



# Peramalan Jumlah Penumpang Kereta Api Fajar Utama Yogya Jurusan Yogyakarta - Pasar Senen Menggunakan Variasi Kalender *ARIMAX* (Studi Kasus di Kantor Pusat PT. Kereta Api Indonesia)

Yoan Sofyana Noer\*, Gungum Darmawan, Resa Septiani Pontoh

Departemen Statistika, FMIPA Universitas Padjadjaran

\*E-mail: yoan.sn@gmail.com

## Abstrak

Tingginya lonjakan jumlah penumpang kereta api pada hari Raya Idul Fitri menyebabkan pola musiman yang bergerak, salah satunya adalah kereta api Fajar Utama Yogya. Pergeseran maju hari raya Idul Fitri tersebut merupakan salah satu efek variasi kalender. Pada penelitian ini efek variasi kalender digunakan sebagai variabel tambahan yang berupa variabel dummy, sehingga disebut dengan model *ARIMAX*. Model terbaik dari *ARIMAX* adalah model dengan MAPE dan RMSE yang paling kecil, yang kemudian akan digunakan untuk mendapatkan hasil peramalan jumlah penumpang kereta api Fajar Utama Yogya ini.

Kata Kunci: jumlah penumpang kereta api, efek variasi kalender, variabel, dummy, model *ARIMAX*

## 1. Pendahuluan

Di Indonesia jenis angkutan darat yang bisa digunakan salah satunya adalah kereta api. Kereta api memiliki banyak kelebihan daripada jenis angkutan darat lainnya. Maka tidak heran jika semakin hari kereta api semakin menarik perhatian masyarakat karena penggunaannya lebih praktis dibandingkan dengan angkutan lain, dapat mengangkut penumpang lebih banyak, harganya lebih ekonomis, menghemat waktu karena terhindar dari kemacetan, serta tingkat keselamatan yang cukup tinggi (Widhianti, 2013).

Berdasarkan hasil rekapan kantor pusat PT KAI, jumlah penumpang kereta api setiap tahunnya mengalami kenaikan baik di pulau Jawa maupun Sumatera. Jumlah penumpang kereta api yang mengalami kenaikan disebabkan oleh banyak hal, salah satunya yaitu disebabkan oleh *event* di hari khusus, seperti pada saat hari raya Idul Fitri. Hari raya Idul Fitri ini merupakan salah satu efek variasi kalender *holiday variation*. Kenaikan jumlah penumpang menjelang hari raya Idul Fitri ini biasanya terjadi pada rentang waktu dua minggu sebelum terjadinya hari raya Idul Fitri atau bahkan 30 hari selama bulan Ramadhan.

Kereta Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen adalah salah satu *unit* kereta api yang mengalami kenaikan (lonjakan) efek hari raya Idul Fitri. Lonjakan ini seringkali menjadi permasalahan dan kendala yang dihadapi PT. Kereta Api (Persero). Hal tersebut terjadi karena keterbatasan kapasitas angkut yang disediakan tidak seimbang. Yakni jumlah permintaan angkutan penumpang kereta api jauh lebih besar dibandingkan dengan kapasitas tempat duduk yang disediakan.

Penelitian ini akan mencoba membentuk model pada data *time series* yang mengalami pelonjakan pada saat hari raya Idul Fitri, yang dilakukan dengan metode peramalan yang didasarkan pada *calendar variation effect* atau efek variasi kalender dengan model *ARIMAX*. Model *ARIMAX* ini digunakan karena dapat menangkap pergeseran hari raya Idul Fitri yang terjadi pada variasi kalender, sedangkan model lainnya seperti *ARIMA* tidak dapat menangkap pergeseran tersebut. Model terbaik dari *ARIMAX* akan digunakan untuk meramalkan data jumlah penumpang kereta api Fajar Utama Yogya jurusan Yogyakarta - Pasar Senen untuk periode yang akan datang.

Bentuk lain dari metode variasi kalender sekarang ini banyak digunakan dalam penelitian, salah satunya adalah model *ARIMAX*. Model *ARIMAX* merupakan perluasan dari model *ARIMA*. Model ini menggunakan variabel tambahan yang diperoleh dari variasi kalender yang dijadikan sebagai variabel dummy pada model. Dalam model ini, faktor-faktor yang mempengaruhi variabel respon  $y$  pada waktu  $t$  dipengaruhi tidak hanya oleh fungsi variabel  $y$  dalam waktu, tetapi juga oleh variabel-variabel predictor lain pada waktu ke- $t$  (Rosadi, 2012).

Penelitian mengenai metode peramalan variasi kalender *ARIMAX* sekarang ini sudah banyak dilakukan diantaranya adalah Afriani (2013) dan Izza (2014) yang menghasilkan keakurasian data ramalan yang baik dan memberikan kesimpulan bahwa hasil peramalan model *ARIMAX* merupakan metode terbaik yang digunakan jika data deret waktu dipengaruhi pergeseran kalender hijriah.



Dengan beberapa penelitian sebelumnya yang sudah disebutkan diatas, maka metode variasi kalender model *ARIMAX* cocok digunakan untuk melihat pola data deret waktu dipengaruhi efek variasi kalender. Oleh karena itu dalam proses memodelkan dan meramalkan jumlah penumpang kereta api Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen, penulis akan menggunakan metode peramalan analisis variasi kalender dengan model *ARIMAX*.

## 2. Metode

Metode yang digunakan pada penelitian ini adalah metode efek variasi kalender model *ARIMAX*.

### 2.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Divisi Komersial Pemasaran Penumpang Kantor Pusat PT Kereta Api Indonesia (KAI) Persero. Data jumlah penumpang kereta api Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen yang digunakan adalah dari bulan Januari tahun 2011 sampai dengan bulan Mei tahun 2016.

Peneliti akan membagi menjadi dua yaitu data untuk *in-sample* dan data untuk *out-sample*. Data *in-sample* sebanyak 60 observasi yaitu data jumlah penumpang kereta api Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen bulan Januari tahun 2011 sampai dengan bulan Desember tahun 2015. Dan data *out-sample* sebanyak 5 observasi yaitu data dari bulan Januari sampai dengan bulan Mei tahun 2016.

Penelitian ini memfokuskan masalah pada variabel jumlah penumpang kereta api Bisnis Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen dan variabel eksogen yaitu variabel *dummy* yang diperoleh dari efek variasi kalender.

### 2.2 Metode Analisis Data

Model *ARIMAX* dengan variabel *dummy* efek variasi kalender ditulis:

$$y_t = \beta_1 L_{1,t} + \beta_2 L_{2,t} + \beta_3 L_{3,t} + \frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t \quad (3.1)$$

Adapun model *ARIMAX* dengan pengaruh musiman adalah sebagai berikut.

$$y_t = \beta_1 L_{1,t} + \beta_2 L_{2,t} + \beta_3 L_{3,t} + \frac{\theta_q(B)\theta_Q(B^S)}{\phi_p(B)\phi_P(B^S)(1-B)^d(1-B^S)^D} a_t \quad (3.2)$$

dengan

$\beta_1, \beta_2, \beta_3$  : koefisien dari efek variasi kalender (bulan sebelum terjadinya hari raya Idul Fitri, bulan saat terjadinya hari raya Idul Fitri, dan bulan setelah terjadinya hari raya Idul Fitri)

$L_{1,t}, L_{2,t}, L_{3,t}$ : dummy variasi kalender (bulan sebelum hari raya Idul Fitri, bulan saat terjadinya hari raya Idul Fitri dan bulan setelah terjadinya hari raya Idul Fitri)

$$\frac{\theta_q(B)}{\phi_p(B)(1-B)^d} : \text{komponen ARIMA}$$

$$\frac{\theta_q(B)\theta_Q(B^S)}{\phi_p(B)\phi_P(B^S)(1-B)^d(1-B^S)^D} : \text{komponen SARIMA}$$

$a_t$ : residual ke-t

Langkah penyelesaian analisis dengan menggunakan model *ARIMAX* adalah sebagai berikut.

1. Menentukan variabel *dummy* untuk model regresi.

Variabel *dummy* untuk variasi kalender dalam penelitian ini yaitu bulan sebelum terjadinya hari raya Idul Fitri ( $L_{1,t}$ ), bulan saat terjadinya hari raya Idul Fitri ( $L_{2,t}$ ), dan bulan setelah terjadinya hari raya Idul Fitri ( $L_{3,t}$ ).  $L_{1,t}, L_{2,t}, L_{3,t}$  diberikan nilai *dummy* 1, sedangkan untuk bulan lainnya diberikan nilai *dummy* 0.

2. Menuliskan persamaan model regresi variabel *dummy*.

Variabel *dummy* sebagai variabel predictor ( $L_{1,t}, L_{2,t}, L_{3,t}$ ), dan data jumlah penumpang kereta api Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen sebagai variabel respon ( $y_t$ ). Sehingga memenuhi persamaan berikut.

$$y_t = \beta_1 L_{1,t} + \beta_2 L_{2,t} + \beta_3 L_{3,t} + \varepsilon_t \quad (3.3)$$

$\beta_{m,t}$ : koefisien dari efek variasi kalender dengan  $m=1,2,3, t=1,2,\dots,t$  dan  $\varepsilon_t$ : residual ke-t.

3. Melakukan estimasi parameter model regresi variabel *dummy*

Estimasi parameter regresi variabel *dummy* dilakukan dengan menggunakan metode *ordinary least square (OLS)*.

4. Melakukan pengujian signifikansi parameter model regresi *dummy*.

Signifikansi parameter dilakukan menggunakan statistik uji t. Jika ada parameter yang tidak signifikan, maka dilakukan estimasi ulang dengan tidak menyertakan parameter yang tidak signifikan. Sehingga diperoleh model regresi *dummy* dengan semua parameter yang signifikan.



5. Melakukan *diagnostic* model pada residual regresi variabel *dummy*.

Residual dikatakan *white noise* apabila telah memenuhi sifat identik dimana variansnya konstan dan independen dimana residualnya tidak saling berkorelasi, dengan rata-rata 0. Statistik uji yang digunakan untuk menguji asumsi *white noise* adalah statistik uji Ljung-Box. Dan untuk menguji apakah deret kekeliruan ( $\varepsilon_t$ ) memiliki distribusi normal, dapat dilakukan Uji Kolmogorov Smirnov.

Jika sudah memenuhi asumsi *white noise* dan memenuhi kenormalan distribusi pada residual, maka dilanjutkan dengan peramalan pada langkah 9. Namun jika tidak *white noise* maka dilanjutkan dengan pendugaan model ARIMA pada residual.

6. Pendugaan model ARIMA/SARIMA (jika residual data tidak *white noise*).

Identifikasi model ARIMA/SARIMA dapat dilakukan dengan melakukan plot fungsi ACF dan plot fungsi PACF pada data yang telah stasioner. Identifikasi model ini akan memperoleh model ARIMA ( $p,d,q$ ) atau SARIMA ( $p,d,q$ )( $P,D,Q$ )<sup>5</sup> yang memungkinkan.

7. Pemodelan ARIMAX/SARIMAX.

Langkah-langkah untuk mendapatkan model ARIMAX/SARIMAX dilakukan dengan langkah sebagai berikut:

7.1 Model ARIMA/SARIMA yang diperoleh dari langkah 7, masing-masing diestimasi secara simultan dengan model regresi *dummy* yang telah diperoleh sebelumnya menggunakan metode penaksiran parameter *Maksimum Likelihood Estimation (MLE)*. Sehingga didapatkan model seperti pada persamaan 3.1.

7.2 Melakukan uji signifikansi parameter-parameter dari model ARIMAX/SARIMAX menggunakan uji *t* (seperti pada langkah 5).

7.3 Melakukan *diagnostic* model (seperti pada langkah 6)

7.4 Ketepatan Model

Model ARIMAX akan dipilih peneliti berdasarkan kriteria ketepatan model sebagai berikut.

a. *Root Mean Square Error (RMSE)*

$$RMSE_{out} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (3.4)$$

Dengan  $Y_t$  adalah nilai data aktual,  $\hat{Y}_t$  nilai ramalan,  $n$  menyatakan banyaknya data yang akan dihitung residualnya.

b. *Mean Absolute Percentage Error (MAPE)*

Jika memiliki nilai yang diramal dan aktual untuk  $n$  periode, MAPE dihitung sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{Z_t} \times 100\% = \frac{1}{n} \sum_{t=1}^n \frac{|Z_t - \hat{Z}_t|}{Z_t} \times 100\%$$

### 3. Hasil dan Pembahasan

Berdasarkan hasil analisis dengan software R didapatkan persamaan regresi variabel *dummy* sebagai berikut:

$$y_t = 10741L_{1,t} + 14233L_{2,t} + 10231L_{3,t} + \varepsilon_t$$

Parameter-parameter diatas sudah signifikan. Namun residual model nya tidak bersifat *white noise* yang dibuktikan dengan adanya garis yang keluar dari garis signifikansi pada plot acf dan pacf pada gambar 3.2.2.

Karena residual regresi variabel *dummy* tidak *white noise*, maka akan dilakukan pendugaan model dengan ARIMA sebagai berikut:

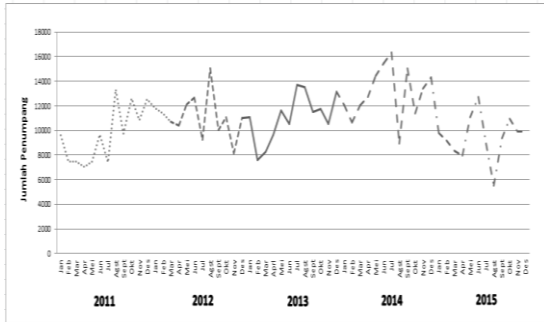
Data yang digunakan ARIMA harus memenuhi kriteria kestasioneran dalam rata-rata dan varians. Hasil pengujian dengan menggunakan software R box.cox test didapatkan lambda 1.99 menunjukkan bahwa data telah stasioner dalam rata-rata. Namun saat dilakukan Adf.test didapatkan hasil p-value=0.05 =  $\alpha=0.05$ . Maka, akan dilakukan proses *differencing* pada data. Proses *differencing* hanya dilakukan satu kali karena data sudah stasioner dalam varians dengan p-value=0.01 <  $\alpha=0.05$ .

Data yang telah di *differencing* akan dilihat plot acf dan pacf nya untuk penentuan orde p dan q. Plot acf dan pacf data tersebut adalah pada gambar 3.2.3. Berdasarkan plot acf dan pacf didapatkan orde p (AR) = 2 dan orde q (MA) = 3.

Setelah itu akan dilakukan pemodelan ARIMAX, yaitu dengan mengestimasi secara simultan (*trial and error*) pada model ARIMA yang memungkinkan tersebut dengan model persamaan regresi variabel *dummy* yang telah didapatkan.

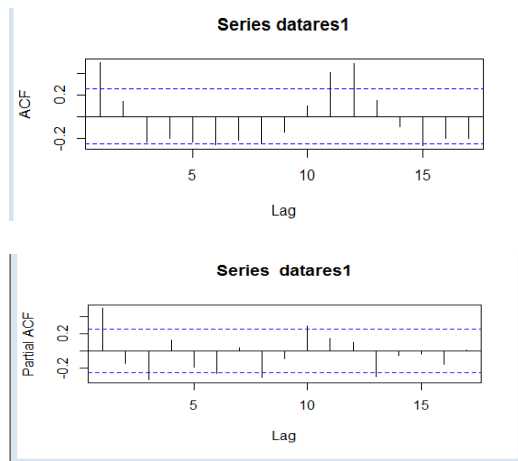
Sehingga didapatkan model ARIMAX yang telah memenuhi syarat *white noise* dan residual berdistribusi normal dengan MAPE sebesar 0.002 adalah ARIMAX (2,1,0). Yang memenuhi persamaan metode variasi kalender model ARIMAX sebagai berikut:

$$y_t = 1074L_{1,t} + 14233L_{2,t} + 10231L_{3,t} + \frac{\theta_q(B)}{2.1064_2(B)(1-B)^1} a_t$$



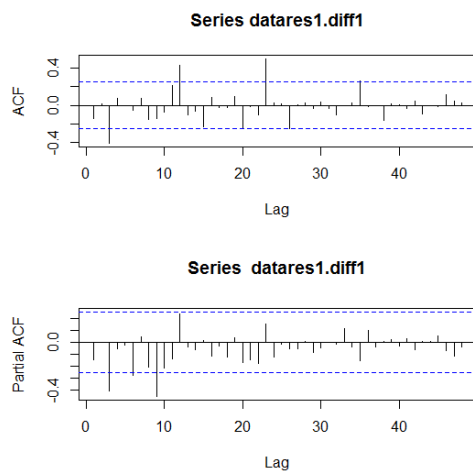
Gambar 1. menunjukkan plot data jumlah penumpang kereta api Fajar Utama Yogya jurusan Yogyakarta – Pasar Senen dari tahun 2011 sampai dengan tahun 2015.

Gambar 1 menunjukkan adanya pelonjakan penumpang efek variasi kalender setiap tahun nya.



Gambar 2. merupakan plot acf dan pacf yang menunjukkan residual dari model regresi variable *dummy* yang belum *white noise*.

Gambar plot acf dan pacf yang ditunjukkan pada Gambar 2 menunjukkan masih adanya garis ke atas yang keluar dari garis *significant* (garis putus-putus) yang menandakan bahwa residual model regresi *dummy* belum *white noise*.



Gambar 3. merupakan plot acf dan pacf yang sudah *white noise*.

Gambar 3 menunjukkan plot acf dan pacf digunakan untuk memnetetukan orde pada pendugaan residual dengan model ARIMA.

#### 4. Kesimpulan

Dari model terbaik (*ARIMAX 2,0,1*) yang telah diperoleh maka dapat dilakukan peramalan untuk bulan Juni 2016 sampai dengan bulan Desember tahun 2016 berturut-turut yaitu 13787, 10318, 9356, 9203, 10180, 13655.

#### Ucapan Terima Kasih

Penulis mengucapkan terimakasih kepada semua pihak yang telah ikut berperan dan membantu baik moril maupun materil dalam proses penyelesaian karya tulis ini ini. Penulis menyadari bahwa dalam penyusunan karya tulis ini masih banyak kekurangan. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun untuk perbaikan karya tulis ini agar dapat disempurnakan agar menjadi bahan perbaikan untuk penelitian selanjutnya.

#### Daftar Pustaka

- Afriani Indri, dkk. 2013. *Pemodelan dan Peramalan Jumlah Pengunjung KBS Menggunakan Model Variasi Kalender ARIMAX*. Statistika, FMIPA, Institut Teknologi Sepuluh Nopember Surabaya.
- Dedi Rosadi. (2012). *Analisis Ekonometrika dan Runtun Waktu Terapan dengan R*. Yogyakarta : Andi Offset
- Izza, Maries Lailatul, dkk. 2014. *Peramalan Penjualan Sepeda Motor Menurut Tipe Dengan Pendekatan Autoregressive Integrated Moving Average With Exogeneous Input (Arimax) Di Kabupaten Banyuwangi*. Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.