kebaikan yang digunakan untuk pemilihan model adalah kriteria *out-sample*.

Tabel 4.7 Pemilihan Model ARIMA Terbaik Berdasarkan Kriteria Kebaikan Model

Model ARIMA	Kriteria <i>In-sample</i>	Kriteria <i>Out-sample</i>	
	RMSE	RMSE	MAPE (%)
(1,1,0)(1,1,0) ¹²	31345,7	338,897	14,74
(1,1,0)(0,1,1) ¹²	1103,1	1041,275	44,74
(0,1,1)(1,1,0) ¹²	27296,5	360,828	15,58

Berdasarkan nilai MAPE, model ARIMA (1,1,0)(1,1,0)¹² memiliki nilai MAPE terkecil. Berdasarkan kriteria data *out-sample*, diketahui bahwa model ARIMA (1,1,0)(1,1,0)¹² memiliki kemampuan yang lebih baik dibandingkan kedua model ARIMA yang lain dalam meramalkan jumlah kejadian DBD di Jawa Timur, dengan nilai MAPE sebesar 14,74% dan nilai RMSE sebesar 338,897. Adapun model ARIMA (1,1,0)(1,1,0)¹² dapat dituliskan sebagai berikut,

$$Z_t = 0,271Z_{t-1} + 0,5828Z_{t-12} - 0,15794Z_{t-13} + a_t$$

Model ARIMA (1,1,0)(1,1,0)¹² menunjukkan bahwa jumlah kejadian DBD di Jawa Timur pada waktu ke t berkaitan dengan jumlah kejadian DBD pada waktu ke $t-1$, $t-12$, dan $t-13$.

4.3 Identifikasi Model INGARCH untuk Meramalkan Jumlah Kejadian Demam Berdarah Dengue di Jawa Timur

Identifikasi model INGARCH dalam penelitian ini digunakan untuk mengetahui model yang tepat dalam meramalkan kejadian DBD di Jawa Timur untuk masa yang akan datang pada data yang telah stasioner dalam *mean* dan variansi.

4.3.1 Identifikasi Model

Identifikasi model INGARCH dilakukan untuk menentukan karakteristik data *out sample*, serta menentukan orde p dan q yang tepat untuk model. Langkah awal sebelum melakukan analisis adalah membuat plot *time series* pada data jumlah kejadian DBD di Jawa Timur untuk data *in-sample*, yang ditampilkan pada Gambar 4.3.

Selanjutnya untuk menentukan orde p dan q pada model INGARCH maka dilakukan analisis terhadap plot ACF dan PACF dari data *in-sample*. Visualisasi plot ACF dan PACF pada data ditampilkan pada Gambar 4.6. Plot ACF pada gambar 4.6 menunjukkan bahwa data bergerak turun cepat dan *cut off* pada lag 1, 2, 3, 4, 12, 13, 18, 19, dan 20. Sedangkan plot PACF menunjukkan bahwa data bergerak turun cepat dan *cut off* pada lag 1, 2, dan 13. Dugaan model sementara yang memungkinkan berdasarkan plot ACF dan PACF pada data jumlah kejadian DBD di Jawa Timur adalah INGARCH (1,0), INGARCH (2,0), INGARCH ([1,2,13],0), INGARCH ([1,13],0), INGARCH (1,12), INGARCH ([1,2,13],1), INGARCH (1,12), INGARCH (1,[1,12]), INGARCH (1,[1,2,12,13]), dan INGARCH (1,[1,13]).

4.3.2 Estimasi dan Pengujian Signifikansi Parameter Model

Setelah dilakukan identifikasi model untuk mendapatkan model dugaan sementara peramalan jumlah kejadian DBD di Jawa Timur, selanjutnya dilakukan pengujian signifikansi parameter untuk menunjukkan apakah parameter sudah signifikan terhadap model atau layak untuk dimasukkan ke dalam model. Estimasi dan pengujian signifikansi parameter dilakukan pada semua model INGARCH yang telah diduga hingga diperoleh tiga model INGARCH yang paling sesuai, yakni model INGARCH

(1,0), INGARCH ([1,13],0), dan INGARCH (1,12). Selain itu, estimasi model juga dilakukan pada model INGARCH- d ([1,12,13],0), yakni orde p dan q dengan *differencing* pertama dan *differencing* musiman sesuai dengan model ARIMA terbaik yang telah terpilih sebelumnya.

Tabel 4.8 Data *Differencing* Pertama dan Musiman Pertama

Data ke-	Data <i>Differencing</i>	Data yang Telah Positif
1	-2575	530
2	-566	2539
3	848	3953
4	-432	2673
5	90	3195
...	...	...
36	-3104	1
...	...	...
91	-240	2865
92	-65	3040
93	5	3110
94	43	3148
95	-180	2925

Adapun estimasi model INGARCH- d yakni model INGARCH berdasarkan hasil model ARIMA terbaik dengan *differencing* pertama dan *differencing* musiman dilakukan secara manual, dikarenakan belum tersedia paket statistik untuk model ini. Untuk melakukan estimasi model INGARCH- d adalah dengan mempositifkan data hasil *differencing* pertama dan musiman pertama. Nilai data yang paling minimum dijumlahkan dengan bilangan bulat positif sehingga data *differencing* bernilai positif. Pada data hasil *differencing* diperoleh nilai paling kecil sebesar -3104, sehingga semua data dijumlahkan dengan bilangan 3105. Selanjutnya dilakukan estimasi parameter pada data yang telah positif. Data hasil *differencing* ditampilkan pada

Tabel 4.8 berikut. Adapun hasil estimasi parameter model ditunjukkan oleh Tabel 4.9 berikut.

Tabel 4.9 Uji Signifikansi Model INGARCH

Model INGARCH	Parameter	Estimasi	Std. Error	CI (lower)	CI (upper)
(1,0)	β_0	248,600	2,700	239,532	257,668
	β_1	0,84900	0,004	0,84100	0,85700
([1,13],0)	β_0	21800,0	5,2846	207.740	22800,0
	β_1	0,87300	0,0041	0,86500	0,88200
	β_{13}	1,34e-09	0,0031	-0,0062	0,00618
(1,12)	β_0	49000,0	7,3172	475,901	50500,0
	β_1	0,68600	0,0041	0,67800	0,69400
	α_{12}	1,25e-10	0,0048	-0,0095	0,00956
([1,12,13],0)	β_0	3090,00	54,592	2980,2	3190,0
	β_1	2,90e-06	0,0090	-0,0178	0,0178
	β_{12}	3,82e-11	0,0090	-0,0177	0,0177
	β_{13}	1,12e-10	0,0107	-0,0210	0,0210

Model GLM yang diperoleh mengikuti distribusi Poisson dengan fungsi link *identity*. Berdasarkan hasil estimasi parameter dari ketiga model, diperoleh hanya model INGARCH (1,0) yang telah sesuai karena semua parameter telah signifikan dengan taraf signifikansi *alpha* 5% atau nilai $|t_{hitung}|$ model INGARCH (1,0) telah lebih besar dari $t_{tabel} = 1,983$. Sedangkan model INGARCH (1,12) dan INGARCH ([1,13],0) tidak sesuai disebabkan ada parameter yang tidak signifikan pada taraf signifikansi *alpha* 5%. Model INGARCH-*d* ([1,12,13],0) juga tidak sesuai disebabkan ada parameter yang tidak signifikan pada taraf signifikansi *alpha* 5%. Namun, model INGARCH-*d* ([1,12,13],0) akan disertakan sebagai model terpilih untuk melakukan peramalan. Sebab, model

ini merupakan hasil dari pemilihan model ARIMA terbaik dengan *differencing* pertama dan *differencing* musiman, walaupun pada model INGARCH terdapat parameter yang tidak signifikan. Maka, peramalan akan dilakukan dengan menyertakan parameter tidak signifikan, apakah hasil ramalan tersebut lebih baik dari hasil peramalan dengan model dengan semua parameter yang telah signifikan.

Peramalan dengan menyertakan parameter model yang tidak signifikan telah dikembangkan oleh Hyndman dan Kostenko (2008), yang mengatakan bahwa peramalan dapat dilakukan tanpa nilai *p-value* dan tingkat signifikansi *alpha*. Mereka juga menyarankan agar lebih memperhatikan pemilihan model yang mampu menghasilkan peramalan terbaik. Selain itu, Armstrong (2007) juga mengatakan bahwa tes signifikansi statistik berbahaya atau mengganggu bagi pengembangan pengetahuan ilmiah, khususnya dalam metode peramalan karena akan mengalihkan perhatian peneliti dari penggunaan metode yang tepat. Dia menyarankan agar lebih memperhatikan ukuran efek, *confident interval*, replikasi/ekstensi, dan meta-analisis. Tahap selanjutnya adalah melakukan peramalan pada data *out-sample* untuk memilih model INGARCH terbaik berdasarkan hasil ramalan terbaik.

4.3.3 Pemilihan Model INGARCH Terbaik

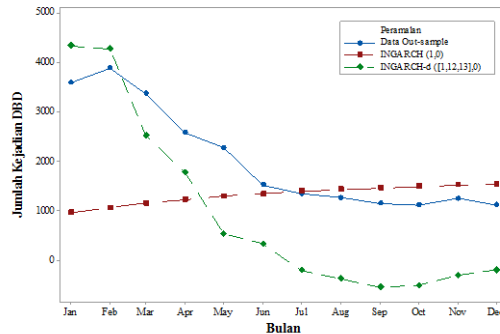
Pemilihan model INGARCH menghasilkan dua model dari pemodelan data jumlah kejadian DBD di Jawa Timur dengan distribusi data mengikuti distribusi Poisson dan fungsi link *identity*. Selanjutnya akan dibandingkan kedua model tersebut untuk memilih model terbaik yang akan digunakan untuk melakukan peramalan di masa mendatang. Suatu model yang digunakan dalam peramalan, dipilih berdasarkan kemampuannya dalam meramalkan nilai-nilai di masa mendatang. Hasil peramalan dari masing-masing model dapat ditunjukkan secara visual menggunakan plot *time series*.

Adapun peramalan model INGARCH- d dilakukan secara manual dengan cara mengembalikan data yang telah dipositifkan kepada data hasil *differencing* pertama dan *differencing* musiman atau data asli hasil *differencing*. Untuk menghitung hasil ramalan maka digunakan rumus berikut,

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13}$$

dimana $Y_{t-1} = \lambda_t - 3105$, λ_t adalah hasil peramalan model dari data *differencing* yang telah dipositifkan, dan Y_{t-1} adalah hasil ramalan data asli yang telah stasioner dengan *differencing* pertama dan *differencing* musiman. Misal diperoleh hasil ramalan dari data yang telah dipositifkan pada periode bulan Januari sebesar $\lambda_{109} = 3092$, maka hasil ramalan pada periode tersebut adalah,

$$Y_{109} = (\lambda_{109} - 3105) + Y_{109-12} - Y_{109-13}$$



Gambar 4.10 Perbandingan Data *Out-Sample* dengan Hasil Peramalan Model INGARCH

Perbandingan hasil peramalan dari kedua model dengan data *out-sample* jumlah kejadian DBD di Jatim sebanyak setahun atau dua belas bulan ditunjukkan oleh plot *time series* pada Gambar 4.10. Berdasarkan plot *time series*, diketahui bahwa peramalan model INGARCH (1,0) memiliki pola yang tidak sama

dengan data *out-sample*, artinya data tidak mengikuti pola data *out-sample*. Hal ini mengindikasikan bahwa model yang didapatkan tidak sesuai. Sementara model INGARCH- d $([1,12,13],0)$ dengan menyertakan paramater yang tidak signifikan memiliki pola yang cenderung sama dengan data *out-sampel*, walaupun ada nilai hasil ramalan yang negatif. Hal ini disebabkan hasil peramalan pada bulan ke tujuh negatif, sehingga memberikan pengaruh besar kepada hasil peramalan untuk periode berikutnya.

Hasil peramalan yang negatif menandakan adanya *outlier* yang besar pada data. Maka, model INGARCH- d $([1,12,13],0)$ hanya bisa digunakan untuk meramalkan kejadian DBD di Jawa Timur untuk periode enam bulan ke depan. Analisis terhadap indikasi adanya *outlier* pada data untuk model INGARCH- d akan dilakukan pada sub bab 4.4. Adapun akurasi model dapat ditunjukkan berdasarkan kriteria kebaikan model. Dalam hal ini, kriteria yang digunakan adalah RMSE dan MAPE model untuk peramalan periode enam bulan ke depan. Kriteria *out-sample* diterapkan pada kedua model INGARCH ditunjukkan oleh Tabel 4.10 berikut.

Tabel 4.10 Pemilihan Model INGARCH Terbaik Berdasarkan Kriteria Kebaikan Model

Pembeda	Model	RMSE	MAPE (%)
-	INGARCH (1,0)	1940,392	53,15
$(1-B)(1-B^{12})Z_t$	INGARCH- d $([1,12,13],0)$	1045,120	40,38

Pemodelan INGARCH yang memberikan nilai MAPE dan RMSE terkecil berdasarkan kriteria *out-sample* dan *in-sample* adalah model INGARCH- d $([1,12,13],0)$, yaitu model yang menyertakan paramater tidak signifikan dengan *differencing* pertama dan *differencing* musiman pertama, dengan nilai MAPE

sebesar 40,38% dan nilai RMSE sebesar 1045,120. Model INGARCH- d ([1,12,13],0) dapat dituliskan sebagai berikut.

$$\lambda_t^d = 3090 + 0,0000029Y_{t-1} + 0,0000000000382Y_{t-12} \\ + 0,000000000112Y_{t-13}$$

dimana Y_t pada periode t diberikan oleh $Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t^d)$. Model INGARCH- d ([1,12,13],0) menunjukkan bahwa jumlah kejadian DBD di Jawa Timur pada waktu ke t berkaitan dengan jumlah kejadian DBD pada waktu ke $t-1$, $t-12$, dan $t-13$. Hasil ramalan dari model adalah nilai median pada paramater λ_t^d . Untuk mengatasi nilai hasil peramalan yang negatif disebabkan adanya data yang *outlier* maka dilakukan analisis intervensi.

4.4 Analisis Intervensi

Untuk mengatasi hasil peramalan model INGARCH- d ([1,12,13],0) yang negatif disebabkan adanya *outlier* pada data, maka dilakukan analisis intervensi pada model. Langkah pertama adalah melakukan deteksi *outlier* pada residual model ARIMA. (1,1,0)(1,1,0)¹². Hasil deteksi *outlier* residual model ditampilkan oleh Tabel 4.10.

Tabel 4.11 Deteksi *Outlier* Residual Model ARIMA Terbaik

Observasi	Tipe	p-value
36	Shift	<0,0001
37	Shift	<0,0001
39	Shift	<0,0001
40	Shift	<0,0001
49	Shift	<0,0001
73	Additive	<0,0001
97	Shift	<0,0001
99	Shift	<0,0001

Berdasarkan table 4.11 terdapat *outlier* pada model ARIMA $(1,1,0)(1,1,0)^{12}$ sehingga didapatkan hasil peramalan model INGARCH- d yakni model INGARCH dengan *differencing* pertama dan *differencing* musiman pertama bernilai negatif. Maka langkah selanjutnya adalah melakukan analisis intervensi pada model INGARCH berdasarkan hasil ARIMA terbaik.

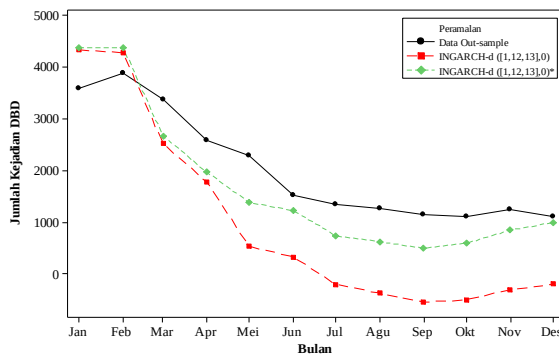
Selanjutnya dilakukan estimasi paramater model dengan menyertakan empat efek intervensi pada model. Intervensi pertama adalah *outlier* pada observasi ke-37 sampai dengan 39, yaitu data pada bulan Januari 2010 sampai dengan Maret 2010. Intervensi kedua padalah *outlier* pada observasi ke-97 dan 98, yaitu data pada bulan Januari 2015 dan Februari 2015. Intervensi ketiga adalah *outlier* pada observasi ke-49 sampai dengan 72, yaitu data bulan Januari 2011 sampai dengan Desember 2012. Intervensi keempat adalah *outlier* pada observasi ke- 85 sampai dengan 96, yaitu data bulan Januari 2014 sampai dengan Desember 2014. Model INGARCH- d dengan menyertakan efek intervensi ini kemudian akan dinyatakan dnegan INGARCH- d^* . Hasil estimasi parameter model dengan menyertakan efek intervensi ditunjukkan oleh Tabel 4.12.

Tabel 4.12 Hasil Estimasi Paramater Model INGARCH- d^*

Model INGARCH	Parameter	Estimasi	Std. Error	CI (lower)	CI (upper)
([1,12,13],0)*	β_0	3080,0	56,638	2970,0	3190,0
	β_1	0,0193	0,0099	-9,61e-05	0,0387
	β_{12}	1,85e-10	0,0092	-0,0180	0,0180
	β_{13}	8,92e-11	0,0108	-0,0213	0,0213
	$\omega_{37 \leq t \leq 39}$	0,0892	34,959	-68,400	68,600
	$\omega_{97;98}$	0,8050	43,170	-83,800	85,400
	$\omega_{49 \leq t \leq 72}$	0,0243	13,828	-27,100	27,100
	$\omega_{85 \leq t \leq 96}$	0,0852	18,041	-35,300	35,400

Hasil estimasi paramater dari model $\text{INGARCH}-d$ $([1,12,13],0)^*$, yaitu model INGARCH dengan fungsi intervensi, *differencing* pertama, dan *differencing* musiman pertama pada Tabel 4.12 akan digunakan untuk melakukan peramalan pada jumlah kejadian DBD di Jawa Timur pada periode 12 bulan ke depan. Secara visual, hasil peramalan jumlah kejadian DBD di Jawa Timur pada periode 12 bulan ke depan ditunjukkan oleh *time series* plot pada Gambar 4.13.

Berdasarkan plot *time series* pada Gambar 4.11, terlihat bahwa pola data hasil peramalan model INGARCH dengan intervensi untuk periode dua belas bulan ke depan sudah mengikuti pola data *out-sample*. Hasil ini mengindikasikan bahwa model yang didapatkan telah sesuai. Secara visual, jika dibandingkan dengan hasil peramalan model $\text{INGARCH}-d$ $([1,12,13],0)$ dengan model $\text{INGARCH}-d$ $([1,12,13],0)^*$, maka model $\text{INGARCH}-d$ $([1,12,13],0)^*$ menunjukkan bahwa hasil peramalan tidak memuat nilai yang negatif, artinya model yang didapatkan telah mampu mengatasi adanya *outlier* pada data.



Gambar 4.11 Perbandingan Data *Out-Sample* dengan Hasil Peramalan Model $\text{INGARCH}-d$ dan Model $\text{INGARCH}-d^*$

Model $\text{INGARCH}-d$ $([1,12,13],0)^*$ dapat dituliskan sebagai berikut.

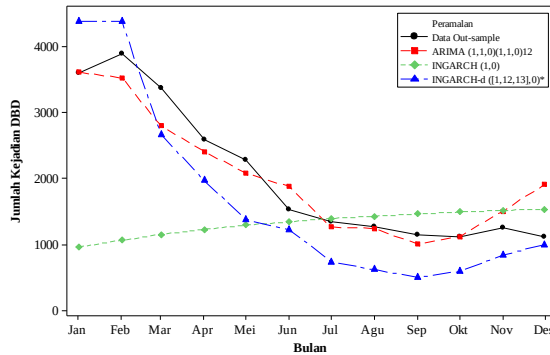
$$\lambda_t^{d*} = 3080 + 0,0193Y_{t-1} + 0,000000000185Y_{t-12} + 0,0000000000892Y_{t-13} \\ + 0,0892(37 \leq t \leq 39) + 0,805(t = 97; 98) + 0,0243(49 \leq t \leq 72) \\ + 0,0852(85 \leq t \leq 96), \quad t = 1, 2, \dots, 108$$

dimana Y_t pada periode t diberikan oleh $Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t^{d*})$. Model INGARCH- d $([1,12,13],0)^*$ menunjukkan bahwa jumlah kejadian DBD di Jawa Timur pada waktu ke t berkaitan dengan jumlah kejadian DBD pada waktu ke $t-1$, $t-12$, dan $t-13$ dengan empat fungsi intervensi pada waktu $37 \leq t \leq 39$, $t = 97; 98$, $49 \leq t \leq 72$, dan $85 \leq t \leq 96$ untuk $t = 1, 2, \dots, 108$. Hasil peramalan dari model adalah nilai median pada parameter λ_t^{d*} . Tahap selanjutnya adalah melakukan perbandingan model INGARCH dengan model ARIMA untuk memperoleh hasil peramalan terbaik.

4.5 Perbandingan Pemodelan ARIMA dan INGARCH

Perbandingan dilakukan terhadap pemodelan ARIMA terbaik, yaitu ARIMA $(1,1,0)(1,1,0)^{12}$, model INGARCH terbaik yaitu INGARCH $(1,0)$ dengan hanya menyertakan parameter yang signifikan, dan INGARCH- d $([1,12,13],0)^*$ dengan menyertakan semua parameter yang tidak signifikan. Secara visual, hasil peramalan ketiga model ditunjukkan oleh plot *time series* pada Gambar 4.12.

Berdasarkan Gambar 4.14, secara visual terlihat bahwa model ARIMA memiliki kemampuan peramalan jumlah kejadian DBD di Jawa Timur yang lebih baik dibandingkan dengan model INGARCH. Hal ini dapat dilihat pada plot *time series* model ARIMA yang lebih mendekati pola data asli jumlah kejadian DBD di Jawa Timur periode Januari 2016 sampai dengan Desember 2016. Model INGARCH- d^* juga mendekati pola data asli, sehingga diindikasikan model ini telah sesuai dan baik untuk digunakan peramalan jumlah kejadian DBD di Jawa Timur.



Gambar 4.12 Plot *Time Series* Perbandingan Hasil Peramalan ARIMA, INGARCH, dan Data Asli

Sementara, model INGARCH yang hanya menyertakan parameter signifikan sangat jauh dari pola data *out sample* atau data asli, sehingga diindikasikan model ini tidak cocok digunakan untuk peramalan jumlah kejadian DBD di Jawa Timur. Di samping itu, ketiga model juga dapat dibandingkan berdasarkan kriteria kebaikan model menggunakan RMSE dan MAPE. Nilai RMSE dan MAPE model untuk peramalan periode dua belas bulan ke depan yakni bulan Januari 2016 sampai dengan Desember 2016 ditampilkan dalam Tabel 4.13.

Tabel 4.13 Perbandingan Kriteria Kebaikan Model untuk *Out-sample* pada Model ARIMA Terbaik dan INGARCH Terbaik

Model	RMSE	MAPE (%)
ARIMA (1,1,0)(1,1,0) ¹²	338,89	14,74
INGARCH (1,0)	1387,73	37,97
INGARCH –d ([1,12,13],0)*	596,59	31,71

Model ARIMA terbaik memiliki nilai MAPE sebesar 14,74% dan RMSE sebesar 338,89. Sedangkan model INGARCH terbaik yang menyertakan parameter signifikan memiliki nilai

MAPE sebesar 37,97% dan nilai RMSE sebesar 1387,73. Adapun model INGARCH- d^* yakni model INGARCH- d dengan intervensi dan menyertakan parameter tidak signifikan memiliki nilai MAPE sebesar 37,97% dan RMSE sebesar 596,59. Berdasarkan kriteria kebaikan model yaitu MAPE dan RMSE yang diterapkan pada data *out-sample* untuk periode dua belas bulan ke depan yakni bulan Januari 2016 sampai dengan Desember 2016, maka dapat diketahui MAPE dan RMSE model ARIMA lebih kecil daripada model INGARCH- d^* , sehingga model ARIMA memiliki kemampuan peramalan jumlah kejadian DBD di Jawa Timur yang lebih baik dibandingkan model INGARCH- d^* . Hasil peramalan periode bulan Januari 2016 sampai dengan Desember 2016 dari ketiga model yaitu ARIMA (1,1,0)(1,1,0)¹², INGARCH (1,0), dan model INGARCH- d ([1,12,13],0)* ditunjukkan oleh Tabel 4.14.

Tabel 4.14 Peramalan *Out-sample* oleh Model ARIMA dan INGARCH Tahun 2016

Periode	Jumlah Kejadian DBD (Orang)			
	Data Asli	ARIMA (1,1,0)(1,1,0) ¹²	INGARCH (1,0)	INGARCH- d ([1,12,13],0)*
Januari	3590	3606	964	4372
Februari	3884	3519	1064	4369
Maret	3368	2796	1152	2664
April	2581	2406	1229	1964
Mei	2278	2082	1292	1378
Juni	1523	1874	1345	1223
Juli	1346	1262	1393	740
Agustus	1264	1243	1428	619
September	1147	1011	1462	501
Oktober	1116	1119	1492	593
November	1249	1499	1517	845
Desember	1115	1913	1536	995

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Analisis yang telah dilakukan terhadap jumlah kejadian DBD di Jawa Timur memberikan beberapa kesimpulan sebagai berikut.

1. Jumlah kejadian DBD di Jawa Timur sepanjang tahun 2007 hingga 2016 cenderung fluktuatif atau naik turun dan terdapat pola inkonsistensi. Adapun rata-rata jumlah kejadian DBD tertinggi terjadi pada bulan Januari yaitu sebesar 2933 kasus, sementara jumlah kejadian paling sedikit terjadi pada bulan September yaitu sebesar 552. Jumlah kejadian DBD sangat tinggi pada bulan Januari, Februari dan Maret disebabkan peralihan musim dari musim penghujan menuju musim kemarau. Adapun jumlah kejadian DBD di Jawa Timur pada bulan Maret hingga Desember cenderung sedikit. Hal ini dipengaruhi oleh siklus perkembangbiakan vektor nyamuk *Aedes aegypti* pada bulan-bulan tertentu di sepanjang tahunnya.
2. Model ARIMA memenuhi asumsi *white noise* dan residual berdistribusi normal. Model ARIMA terbaik adalah $(1,1,0)(1,1,0)^{12}$ dengan nilai MAPE sebesar 14,74% dan RMSE sebesar 338,897. Model ARIMA $(1,1,0)(1,1,0)^{12}$ menunjukkan bahwa jumlah kejadian DBD di Jawa Timur pada waktu ke t berkaitan dengan jumlah kejadian DBD pada waktu ke $t-1$, $t-12$, dan $t-13$ dengan model sebagai berikut,

$$Z_t = 0,271Z_{t-1} + 0,5828Z_{t-12} - 0,15794Z_{t-13} + a_t$$

3. Model INGARCH terbaik yang terpilih adalah INGARCH- d $([1,12,13],0)$ dengan semua parameter yang signifikan. Fungsi distribusinya adalah distribusi Poisson dengan fungsi *link identity*, serta dilakukan *differencing* reguler pertama dan *differencing* musiman pertama. Hasil peramalan pada model INGARCH- d $([1,12,13],0)$ memuat nilai negatif yang

menandakan adanya *outlier* pada data. Untuk mengatasi hal tersebut, dilakukan analisis intervensi sehingga diperoleh nilai hasil peramalan model INGARCH- d $([1,12,13],0)^*$ yang positif dengan nilai MAPE sebesar 31,71% dan RMSE sebesar 596,59. Y_t pada periode t berdistribusi Poisson dengan $Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t^{d*})$. Model INGARCH- d $([1,12,13],0)^*$ mampu mengatasi adanya *outlier* pada data dan mampu melakukan peramalan dengan baik. Model INGARCH- d $([1,12,13],0)^*$ menunjukkan bahwa jumlah kejadian DBD di Jawa Timur pada waktu ke t terkait dengan jumlah kejadian DBD pada waktu ke $t-1$, $t-12$, dan $t-13$ dengan empat fungsi intervensi pada waktu $37 \leq t \leq 39$, $t=97;98$, $49 \leq t \leq 72$, dan $85 \leq t \leq 96$ untuk $85 \leq t \leq 96$. Model INGARCH- d $([1,12,13],0)^*$ dapat dituliskan sebagai berikut.

$$\lambda_t^{d*} = 3080 + 0,0193Y_{t-1} + 0,000000000185Y_{t-12} + 0,0000000000892Y_{t-13} \\ + 0,0892(37 \leq t \leq 39) + 0,805(t=97;98) + 0,0243(49 \leq t \leq 72) \\ + 0,0852(85 \leq t \leq 96), \quad t = 1, 2, \dots, 108$$

4. Peramalan jumlah kejadian DBD di Jawa Timur dapat dilakukan dengan lebih baik jika menggunakan model ARIMA. Hal ini ditinjau dari nilai MAPE dan RMSE yang didapatkan lebih kecil dari nilai MAPE dan RMSE model INGARCH-d* untuk peramalan periode 12 bulan ke depan. Analisis intervensi pada model INGARCH-d dapat membantu mengatasi *outlier* pada data.

5.2 Saran

Untuk melakukan peramalan jumlah kejadian DBD di Jawa Timur dapat dilakukan dengan menggunakan metode ARIMA, sehingga pihak Dinas Kesehatan Provinsi Jawa Timur dapat melakukan tindakan preventif yang tepat guna menekan jumlah kasus DBD di masyarakat Jawa Timur. Data yang digunakan dalam penelitian ini merupakan data yang tidak stasioner. Maka paket statistik untuk melakukan pemodelan INGARCH pada data yang tidak stasioner dapat dikembangkan.

DAFTAR PUSTAKA

- Armstrong, J.S., (2007). "Significance Tests Harm Progress in Forecasting". *International Journal of Forecasting*, 23, 321–327.
- Asrirawan, Suhartono, (2015). "The Estimation of Generalized Seasonal Autoregressive Integrated Moving Average (GSARIMA) Models Using Bayesian Approach for Forecasting the Number of Dengue Hemorrhagic Fever (DHF) Patients". International Conference on Statistics, Mathematics, Teaching, and Research.
- Box, G.E.P., Tiao, G.C. (1975). "Intervention Analysis with Applications to Economic and Environmental Problems". *Journal of the American Statistical Association*, 70 (349), 70–79. doi:10.2307/2285379.
- Christou, V., Fokianos, K. (2015). "On Count Time Series Prediction". *Journal of Statistical Computation and Simulation*, 85 (2), 357–373. doi:10.1080/00949655.2013.823612.
- Cryer, J.D., dan Chan, K.S. (2008). "Time Series Analysis: With Application in R: Second Edition". USA: Spinger Science dan Business Media, LLC.
- Dinkes. (2009). *Vektor Demam Berdarah dan cara Pengendalian*. Jakarta: Direktorat Jendral Pengendalian Penyakit dan Penyehatan Lingkungan.
- Doukhan, P., Fokianos, K., Tjøstheim, D. (2012). "On Weak Dependence Conditions for Poisson Autoregressions". *Statistics & Probability Letters*, 82 (5), 942–948. doi:10.1016/j.spl.2012.01.015.
- Ferland, R., Latour A., Oraichi, D. (2006). "Integer-Valued GARCH Process". *Journal of Time Series Analysis*, 27 (6), 923–942. doi:10.1111/j.1467-9892.2006.00496.x.
- Fokianos, K., Fried, R. (2010). "Interventions in GARCH Processes". *Journal of Time Series Analysis*, 210-225. Doi : 10.1111/j.1467-9892.2010.00657.x.

- Fokianos, K., Fried, R. (2012). "Interventions in Log-Linear Poisson Autoregression". *Statistical Modelling*, 12 (4), 299–322. doi:10.1177/1471082X1201200401.
- Gneiting, T., Balabdaoui, F., Raftery, A.E. (2007) "Probabilistic Forecasts, Calibration and Sharpness". *Journal of the Royal Statistical Society B*, 69 (2), 243–268. doi:10.1111/j. 1467-9868.2007.00587.x.
- Heinen, A. (2003). "Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model". *CORE Discussion Paper*, 62. doi:10.2139/ssrn.1117187.
- Hyndman, R.J., & Koehler, A.B. (2006). "Another Look at Measures of Forecast Accuracy". *International Journal of Forecasting*, 22 (4), 679-688.
- Kejadian Luar Biasa Demam Berdarah di Jawa Timur. KOMPAS.COM. 26 Januari 2015.
<<http://lifestyle.kompas.com/read/2015/01/26/102000423/Kejadian.Luar.Biasa.Demam.Berdarah.di.Jawa.Timur>>
- Kementerian Kesehatan Republik Indonesia. (2010). *Demam Berdarah Dengue di Indonesia Tahun 1968-2009*. Jakarta: Kementerian Kesehatan RI.
- Kementerian Kesehatan Republik Indonesia. (2016). *Situasi DBD di Indonesia*. Jakarta: Kementerian Kesehatan RI.
- Kostenko, A.V., & Hyndman, R.J. (2008). "Forecasting Without Significance Tests?". Monash University (Australia): Department of Econometrics and Business Statistics.
- Liboschik, T., Fokianos, K., Fried, R. (2016). "tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models". *Vignette of R Package tscount Version 1.3.0*, doi:10.17877/DE290R-7239.
- McCullagh, P., & Nelder, J.A. (1989). "Generalized Linier Model Second Edition". New York: Chapman & Hall
- Pangestu, S. (1986). "Forecasting Konsep dan Aplikasi". Yogyakarta: BPPE UGM.

- Satari, H.I., dan Meiliasari, M. (2004). *Demam Berdarah*. Jakarta : Puspa Swara, Anggota IKAPI.
- Soedarto. (2012). *Demam Berdarah Dengue*. Jakarta: CV. Sagung Seto.
- Supranto, J. (1984). "Ekonomic: Buku Dua". Bogor : Ghalia Indonesia
- Wei, W.W.S. (1994). "*Time Series Analysis: Univariate and Multivariate Method*". 2nd Edition. New York : Pearson Education, Inc.
- Widagdo. (2012). *Masalah dan Tata Laksana Penyakit Anak dengan Demam*. Jakarta : CV. Sagung Seto.
- Winahju, W.S., Mukarromah, A., dan Rahayu, D. (2012). "*Pemodelan Pengaruh Iklim Terhadap Angka Kejadian Demam Berdarah Dengue di Surabaya*". Jurnal Sains dan Seni ITS Vol.1, ISSN: 2301928X
- World Health Organization. (2001). *Prevention and Control of Dengue and Dengue Haemorrhagic Fever : Comprehensive Guidelines*. Salmiyatun (Ed.) *Pencegahan dan Pengendalian Dengue dan Demam Berdarah Dengue : Panduan Lengkap*. Jakarta : Penerbit Buku Kedokteran EGC.
- Zeger, S.L., Qaqish, B. (1988). "*Markov Regression Models for Time Series: A Quasi-Likelihood Approach*" *Biometrics*. Vol. 44, No. 4 (Dec., 1988), pp. 1019-1031.

Halaman ini sengaja dikosongkan

LAMPIRAN

A. Output Minitab Model ARIMA

1. Model ARIMA (1,1,0)(1,1,0)¹²

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0,2710	0,1003	2,70	0,008
SAR	12	-0,5828	0,0875	-6,66	0,000

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 108, after differencing 95

Residuals: SS = 5,36849 (backforecasts excluded)
MS = 0,05773 DF = 93

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	4,4	14,2	24,2	33,6
DF	10	22	34	46
P-Value	0,927	0,893	0,894	0,914

Forecasts from period 108

Period	Forecast	95% Limits	
		Lower	Upper
109	8,19040	7,71939	8,66141
110	8,16572	7,40400	8,92744
111	7,93609	6,94553	8,92665
112	7,78574	6,60500	8,96648
113	7,64126	6,29577	8,98676
114	7,53605	6,04358	9,02851
115	7,14019	5,51391	8,76648
116	7,12502	5,37510	8,87493
117	6,91846	5,05309	8,78383
118	7,02040	5,04631	8,99450
119	7,31258	5,23545	9,38972
120	7,55641	5,38111	9,73170

2. Model ARIMA (1,1,0)(0,1,1)¹²

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0,2089	0,0991	2,11	0,038
SMA	12	0,8456	0,0832	10,17	0,000

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 108, after differencing 95

Residuals: SS = 4,35992 (backforecasts excluded)
MS = 0,04688 DF = 93

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	6,8	17,1	27,3	38,7
DF	10	22	34	46
P-Value	0,745	0,756	0,787	0,769

Forecasts from period 108

Period	Forecast	95% Limits	
		Lower	Upper
109	7,73425	7,30978	8,15871
110	7,60552	6,93957	8,27147
111	7,41026	6,55810	8,26241
112	7,23377	6,22730	8,24024
113	7,10680	5,96633	8,24726
114	6,85267	5,59231	8,11304
115	6,56047	5,19064	7,93029
116	6,32616	4,85499	7,79733
117	6,14742	4,58146	7,71339
118	6,31662	4,66128	7,97197
119	6,62576	4,88562	8,36589
120	6,81591	4,99493	8,63689

3. Model ARIMA (0,1,1)(0,1,1)¹²

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	-0,1925	0,1023	-1,88	0,063
SMA	12	0,8057	0,0937	8,60	0,000

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 108, after differencing 95

Residuals: SS = 4,49950 (backforecasts excluded)
MS = 0,04838 DF = 93

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	7,2	18,1	28,4	41,1
DF	10	22	34	46
P-Value	0,710	0,702	0,739	0,678

Forecasts from period 108

Period	Forecast	95% Limits	
		Lower	Upper
109	7,75665	7,32545	8,18786
110	7,62749	6,95640	8,29858
111	7,42030	6,57486	8,26574
112	7,24099	6,25145	8,23053
113	7,10204	5,98686	8,21721
114	6,85308	5,62506	8,08109
115	6,55121	5,21988	7,88254
116	6,31793	4,89074	7,74511
117	6,13578	4,61878	7,65278
118	6,29970	4,69792	7,90148
119	6,60228	4,91999	8,28458
120	6,75679	4,99766	8,51592

4. Model ARIMA (1,1,0)(1,1,0)¹²

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
SAR	12	-0,5953	0,0870	-6,85	0,000
MA	1	-0,2372	0,1006	-2,36	0,021

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 108, after differencing 95

Residuals: SS = 5,41634 (backforecasts excluded)
MS = 0,05824 DF = 93

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	5,5	16,4	28,6	38,3
DF	10	22	34	46
P-Value	0,853	0,796	0,731	0,783

Forecasts from period 108

Period	Forecast	95% Limits	
		Lower	Upper
109	8,17409	7,70099	8,64719
110	8,13233	7,37972	8,88493
111	7,90354	6,95013	8,85695
112	7,75522	6,63648	8,87396
113	7,61487	6,35226	8,87747
114	7,51023	6,11856	8,90190
115	7,11704	5,60730	8,62679
116	7,10643	5,48719	8,72566
117	6,90174	5,17997	8,62351
118	7,00328	5,18474	8,82182
119	7,29408	5,38367	9,20449
120	7,53998	5,54192	9,53804

B. Output R Model INGARCH

1. Model INGARCH Tanpa *Differencing*

```
Call:
tsglm(ts = dbd, model = list(past_obs = 1))
```

Coefficients:

	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	248.600	4.626	239.532	257.668
beta_1	0.849	0.004	0.841	0.857

Standard errors and confidence intervals (level = 95 %) obtained by normal approximation.

Link function: identity
 Distribution family: poisson
 Number of coefficients: 2
 Log-likelihood: -13881.56
 AIC: 27767.13
 BIC: 27772.49
 QIC: 27766.77

```
Call:
tsglm(ts = dbd, model = list(past_obs = 1, past_mean = 12))
```

Coefficients:

	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	4.90e+02	7.31722	475.90086	5.05e+02
beta_1	6.86e-01	0.00413	0.67793	6.94e-01
alpha_12	1.25e-10	0.00488	-0.00956	9.56e-03

Standard errors and confidence intervals (level = 95 %) obtained by normal approximation.

Link function: identity
 Distribution family: poisson
 Number of coefficients: 3
 Log-likelihood: -15745.61
 AIC: 31497.21
 BIC: 31505.26
 QIC: 31496.67

```
Call:
tsglm(ts = dbd, model = list(past_obs = c(1, 13)))
```

Coefficients:

	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	2.18e+02	5.28468	207.74085	2.28e+02
beta_1	8.73e-01	0.00419	0.86509	8.82e-01
beta_13	1.34e-09	0.00315	-0.00618	6.18e-03

Standard errors and confidence intervals (level = 95 %) obtained by normal approximation.

Model INGARCH Tanpa *Differencing* (Lanjutan)

```
Link function: identity
Distribution family: poisson
Number of coefficients: 3
Log-likelihood: -13785.03
AIC: 27576.06
BIC: 27584.11
QIC: 27575.34
```

2. Model INGARCH (1,0)

```

$pred
      Jan      Feb      Mar      Apr      May      Jun
2016 963.6263 1066.9112 1154.6208 1229.1038 1292.3547 1346.0673
      Jul      Aug      Sep      Oct      Nov      Dec
2016 1391.6801 1430.4144 1463.3076 1491.2405 1514.9611 1535.1047

$interval
      lower upper
Jan 2016    881  1053
Feb 2016    951  1182
Mar 2016   1017  1293
Apr 2016   1079  1371
May 2016   1107  1443
Jun 2016   1171  1529
Jul 2016   1213  1598
Aug 2016   1256  1635
Sep 2016   1273  1651
Oct 2016   1293  1688
Nov 2016   1341  1704
Dec 2016   1343  1735

$level
[1] 0.95

$global
[1] TRUE

$type
[1] "quantiles"

$method
[1] "bootstrap"

$B
[1] 1000

$estim
[1] "ignore"

$B_estim
[1] 0

$warning_messages
character(0)

$median
      Jan      Feb      Mar      Apr      May      Jun
2016 964.0 1064.0 1151.5 1228.5 1292.0 1345.0
      Jul      Aug      Sep      Oct      Nov      Dec
2016 1392.5 1427.5 1462.0 1491.5 1516.5 1536.0

```

3. Model INGARCH ([1,13],0)

```

$pred
      Jan      Feb      Mar      Apr      May      Jun
2016 953.4138 1050.7109 1135.6802 1209.8834 1274.6848 1331.2756
      Jul      Aug      Sep      Oct      Nov      Dec
2016 1380.6961 1423.8549 1461.5452 1494.4600 1523.2044 1548.3067

$interval
      lower upper
Jan 2016    863   1037
Feb 2016    926   1164
Mar 2016    998   1262
Apr 2016   1069   1366
May 2016   1111   1441
Jun 2016   1153   1504
Jul 2016   1203   1567
Aug 2016   1243   1627
Sep 2016   1279   1658
Oct 2016   1285   1727
Nov 2016   1315   1758
Dec 2016   1350   1794

$level
[1] 0.95

$global
[1] TRUE

$type
[1] "quantiles"

$method
[1] "bootstrap"

$B
[1] 1000

$estim
[1] "ignore"

$B_estim
[1] 0

$warning_messages
character(0)

$median
      Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep
2016 952.0 1049.0 1133.0 1207.0 1274.0 1330.0 1380.0 1423.0 1458.0
      Oct      Nov      Dec
2016 1492.5 1522.0 1547.0

```


4. Output R Hasil Estimasi Parameter dan Peramalan Model INGARCH-d ([1,12,13],0)

```
Call:
tsglm(ts = DBD, model = list(past_obs = c(1, 12, 13)))

Coefficients:
              Estimate Std. Error CI(lower) CI(upper)
(Intercept)  3.09e+03   54.68306  2982.1843  3.20e+03
beta_1       2.75e-06   0.00908   -0.0178   1.78e-02
beta_12      2.15e-11   0.00904   -0.0177   1.77e-02
beta_13      9.01e-10   0.01073   -0.0210   2.10e-02
Standard errors and confidence intervals (level = 95 %) obtained
by normal approximation.

Link function: identity
Distribution family: poisson
Number of coefficients: 4
Log-likelihood: -11044.71
AIC: 22097.42
BIC: 22108.15
QIC: 22097.42

$pred
      Jan      Feb      Mar      Apr      May      Jun      Jul
Aug 2016 3089.369 3089.370 3089.370 3089.370 3089.370 3089.370 3089.370
3089.370
      Sep      Oct      Nov      Dec
2016 3089.370 3089.370 3089.370 3089.370

$interval
      lower upper
Jan 2016  2937  3258
Feb 2016  2924  3234
Mar 2016  2953  3248
Apr 2016  2934  3237
May 2016  2943  3235
Jun 2016  2941  3268
Jul 2016  2922  3254
Aug 2016  2908  3254
Sep 2016  2944  3243
Oct 2016  2939  3247
Nov 2016  2956  3237
Dec 2016  2936  3239

$level
[1] 0.95

$global
[1] TRUE

$type
[1] "quantiles"

$method
[1] "bootstrap"

$B
[1] 1000

$estim
[1] "ignore"
```

Output R Hasil Estimasi Parameter dan Peramalan Model INGARCH-d ([1,12,13],0) (Lanjutan)

```
$B_estim
[1] 0

$warning_messages
character(0)

$median
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2016 3092 3085 3090 3092 3085 3090 3091 3091 3089 3087 3088 3094
```

5. Output R Hasil Estimasi Paramater dan Peramalan Model INGARCH-d $[[1,12,13],0]*/$ Model INGARCH dengan Intervensi

```
Call:
tsglm(ts = dbd_new_, model = list(past_obs = c(1, 12, 13)), xreg =
dbdinterventions,
distr = "poisson")

Coefficients:
              Estimate Std. Error CI(lower) CI(upper)
(Intercept)  3.08e+03   56.63790   2.97e+03   3.19e+03
beta_1       1.93e-02    0.00991  -9.61e-05   3.87e-02
beta_12      1.85e-10    0.00917  -1.80e-02   1.80e-02
beta_13      8.92e-11    0.01086  -2.13e-02   2.13e-02
Interv_1     8.61e-02   34.95923  -6.84e+01   6.86e+01
Interv_2     8.05e-01   43.17037  -8.38e+01   8.54e+01
Interv_3     2.43e-02   13.82885  -2.71e+01   2.71e+01
Interv_4     8.52e-02   18.04131  -3.53e+01   3.54e+01
Standard errors and confidence intervals (level = 95 %) obtained
by normal approximation.

Link function: identity
Distribution family: poisson
Number of coefficients: 8
Log-likelihood: -10931.91
AIC: 21879.82
BIC: 21900.34
QIC: 21880.78

$pred
      Jan      Feb      Mar      Apr      May      Jun      Jul
2016 3135.900 3139.974 3140.053 3140.054 3140.055 3140.055 3140.055
      Aug      Sep      Oct      Nov      Dec
2016 3140.055 3140.055 3140.055 3140.055 3140.055

$interval
      lower upper
Jan 2016  2988  3281
Feb 2016  2997  3296
Mar 2016  2991  3296
Apr 2016  2989  3317
May 2016  2979  3293
Jun 2016  2983  3277
Jul 2016  2969  3310
Aug 2016  2984  3297
Sep 2016  2986  3286
Oct 2016  2988  3307
Nov 2016  2982  3292
Dec 2016  2984  3305

$level
[1] 0.95

$global
[1] TRUE

$type
[1] "quantiles"

$method
```

Output R Hasil Estimasi Paramater dan Peramalan Model INGARCH-d ([1,12,13],0)/Model INGARCH dengan Intervensi (Lanjutan)*

```
[1] "bootstrap"

$B
[1] 1000

$estim
[1] "ignore"

$B_estim
[1] 0

$warning_messages
character(0)

$median
      Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep
2016 3135.5 3138.0 3139.0 3144.0 3135.5 3137.0 3142.0 3139.0 3141.0
      Oct      Nov      Dec
2016 3139.0 3139.5 3140.0
```

BIODATA PENULIS



Penulis dilahirkan di pulau Kangean, Kabupaten Sumenep pada tanggal 13 September 1992, merupakan anak pertama dari lima bersaudara. Penulis telah menempuh pendidikan formal yaitu TK Al-Qur'an Islamiyah Arjasa Sumenep; SDN Pandeman 1 Arjasa Sumenep, dan SMA Muhammadiyah 1 Sumenep. Semasa kuliah, penulis aktif di berbagai lembaga kampus, di antaranya Lembaga Dakwah Jurusan (LDJ) Statistika ITS, Lembaga Dakwah Kampus Jama'ah Masjid Manarul Ilmi (JMMI) ITS, dan Unit Kegiatan Mahasiswa Lembaga Pers Mahasiswa (LPM) 1-0. Penulis juga aktif mengirimkan tulisan berupa opini dan surat pembaca ke berbagai media baik cetak maupun online. Di antaranya pernah dimuat di islampos.com, suarasurabaya.net, dll. Jika ingin kenal lebih dekat dengan penulis, bisa menghubungi alamat email zhakie.zz@gmail.com.

