

Peramalan Banyaknya Pengunjung Pantai Glagah Menggunakan Metode *Autoregressive Integrated Moving Average Exogenous (ARIMAX)* dengan Efek Variasi Kalender

Solikhah Novita Intan¹, Etik Zukhronah², and Supriyadi Wibowo³

^{1,3}Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Sebelas Maret

²Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Sebelas Maret

etikzukhronah@staff.uns.ac.id

ABSTRACT. Glagah Beach is one of the tourist destinations in Kulon Progo Regency, Yogyakarta which is the most visited by tourists. Glagah Beach visitors data show that in the month of Eid Al-Fitr there was a significant increase. This shows that there is an effect of the calendar variation of Eid al-Fitr. Therefore, it is needed a method that can be used to analyze time series data which contains effects of calendar variation, that is ARIMAX method. The aims of this study are to find the best ARIMAX model and to predict the number of Glagah Beach tourists. The result shows that the best ARIMAX model was ARIMAX([24],0,0). Forecasting from January to September 2016 are 37211, 21306, 26247, 24148, 28402, 29309, 81724, 26029, and 23688 visitors.

Keywords: Glagah Beach, variation of calendar, Eid al-Fitr, ARIMAX.

1. Pendahuluan

Pariwisata adalah berbagai macam kegiatan wisata dan didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, pemerintah, dan pemerintah daerah [1]. Menurut Demartoto [2], pariwisata merupakan fenomena kemasyarakatan yang menyangkut manusia, masyarakat, kelompok, organisasi, kebudayaan, dan sebagainya yang merupakan objek kajian sosiologi. Bidang kegiatan yang terlibat dalam sektor pariwisata membuat sektor pariwisata mampu menjadi faktor penting dalam pembangunan ekonomi suatu daerah.

Provinsi Daerah Istimewa Yogyakarta (DIY) mempunyai sejumlah tempat wisata pantai. Salah satunya Pantai Glagah yang terletak di Kabupaten Kulon Progo. Pantai Glagah menyajikan pemandangan laut dengan laguna dan tetrapod di sepanjang pantainya yang tidak dapat ditemukan di pantai lain di Provinsi DIY. Berdasarkan data, banyaknya pengunjung Pantai Glagah setiap tahunnya naik pada bulan-bulan tertentu yang terdapat hari libur dan bulan yang terdapat hari raya Idul Fitri. Namun, kenaikan yang paling signifikan yaitu bulan yang di dalamnya terdapat hari raya Idul Fitri. Hal ini disebabkan karena mayoritas penduduk Indonesia beragama Islam. Hari raya Idul Fitri merupakan

hari raya terbesar umat Islam yang umumnya memiliki masa libur lebih panjang daripada hari libur keagamaan lainnya maupun hari libur nasional lainnya.

Adanya kenaikan pada bulan-bulan tertentu pada data pengunjung Pantai Glagah dari waktu ke waktu tersebut disebabkan oleh efek variasi kalender. Ada dua macam efek variasi kalender, yaitu efek perdagangan dan liburan [3]. Penentuan hari raya Idul Fitri tidak mengikuti kalender Masehi, melainkan berdasarkan kalender Hijriah. Perbedaan waktu hari raya Idul Fitri mengakibatkan setiap tiga tahun sekali hari raya Idul Fitri terjadi di bulan yang berbeda. Hal ini menyebabkan adanya efek variasi kalender.

Penelitian tentang data runtun waktu yang mengandung efek variasi kalender telah dilakukan oleh Bell dan Hillmer [4] yaitu dengan menggunakan efek variasi kalender di hari libur paskah dan model *ARIMA*. Kemudian Chowdhury [5] meneliti tentang efek Idul Adha di *Dhaka Stock Exchange* yang memengaruhi indeks pasar. Selanjutnya Wulansari *et al.* [6] melakukan peramalan *netflow* uang kartal dengan metode *ARIMAX* dan *radial basis function network (RBFN)* yang menunjukkan model *ARIMAX* menghasilkan peramalan yang lebih baik. Pada penelitian ini dilakukan peramalan banyaknya pengunjung Pantai Glagah menggunakan metode *ARIMAX* dengan efek variasi kalender.

2. Variabel *Dummy*

Regresi linear mempunyai bentuk umum yang sama dengan regresi dalam konteks runtun waktu. Pada analisis regresi, variabel dependen dipengaruhi oleh variabel independen yang berupa variabel kualitatif seperti nama bulan, jenis kelamin, agama, dan sebagainya. Variabel kualitatif tersebut dapat diubah menjadi variabel kuantitatif yang bernilai 0 atau 1, atau sering disebut dengan variabel *dummy*. Nilai tersebut digunakan untuk mengindikasikan sifat, dengan 0 untuk tidak adanya sifat, dan 1 untuk adanya sifat. Menurut Gujarati [7], bentuk umum model regresi variabel *dummy* dapat ditulis

$$Z_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_j X_{j,t} + a_t$$

dengan Z_t adalah variabel respon ke- t , $t = 1, 2, \dots, n$, β_0 adalah konstanta, β_i adalah parameter variabel *dummy*, X_i adalah variabel independen, dengan $i = 1, 2, \dots, j$, dan a_t adalah eror ke- t . Sedangkan menurut Suhartono *et al.* [8] model regresi variabel *dummy* untuk efek liburan dapat ditulis

$$y_t = \gamma_0 + \gamma_1 V_{1,t} + \gamma_2 V_{2,t} + \dots + \gamma_l V_{l,t} + w_t$$

dengan y_t adalah nilai pengamatan ke- t , γ_l adalah parameter variabel *dummy* dengan efek liburan, $V_{l,t}$ adalah variabel *dummy* dengan efek liburan, dan w_t adalah eror ke- t .

3. Model ARIMA

Metode *autoregressive integrated moving average* (ARIMA) merupakan penggabungan antara *autoregressive* (AR) dan *moving average* (MA) serta proses *differencing* (orde d untuk data non musiman, orde D untuk data musiman) terhadap data runtun waktu [9]. Secara umum, model ARIMA dituliskan sebagai ARIMA (p, d, q) (P, D, Q) dengan model matematis sebagai berikut :

$$\phi_p B \Phi_P B^S (1 - B)^d (1 - B^S)^D y_t = \theta_q(B) \Theta_Q(B^S) a_t,$$

dengan

(p, d, q) : orde AR (p), *differencing* (d), MA (q) untuk pola non musiman,

$(P, D, Q)^D$: orde AR (P), *differencing* (D), MA (Q) untuk pola musiman,

B : operator *backshift*,

$\phi_p B$: koefisien komponen AR non musiman dengan derajat p ,

$$\phi_p B = (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p)$$

$\Phi_P B^S$: koefisien komponen AR musiman S ,

$$\Phi_P B^S = (1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS})$$

$(1 - B)^d$: operator untuk *differencing* orde d ,

$(1 - B^S)^D$: operator untuk *differencing* musiman S orde D ,

$\theta_q(B)$: koefisien komponen MA non musiman dengan derajat q ,

$$\theta_q B = (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q)$$

$\Theta_Q(B^S)$: koefisien komponen MA musiman S ,

$$\Theta_Q B^S = (1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS})$$

a_t : nilai residu pada waktu ke- t .

Model ARIMA dapat diidentifikasi dengan melihat plot *autocorrelation function* (ACF) dan plot *partial autocorrelation function* (PACF) data yang telah stasioner. Orde proses ARIMA ditentukan oleh *lag* yang keluar pada plot ACF dan plot PACF.

4. Model ARIMAX

Model ARIMAX adalah model ARIMA dengan tambahan variabel *exogenous*. Variabel *exogenous* yang digunakan dapat berupa variabel *dummy* untuk efek variasi kalender dan tren deterministik. Variasi kalender merupakan pola musiman dengan

panjang periode yang bervariasi. Menurut Lee *et al.* [10], variasi kalender dapat disebabkan oleh adanya variasi hari kerja dan variasi hari besar suatu agama/kebudayaan tertentu dari bulan ke bulan hingga tahun ke tahun. Model *ARIMAX* deterministik dinyatakan dengan Persamaan 4.1.

$$y_t = \delta t + \sum_{g=1}^4 \gamma_g V_{g,t} + \sum_{i=1}^{12} \beta_i M_{i,t} + \frac{\theta_q(B)\theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} a_t \quad (4.1)$$

dengan t adalah variabel tren dan M_j adalah *dummy* bulan, $j = 1, 2, \dots, 12$.

5. Uji Signifikansi Parameter

Model *ARIMAX* yang baik residunya harus memenuhi asumsi *white noise* dan berdistribusi normal. Kondisi *white noise* dapat diuji dengan pengujian Ljung-Box dengan hipotesis

$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$ (tidak terdapat korelasi antar residu)

$H_1 :$ paling tidak terdapat satu $\rho_i \neq 0$ untuk $i = 1, 2, \dots, m$ (terdapat korelasi antar residu).

Untuk menguji hipotesis tersebut digunakan statistik uji Ljung-Box yaitu

$$Q_k = n(n+2) \sum_{k=1}^m \frac{(\hat{\rho}_k)^2}{(n-k)},$$

dengan n adalah banyaknya data, k adalah *lag*, $\hat{\rho}_k$ adalah nilai autokorelasi sampai *lag* ke- k , dan m adalah *lag* maksimum yang diuji. Daerah kritis pada uji ini adalah H_0 ditolak

jika nilai $Q > \chi^2_{1-\alpha; k-p-q}$ atau $p\text{-value} < \alpha = 0,05$ [11].

Pengujian distribusi normal untuk residu dapat diperiksa menggunakan uji Kolmogorov-Smirnov dengan hipotesis

$H_0 : F(a_t) = F_0(a_t)$ (residu berdistribusi normal)

$H_1 : F(a_t) \neq F_0(a_t)$ (residu tidak berdistribusi normal).

Statistik uji Kolmogorov-Smirnov adalah sebagai berikut

$$KS = \text{Sup} |F(a_t) - F_0(a_t)|$$

dengan $F(a_t)$ adalah fungsi distribusi kumulatif residu dan $F_0(a_t)$ adalah fungsi distribusi kumulatif dari distribusi normal. H_0 ditolak jika $KS > KS_{1-\alpha; n}$ atau $p\text{-value} < \alpha = 0,05$.

6. Pemilihan Model Terbaik

Model-model *ARIMAX* yang sudah dibentuk kemudian dievaluasi modelnya untuk dipilih sebagai model terbaik dengan membandingkan nilai *Root Mean Square Error (RMSE)*. Model terbaik mempunyai nilai *RMSE* terkecil. Perhitungan *RMSE in-sample* dan *RMSE out-sample* adalah sebagai berikut.

$$RMSE_{in} = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n-p}} \text{ dan } RMSE_{out} = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}$$

dengan n adalah banyaknya data pengamatan dan p adalah banyaknya parameter dalam model.

7. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah studi kasus menggunakan data bulanan banyaknya pengunjung Pantai Glagah, Kulon Progo, Yogyakarta dari tahun 2009 sampai tahun 2015 yang diperoleh dari buku Statistik Kepariwisata Yogyakarta. Langkah-langkah dalam penelitian ini adalah sebagai berikut :

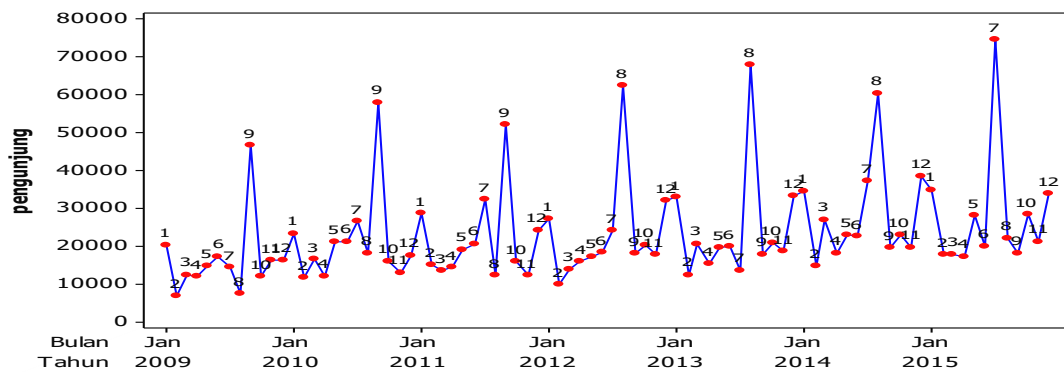
1. Membagi data pengunjung Pantai Glagah menjadi data *in-sample* dari tahun 2009 sampai tahun 2014 dan data *out-sample* menggunakan data tahun 2015.
2. Membuat plot data *in-sample*.
3. Melakukan pemodelan menggunakan metode *ARIMAX* dengan langkah-langkah sebagai berikut :
 - a. Menentukan variabel tren dan variabel *dummy* berdasarkan efek musiman (bulan) dan efek variasi kalender (pada saat hari raya Idul Fitri dan setelah hari raya Idul Fitri).
 - b. Melakukan regresi *time series* untuk menghilangkan efek tren, variasi kalender, dan efek musiman dengan variabel *dummy*. Estimasi parameter menggunakan metode *ordinary least square (OLS)* dan uji signifikansi menggunakan uji t .
 - c. Melakukan pemeriksaan signifikansi parameter dan melakukan eliminasi variabel yang tidak signifikan dan melakukan estimasi ulang model.
 - d. Melakukan uji asumsi *white noise* menggunakan uji Ljung-Box dan uji distribusi kenormalan menggunakan uji Kolmogorov-Smirnov.
 - e. Menentukan model *ARIMAX* melalui plot *ACF* dan plot *PACF*. Kemudian melakukan uji asumsi *white noise* dan kenormalan.
 - f. Melakukan pemilihan model terbaik berdasarkan nilai *RMSE* terkecil.

4. Menghitung peramalan banyaknya pengunjung Pantai Glagah menggunakan model terbaik.

8. Hasil dan Pembahasan

Pada pembahasan ini diuraikan tentang plot data *time series*, variabel *dummy* untuk efek variasi kalender, pembentukan model *ARIMAX*, pengujian signifikansi parameter, dan peramalan untuk banyaknya pengunjung Pantai Glagah tahun 2016.

8.1. **Plot Data.** Plot data banyaknya pengunjung Pantai Glagah dari tahun 2009 sampai tahun 2015 dapat dilihat pada Gambar 1.



Gambar 1. Plot Data Pengunjung Pantai Glagah

Gambar 1 menunjukkan bahwa pada bulan September tahun 2009 terjadi peningkatan yang signifikan, kemudian diikuti bulan September tahun 2010 dan tahun 2011. Selanjutnya bulan Agustus tahun 2012 terjadi peningkatan yang signifikan dan diikuti bulan September tahun 2013 dan tahun 2014. Hal ini mengindikasikan adanya efek variasi kalender pada data pengunjung Pantai Glagah akibat adanya hari raya Idul Fitri yang setiap tahunnya terjadi 11 hari lebih awal dari tahun sebelumnya, sehingga selalu terjadi perubahan setiap tiga tahun sekali.

8.2. **Variabel Dummy untuk Efek Variasi Kalender.** Tren positif pada Gambar 1 dinyatakan dengan variabel t yang merupakan urutan periode data bulanan, sehingga nilai t adalah 1, 2, ..., 84. Variabel yang mewakili efek variasi kalender adalah variabel *dummy* dengan periode mingguan pada saat libur hari raya Idul Fitri dan satu bulan setelah libur hari raya Idul Fitri.

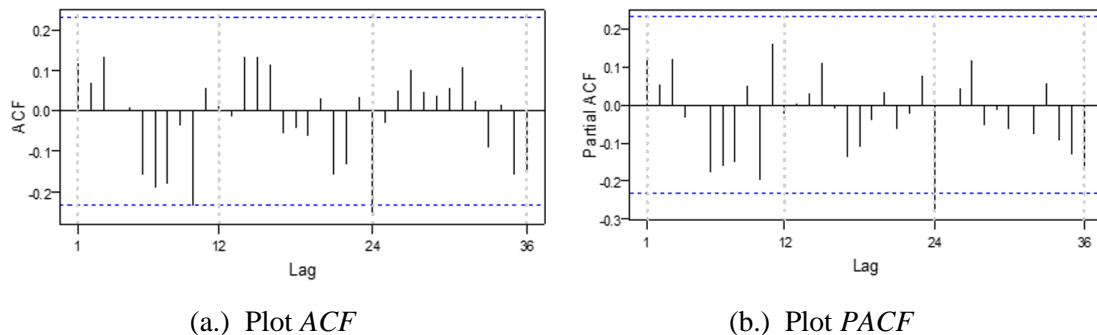
$$V_{i,t} = \begin{cases} 1, & \text{bulan ke } -t \text{ dengan kejadian Idul Fitri di minggu ke } -i \\ 0, & \text{bulan lainnya} \end{cases}$$

$$V_{i,t+1} = \begin{cases} 1, & \text{bulan ke } -(t+1) \text{ dengan kejadian Idul Fitri di minggu ke } -i \\ 0, & \text{bulan lainnya} \end{cases}$$

dengan $i = 1, 2, 3, 4$. Sedangkan variabel *dummy* untuk efek musiman adalah $S_{1,t}$ bernilai 1 untuk bulan Januari dan bernilai 0 untuk bulan yang lain, $S_{2,t}$ bernilai 1 untuk bulan Februari dan bernilai 0 untuk bulan yang lain, dan seterusnya sampai bulan Desember.

8.3. Pembentukan Model ARIMAX. Tahap awal untuk membentuk model ARIMAX yaitu melakukan estimasi regresi *time series* menggunakan metode OLS dan melihat apakah residu sudah stasioner menggunakan uji ADF. Hasil estimasi regresi *time series* menunjukkan bahwa residu sudah stasioner sehingga tidak dilakukan *differencing*. Kemudian melakukan uji parameter menggunakan uji t .

Berdasarkan hasil estimasi dan uji parameter terdapat variabel yang tidak signifikan dengan menggunakan metode *backward elimination* diperoleh model terbaik dengan semua parameter yang telah signifikan yaitu $\delta t, S_{1,t}, S_{2,t}, S_{3,t}, S_{4,t}, S_{5,t}, S_{6,t}, S_{7,t}, S_{8,t}, S_{9,t}, S_{10,t}, S_{11,t}, S_{12,t}, V_{2,t}, V_{3,t}$, dan $V_{4,t+1}$. Model yang memuat parameter tersebut telah memenuhi asumsi *white noise* serta berdistribusi normal untuk residu. Untuk menentukan model ARIMAX dapat dilihat melalui plot ACF dan PACF residu seperti pada Gambar 2.



Gambar 2. Plot ACF dan PACF Residu Regresi *Time Series* yang Telah Signifikan

Gambar 2 menunjukkan model yang didapat yaitu $ARIMAX([24],0,0)$, $ARIMAX(0,0,[24])$, dan $ARIMAX([24],0,[24])$. Model $ARIMAX([24],0,0)$ dan model $ARIMAX(0,0,[24])$ semua parameternya signifikan dan memenuhi asumsi *white noise* serta berdistribusi normal untuk residu, sehingga kedua model tersebut selanjutnya dapat digunakan untuk pemilihan model terbaik.

8.4. **Pemilihan Model Terbaik.** Pemilihan model terbaik berdasarkan nilai dari *RMSE*. Nilai *RMSE out-sample* dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan Nilai *RMSE out-sample*

Model	<i>RMSE out-sample</i>
<i>ARIMAX</i> ([24],0,0)	4428.12
<i>ARIMAX</i> (0,0,[24])	4784,23

Tabel 1 menunjukkan bahwa untuk meramalkan banyaknya pengunjung Pantai Glagah dipilih yang memiliki nilai *RMSE out-sample* yang terkecil yaitu model *ARIMAX*([24],0,0), dengan persamaan sebagai berikut.

$$y_t = 171,57t + 22627,027S_{1,t} + 6550,23S_{2,t} + 11320,16S_{3,t} + 9048,80S_{4,t} + 13131,71S_{5,t} + 13866,55S_{6,t} + 18975,93S_{7,t} + 10243,40S_{8,t} + 7731,70S_{9,t} + 11306,44S_{10,t} + 9254,15S_{11,t} + 20024,92S_{12,t} + 47135,02V_{2,t} + 42608,44V_{3,t} + 37971,57V_{4,t+1} + \frac{1}{(1+0,493B^{24})} a_t \quad (8.1)$$

8.5. **Peramalan *ARIMAX*.** Hasil peramalan *ARIMAX* periode bulan Januari sampai bulan September tahun 2016 dapat dilihat pada Tabel 2.

Tabel 2. Hasil Peramalan *ARIMAX*

Periode	Peramalan	Aktual	Periode	Peramalan	Aktual
Januari	37211	39024	Juni	29309	9783
Februari	21306	18303	Juli	81724	59687
Maret	26247	16537	Agustus	26029	17638
April	24148	14066	September	23688	10835
Mei	28402	21033			

9. Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan sebagai berikut.

1. Model yang sesuai untuk meramalkan banyaknya pengunjung Pantai Glagah adalah *ARIMAX*([24],0,0) yaitu

$$y_t = 171,57t + 22627,027S_{1,t} + 6550,23S_{2,t} + 11320,16S_{3,t} + 9048,80S_{4,t} + 13131,71S_{5,t} + 13866,55S_{6,t} + 18975,93S_{7,t} + 10243,40S_{8,t} + 7731,70S_{9,t} + 11306,44S_{10,t} + 9254,15S_{11,t} + 20024,92S_{12,t} + 47135,02V_{2,t} + 42608,44V_{3,t} + 37971,57V_{4,t+1} + \frac{1}{(1+0,493B^{24})} a_t.$$

2. Hasil peramalan dari bulan Januari sampai September tahun 2016 secara berturut-turut adalah 37211, 21306, 26247, 24148, 28402, 29309, 81724, 26029, dan 23688 pengunjung.

Daftar Pustaka

- [1] Republik Indonesia, *Undang-Undang Republik Indonesia Nomor 10 Tentang Kepariwisata*, Jakarta. 2009.
- [2] Demartoto, A.. *Pembangunan Pariwisata Berbasis Masyarakat*. UNS Press. Surakarta. 2009.
- [3] Liu, L. M., Identification of Time Series Models in the Presence of Calendar Variation. *International Journal and Forecasting*. pp. 357-372.
- [4] Bell, W. R., and S. Hillmer. Modelling Time Series with Calendar Variation. *Journal of American Statistical Association*. pp. 526-534. 1983.
- [5] Chowdhury, T. S.. *Impact of Eid-ul-Azha on Market Return in Dhaka Stock Exchange*. Stamford University. Bangladesh. 2015.
- [6] Wulansari, E. R., dan Suhartono. Peramalan Netflow Uang Kartal Dengan Metode ARIMAX dan Radial Basis Function Network (Studi Kasus Di Bank Indonesia). *Jurnal Sains dan Seni POMITS*. No.2, pp. 73-78. 2014.
- [7] Gujarati, D. N., *Basic Econometrics*, 4 ed., The McGraw-Hill Companies, New York, 2004.
- [8] Suhartono, M. H. Lee, and N. A. Hamzah, *Calendar Variation Model Based on Time Series Regression for Sales Forecasts The Ramadhan Effects*, Prociding of The Regional Conference on Statistical Science. pp. 30-40. 2010.
- [9] Wei, W. W. *Time Series Analysis: Univariate and Multivariate Methods*, 2 ed.. Pearson. New York. 2006.
- [10] Lee, M. H., Suhartono, and N.A. Hamzah, *Calendar Variation Model Based on ARIMAX for Forecasting Sales Data with Ramadhan Effect*, Proceedings of the Regional Conference on Statistical Sciences (2010), pp. 349-361.
- [11] Tsay, R. S. *Analysis of Financial Time Series, Second Edition*. John Wiley & Sons, Inc.. Canada. 2005.