 http://dx.doi.org/10.29313/.v6i2.23057

# Pemodelan ARIMAX (*Autoregressive Integrated Moving Average with Exogenous Variable*)

**Tasya Elba Siswanti*, Teti Sofia Yanti**

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*tasya.els@gmail.com

**Abstract.** Time series data are analyzed to obtain a measure that can be used for forecasting. Forecasting is the prediction of the values of a variable based on the values obtained in the past. One of the time series models that can be used to forecast is ARIMA (Autoregressive Integrated Moving Average). In forecasting, there are variables that affect the model which can cause some observation values to increase or decrease drastically and occur repeatedly with different time ranges, so a special model is needed to forecast with these criteria, the model is ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variable). The ARIMAX model is an extension of the ARIMA model with additional variables or exogenous variables that are considered to have a significant effect on the data to increase forecasting accuracy.

**Keywords: Exogenous Variable, ARIMAX Model, Forecasting.**

**Abstrak.** Data deret waktu dianalisis untuk mendapatkan ukuran yang dapat digunakan untuk peramalan. Peramalan merupakan prediksi nilai-nilai sebuah variabel berdasarkan nilai yang diperoleh pada masa lampau. Salah satu model deret waktu yang dapat digunakan untuk melakukan peramalan adalah ARIMA (*Autoregressive Integrated Moving Average*). Dalam melakukan peramalan terdapat variabel lain yang diduga mempengaruhi model yang dapat menyebabkan beberapa nilai observasi mengalami kenaikan atau penurunan secara drastis dan terjadi secara berulang dengan rentang waktu yang berbeda, sehingga dibutuhkan model khusus untuk melakukan peramalan dengan kriteria tersebut yaitu model ARIMAX (*Autoregressive Integrated Moving Average with Exogenous Variable*). Model ARIMAX merupakan perluasan dari model ARIMA dengan variabel tambahan atau variabel eksogen yang dianggap memiliki pengaruh yang signifikan terhadap data untuk memperoleh model yang lebih baik dan menambah akurasi peramalan.

**Kata Kunci: *Exogenous Variable*, Model ARIMAX, Peramalan.**

## 1. Pendahuluan

Deret waktu merupakan serangkaian observasi terhadap suatu peristiwa yang diambil dari waktu ke waktu, dicatat sesuai urutan terjadinya dan kemudian disusun sebagai data statistik (Hanke, dkk., 2003). Deret waktu dianalisis untuk mendapatkan ukuran yang dapat digunakan untuk membuat keputusan masa kini, untuk peramalan, dan perencanaan operasional di masa yang akan datang (Yanti, 2010). Peramalan digunakan untuk memprediksikan suatu peristiwa di masa yang akan datang berdasarkan data lampau yang dianalisis secara ilmiah (Makridakis, dkk., 2003).

Salah satu alasan penting dari peramalan adalah untuk memprediksi kejadian di masa yang akan datang sehingga dapat dipersiapkan beberapa perencanaan untuk menghadapi kemungkinan tersebut. Menurut Makridakis, dkk. (2003), metode peramalan terbagi menjadi dua kategori utama, yang pertama ialah metode kualitatif yang dilakukan apabila data masa lalu tidak tersedia atau data tersedia tetapi jumlahnya tidak mencukupi sehingga dibutuhkan pemikiran, pertimbangan, dan pendapat dari penyusunnya dan yang kedua ialah metode kuantitatif yang dilakukan apabila data masa lalu tersedia. Metode kuantitatif terbagi atas dua model, yaitu model kausal dan model deret waktu yaitu peramalan yang dilakukan berdasarkan nilai yang diperoleh dari masa lalu untuk memperoleh pola dari data masa lalu yang kemudian pola tersebut digunakan untuk peramalan pada masa yang akan datang.

Menurut Box dan Jenkins (1976), terdapat beberapa macam model deret waktu, diantaranya model *Autoregressive* (AR), *Moving Average* (AR), *Autoregressive Moving-Average* (ARMA), dan *Autoregressive Integrated Moving Average* (ARIMA). Adapun model deret waktu yang dapat menjelaskan pengaruh musiman pada data, model tersebut ialah *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Seiring berkembangnya ilmu pengetahuan, model deret waktu ARIMA mengalami perluasan, sehingga terdapat satu model baru yang disebut model ARIMAX atau *Autoregressive Integrated Moving Average with Exogenous Variable*. Model-model deret waktu tersebut dapat digunakan untuk melakukan peramalan.

Dalam melakukan peramalan terdapat variabel lain yang diduga mempengaruhi model. Adanya variabel yang mempengaruhi model deret waktu dapat menyebabkan beberapa nilai observasi mengalami kenaikan atau penurunan yang drastis dan dapat terjadi secara berulang dengan rentang waktu yang berbeda, sehingga dibutuhkan model khusus untuk melakukan peramalan dengan kriteria tersebut. Model deret waktu yang dapat menyelesaikan masalah dengan kriteria tersebut yaitu model ARIMAX. Model ARIMAX merupakan perluasan dari model ARIMA yang menggunakan variabel tambahan atau *exogenous variable* yang dianggap memiliki pengaruh signifikan terhadap data (Cryer & Chan, 2008).

## 2. Landasan Teori

Deret waktu merupakan serangkaian pengamatan terhadap suatu variabel yang diambil dari waktu ke waktu secara konstan dan dicatat sesuai urutan waktu kejadian dimana setiap pengamatan sebagai variabel random $Y_t$ yang didapatkan berdasarkan indeks waktu tertentu ($t_i$) sebagai urutan waktu pengamatan, sehingga penulisan data deret waktu adalah $Y_1, Y_2, \ldots, Y_n$ (Wei, 2006). Data deret waktu adalah sekumpulan hasil pengamatan yang diperoleh menurut urutan kronologis, biasanya dalam interval waktu yang sama. Dari data deret waktu dapat diramalkan bagaimana keadaannya di masa depan (Yanti, 2010).

Data deret waktu yang stasioner merupakan data deret waktu dengan nilai rata-rata yang konstan. Terdapat beberapa metode yang dapat dilakukan untuk pengujian stasioneritas, diantaranya dengan uji akar unit dan teknik grafik (Yanti, 2009). Untuk memeriksa apakah terdapat akar unit atau tidak dapat menggunakan uji *Augmented Dickey Fuller* (ADF) dengan hipotesis uji sebagai berikut:

$H_0$: $\delta = 0$; data mengandung akar unit atau tidak memenuhi asumsi stasioner.
$H_1$: $\delta \neq 0$; data tidak mengandung akar unit atau memenuhi asumsi stasioner.
kemudian statistik uji ADF dapat dihitung dengan rumus sebagai berikut:

$$t_\delta = \frac{\hat{\delta}}{se(\hat{\delta})} \qquad \ldots(1)$$

dengan $\alpha$=5%, hipotesis $H_0$ ditolak jika $|t_\delta| \geq t_{(\alpha,n)}$ *Dickey Fuller* atau *p-value* < $\alpha$.

Aswi dan Sukarna (dalam Laga, dkk., 2018) menjelaskan bahwa untuk mengidentifikasi kestasioneran pada data juga dapat menggunakan diagram ACF. Selain itu, fungsi autokorelasi (ACF) juga digunakan untuk menentukan orde dari AR pada model ARIMA (Baser, dkk., 2018). Adapun persamaan dari ACF ialah sebagai berikut:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \qquad \text{…(2)}$$

Fungsi autokorelasi parsial (PACF) merupakan suatu fungsi yang menunjukkan besarnya korelasi parsial antara pengamatan pada waktu ke-t dengan pengamatan pada waktu sebelumnya (t-1, t-2, …, t-k). PACF juga digunakan untuk menentukan orde dari MA pada model ARIMA. Persamaan PACF ialah sebagai berikut:

$$\hat{\varphi}_{k+1,k+1} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\varphi}_{kj}\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k-1} \hat{\varphi}_{kj}\hat{\rho}_j} \qquad \text{…(3)}$$

ACF dan PACF digunakan untuk menentukan orde-orde dari model ARIMA. Wei (2006) menjelaskan bahwa model ARIMA adalah penggabungan antara model *Autoregressive* (AR) dan *Moving Average* (MA) yang mengalami proses pembedaan atau *differencing* yang dinyatakan dengan *Integrated* (I). Pada praktiknya, data deret waktu tidak selalu stasioner sehingga proses pembedaan atau *differencing* ini digunakan untuk mengatasi data yang belum stasioner.

Menurut Lemke dan Gabrys (2008), model ARIMA merupakan sebuah model yang kompleks untuk memodelkan dan meramalkan deret waktu. Estimasi parameter dalam model ARIMA dapat dilakukan dengan beberapa metode, salah satu metode yang dapat digunakan ialah *Maximum Likelihood Estimator* (MLE). Model ARIMA dinotasikan dengan ARIMA (p,d,q) yang terdiri atas tiga bagian, yaitu:

1. AR (p) atau notasi p merupakan orde untuk proses *autoregressive*.
2. I (d) atau notasi d menjelaskan bahwa data mengalami proses pembedaan atau *differencing*. Pembedaan atau *differencing* digunakan untuk mencapai kestasioneran pada data deret waktu.
3. MA (q) atau notasi q menyatakan orde untuk proses *moving average*.

Persamaan umum dari model ARIMA adalah sebagai berikut:

$$\varphi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t \qquad \text{…(4)}$$

Bentuk model ARMA yang tepat dalam menggambarkan sifat-sifat data dapat ditentukan dengan cara membandingkan plot ACF dan PACF dengan sifat-sifat fungsi ACF dan PACF teoretis dari model ARMA sebagai berikut (Rosadi, 2011):

**Tabel 1.** Cara Membandingkan Plot ACF dan PACF

| Proses | Sampel ACF | Sampel PACF |
|---|---|---|
| *White Noise* | Tidak ada yang melewati batas interval pada lag >0 | Tidak ada yang melewati batas interval pada lag >0 |
| AR (p) | Meluruh menuju nol secara eksponensial | Di atas batas interval maksimum sampai lag ke p dan di bawah batas pada lag >p |
| MA (q) | Di atas batas interval maksimum sampai lag ke q dan di bawah batas pada lag >q | Meluruh menuju nol secara eksponensial |
| ARMA (p,q) | Meluruh menuju nol secara eksponensial | Meluruh menuju nol secara eksponensial |

Dalam melakukan peramalan terdapat variabel lain yang diduga mempengaruhi model. Adanya variabel yang mempengaruhi model tersebut dapat menyebabkan beberapa nilai observasi mengalami kenaikan atau penurunan yang drastis dan dapat terjadi secara berulang dengan rentang waktu yang berbeda. Kondisi tersebut dapat diatasi dengan model deret waktu ARIMAX. Model ARIMAX merupakan perluasan dari model ARIMA dengan variabel tambahan atau variabel eksogen (exogenous variable) yang dianggap memiliki pengaruh yang signifikan terhadap data untuk menambah akurasi peramalan (Cryer & Chan, 2008). Dalam melakukan pemodelan ARIMAX, terlebih dahulu perlu pengujian stasioneritas pada variabel eksogen (Paul, 2015). Menurut Cools (2009), secara umum, bentuk model ARIMAX (p,d,q) dapat ditulis dengan persamaan sebagai berikut:

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_p X_{p,t} + \frac{\theta_q(B)}{\varphi_p(B)(1-B)^d}\varepsilon_t \quad \ldots(5)$$

Adapun persamaan model ARIMAX dengan tren stokastik dan terdapat unsur musiman adalah sebagai berikut:

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_p X_{p,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\varphi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\varepsilon_t \quad \ldots(6)$$

Sedangkan, persamaan model ARIMAX dengan tren deterministik (tanpa order differencing) adalah sebagai berikut:

$$Y_t = \gamma t + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_p X_{p,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\varphi_p(B)\Phi_P(B^S)^D}\varepsilon_t \quad \ldots(7)$$

Langkah yang perlu dilakukan untuk mengetahui apakah model yang diperoleh sesuai atau tidak ialah pemeriksaan diagnostik. Rosadi (2011) menjelaskan bahwa pemeriksaan diagnostik (diagnostic checking) dilakukan untuk memverifikasi kesesuaian model dengan sifat-sifat data. Dengan demikian, residual yang dihitung berdasarkan model yang telah diestimasi mengikuti asumsi dari error dari model teoretis, seperti sifat white noise dan normalitas dari residual. Uji asumsi white noise dilakukan dengan uji Ljung-Box dan uji asumsi residual berdistribusi normal dilakukan dengan menggunakan uji Kolmogorov-Smirnov.

Proses *white noise* adalah barisan variabel *random* yang tidak berkorelasi dan berdistribusi normal dengan rata-rata konstan E($\varepsilon_t$) = 0, variansi konstan Var ($\varepsilon_t$) $= \sigma^2$ dan $\gamma_k = Cov\ (\varepsilon_t, \varepsilon_{t+k}) = 0$ untuk $k \neq 0$. Anggraeni, dkk. (2017) menjelaskan bahwa uji asumsi *white noise* merupakan salah satu metode untuk mengidentifikasi varians kekeliruan homogen dan independen (antara kekeliruan tidak berautokorelasi). Untuk mengetahui apakah residual memenuhi asumsi *white noise* atau tidak, dapat dilakukan dengan statistik uji *Ljung-Box* dengan hipotesis uji sebagai berikut:

$H_0$: $\rho_1 = \rho_2 = \cdots = \rho_k = 0$; model memenuhi asumsi *white noise*

$H_1$: minimal terdapat satu $\rho_k \neq 0$; model tidak memenuhi asumsi *white noise*

dengan statistik uji sebagai berikut:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\rho_k^2}{(n-k)} \quad \ldots(8)$$

dimana *m* merupakan ukuran lag maksimum, *n* merupakan ukuran nilai pengamatan, dan $\rho_k$ merupakan koefisien autokorelasi. Dengan $\alpha = 5\%$, maka kriteria uji ialah $H_0$ ditolak jika $Q < \chi^2_{(\alpha;k-p-q)}$ atau $p - value < \alpha$, dimana *p* merupakan orde PACF dan *q* merupakan orde ACF.

Dalam pemeriksaan diagnostik dilakukan pengujian asumsi residual berdistribusi normal. Metode yang digunakan untuk pengujian asumsi ini ialah uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut:

$H_0$: residual berdistribusi normal

$H_1$: residual tidak berdistribusi normal

statistik uji untuk uji *Kolmogorov-Smirnov* ialah:

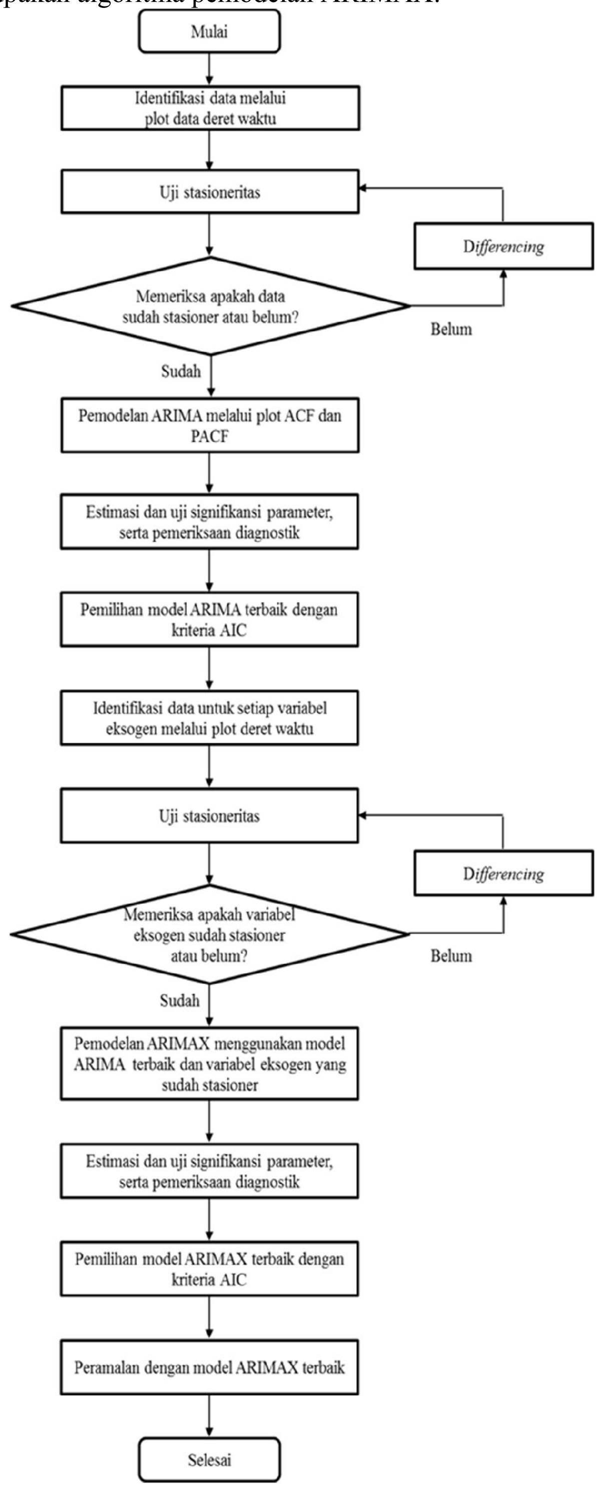
$$D = max|F_0(\varepsilon_t) - S_n(\varepsilon_t)| \quad \ldots(9)$$

dengan $\alpha = 5\%$ maka kriteria uji untuk uji *Kolmogorov-Smirnov* ialah tolak $H_0$ jika $D > D_{(\alpha,n)}$ atau $p - value < \alpha$.

Perihal yang mungkin terjadi pada hasil pemodelan deret waktu ialah terdapat beberapa model dengan semua parameter signifikan, residual memenuhi asumsi *white noise*, dan berdistribusi normal. Sehingga, untuk menentukan model terbaik dari beberapa model tersebut, salah satunya dapat digunakan kriteria *Akaike Information Criterion* (AIC). Model dengan nilai AIC terkecil merupakan model terbaik. Persamaan dari AIC ialah sebagai berikut (Montgomery, Jennings, & Kulahci, 2015):

$$AIC = ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{2p}{T} \qquad \ldots(10)$$

Berikut merupakan algoritma pemodelan ARIMAX:




**Gambar 1.** Algoritma Pemodelan ARIMAX

**Daftar Pustaka**

[1] Anggraeni, W., & dkk. (2017). The Performance of ARIMAX Method in Forecasting Number

of Tuberculosis Patients in Malang Regency, Indonesia. *International Journal of Applied Engineering Research*, 6806-6813.

[2] Baser, U., Bozoglu, M., Eroglu, N. A., & Topuz, B. K. (2018). Forecasting Chestnut Production and Export of Turkey Using ARIMA Model. *Turkish Journal of Forecasting*, 27-33.

[3] Box, G., & Jenkins, G. (1976). *Time Series Analysis, Forecasting, and Control.* San Fransisco.

[4] Cools, M., Elke, & Wets. (2009). Investigating The Variability in Daily Traffic Counts using ARIMAX dan SARIMA(X) Models: Assessing Impact of Holidays on Two Divergent Site Locations.

[5] Cryer, J. D., & Chan, K. (2008). *Time Series Analysis with Application in R.* New York: Springer.

[6] Hanke, John, & dkk. (2003). *Peramalan Bisnis.* Jakarta: Pearson Education Asia Ptc. Ltd dan PT. Pren hallindo.

[7] Laga, A., Wahyuningsih, S., & Hayati, M. N. (2018). Forecasting Clothing Sales with Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX). *Jurnal Eksponensial* , 111-117.

[8] Lemke, C., & Gabrys, B. (2008). Forecasting and Forecast Combination in Airline Revenue Management Applications.

[9] Makridakis, S., Wheelwright, S. C., & McGee, V. E. (2003). *Metode dan Aplikasi Peramalan. Jilid 1. Edisi Revisi.* Jakarta: Binarupa Aksara.

[10] Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting: Second Edition.* New Jersey: John Wiley & Sons, Inc.

[11] Paul, R. K. (2015). ARIMAX-GARCH-WAVELET Model for Forecasting Volatile Data. *Model Assisted Statistics and Applications* , 243-252.

[12] Rosadi, D. (2011). *Analisis Ekonometrika dan Runtun Waktu Terapan dengan R.* Yogyakarta: ANDI.

[13] Wei, W. W. (2006). *Time Series Analysis.* New York: Addison Wesley.

[14] Yanti, T. S. (2010). *Analisis Deret Waktu.* Jakarta-Bandung.

[15] Yanti, T. S. (2009). Model Pengganda Uang untuk Menentukan Jumlah Uang Beredar di Indonesia Menggunakan Model ARIMA Komponen. *Jurnal Statistika , 9*, 25-32.