
Model Variasi Kalender pada Regresi Runtun Waktu untuk Peramalan Jumlah Pengunjung Grojogan Sewu

Etik Zukhronah¹, Winita Sulandari², Isnandar Slamet³, Sugiyanto⁴, Irwan Susanto⁵
^{1,2,3,4,5}Program Studi Statistika, Universitas Sebelas Maret

etikzukhronah@staff.uns.ac.id

Abstract. Grojogan Sewu visitors experience a significant increase during school holidays, year-end holidays, and also Eid al-Fitr holidays. The determination of Eid Al-Fitr uses the Hijriyah calendar so that the occurrence of Eid al-Fitr will progress 10 days when viewed from the Gregorian calendar, this causes calendar variations. The objective of this paper is to apply a calendar variation model based on time series regression and SARIMA models for forecasting the number of visitors in Grojogan Sewu. The data are Grojogan Sewu visitors from January 2009 until December 2019. The results show that time series regression with calendar variation yields a better forecast compared to the SARIMA model. It can be seen from the value of root mean square error (*RMSE*) out-sample of time series regression with calendar variation is less than of SARIMA model.

Keywords: Calendar variation, time series regression, SARIMA, Grojogan Sewu

1. Pendahuluan

Grojogan Sewu terletak di Kabupaten Karanganyar Jawa Tengah. Tempat wisata ini dikelola oleh Propinsi Jawa Tengah. Di dekat Grojogan Sewu terdapat Bumi Perkemahan Camping Lawu Resort dan Sekipan. Selain itu di Kabupaten Karanganyar terdapat obyek wisata perkebunan teh yang berada di daerah kemuning.

Pada umumnya, wisatawan yang berkunjung ke Karanganyar menjadikan Grojogan Sewu sebagai tujuan utama dan dilanjutkan kunjungan ke Taman Ria Balekambang dan Sekipan [1]. Pengunjung Grojogan Sewu mengalami kenaikan pada saat libur sekolah, libur akhir tahun dan juga libur Idul Fitri, terutama pada saat libur Idul Fitri pengunjung mengalami kenaikan yang signifikan. Penentuan Hari Raya Idul Fitri menggunakan kalender Hijriyah sehingga terjadinya Hari Raya Idul Fitri akan mengalami kemajuan 10 hari jika dilihat dari kalender Masehi, hal ini menyebabkan adanya variasi kalender. Data yang mengandung variasi kalender dapat dianalisis menggunakan regresi runtun waktu [2].

Data pengunjung Grojogan Sewu juga mengandung pola musiman, karena pada bulan Juni, Juli dan Desember mengalami kenaikan yang signifikan. Model *Seasonal Autoregressive Moving Average (SARIMA)* dapat digunakan untuk memodelkan data yang mempunyai pola musiman [3-6]. Pada artikel ini dilakukan analisis data pengunjung

Grojogan sewu menggunakan regresi runtun waktu dan *SARIMA* dan selanjutnya dibandingkan keakuratan dari kedua metode tersebut dilihat dari nilai *RMSE out-sample* dan *in-sample*.

2. Landasan Teori

2.1. Regresi Runtun Waktu. Model regresi untuk data runtun waktu mempunyai bentuk persamaan seperti pada persamaan regresi linier pada umumnya. Jika terdapat pola tren pada data y_t maka modelnya seperti berikut

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

dengan ε_t merupakan sesatan yang diasumsikan berdistribusi Normal dan independen dengan rata-rata 0 dan variansi σ_ε^2 . Jika datanya mempunyai pola musiman tanpa konstanta, maka modelnya dapat dinyatakan sebagai

$$y_t = \beta_1 S_{1,t} + \beta_2 S_{2,t} + \dots + \beta_s S_{s,t} + \varepsilon_t$$

dengan $S_{1,t}, S_{2,t}, \dots, S_{s,t}$, merupakan variabel *dummy* untuk pola data musiman. Untuk data dengan variasi kalender dapat dimodelkan dengan regresi linier tanpa konstanta seperti berikut

$$y_t = \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_p V_{p,t} + \varepsilon_t$$

dengan $V_{p,t}$ merupakan variabel *dummy* untuk pengaruh variasi kalender ke- p [2].

2.2. Seasonal Autoregressive Moving Average (SARIMA). *SARIMA* dapat digunakan untuk memodelkan data runtun waktu yang mengandung pola musiman. Menurut Shumway dan Stoffer [7], model *SARIMA*(p, d, q)(P, D, Q)^s dapat ditulis seperti berikut

$$\Phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

dengan

- $\Phi_p(B)$: operator untuk parameter *AR* non musiman orde p
- $\Phi_P(B^s)$: operator untuk parameter *AR* musiman orde P
- $\theta_q(B)$: operator untuk parameter *MA* non musiman orde q
- $\Theta_Q(B^s)$: operator untuk parameter *MA* musiman orde Q
- $(1-B)^d$: operator untuk *differencing* non musiman orde d
- $(1-B^s)^D$: operator untuk *differencing* musiman orde D
- y_t : data ke- t

p	: orde AR
q	: orde MA
P	: orde AR musiman
Q	: orde MA musiman
S	: orde untuk musiman

2.3. Uji Autokorelasi. Untuk menentukan apakah terdapat autokorelasi diantara sesatan dapat digunakan uji autokorelasi. Model yang baik adalah model yang sesatannya tidak saling berautokorelasi. Uji Ljung-Box dapat digunakan untuk mendeteksi adanya autokorelasi [6], dengan langkah sebagai berikut.

- i. $H_0 : \rho = 0$ (tidak ada autokorelasi diantara sesatan).
 $H_1 : \rho > 0$ (ada autokorelasi diantara sesatan).
- ii. $\alpha = 0,05$.
- iii. Daerah kritis : H_0 ditolak jika $Q > X^2_{(\alpha,m)}$ atau nilai $p < 0,05$.
- iv. Statistik uji yang digunakan adalah

$$Q = n(n + 2) \sum_{k=1}^m (n - k)^{-1} r_k^2$$

n merupakan banyaknya pengamatan, r_k adalah autokorelasi sesatan pada lag ke- k , dan m merupakan jumlah maksimum lag.

- v. Kesimpulan

2.4. Ukuran Kesalahan. Model dengan nilai ukuran kesalahan kecil merupakan model yang diharapkan. Salah satu ukuran kesalahan yang dapat digunakan untuk mengevaluasi model peramalan adalah *Root Mean Square Error (RMSE)* [2]. Rumus *RMSE in-sample* dan *RMSE out-sample* sebagai berikut

$$RMSE \text{ in-sample} = \sqrt{\frac{1}{(n-p)} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

$$RMSE \text{ out-sample} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

3. Metode Penelitian

3.1 Data. Data pengunjung tempat wisata Grojogan Sewu diambil dari Dinas Pariwisata dan Olah Raga Kabupaten Karanganyar yang dapat di download di disparpora.karanganyarkab.go.id. Data dibagi dua yaitu data *in-sample* dan *out-sample*. Yang digunakan sebagai data *in-sample* yaitu data bulan Januari 2009 sampai dengan bulan Desember 2018, sedangkan data *out-sample* yaitu data bulan Januari sampai dengan Desember 2019.

3.2 Analisis Data Data dianalisis dengan menggunakan 2 metode yaitu regresi runtun waktu dan *SARIMA*. Tahapan yang dilakukan pada regresi runtun waktu sebagai berikut.

- 1) Membuat plot data pengunjung Grojogan Sewu mulai bulan Januari 2009 sampai dengan bulan Desember 2018.
- 2) Menentukan variabel *dummy* pengaruh variasi kalender dan pengaruh musiman. Untuk menentukan banyaknya variabel *dummy* pengaruh variasi kalender dan pengaruh musiman berdasarkan pada plot data runtun waktu pada tahap 1. Variabel *dummy* untuk pengaruh variasi kalender Idul Fitri (V_t) bernilai 1 untuk bulan yang terdapat Idul Fitri dan bernilai 0 untuk bulan lainnya. Untuk variabel *dummy* pengaruh musiman bernilai 1 untuk bulan Januari ($S_{1,t}$) dan 0 untuk bulan yang lainnya, dan seterusnya sampai bulan Desember ($S_{12,t}$) bernilai 1 dan 0 untuk bulan lainnya.
- 3) Melakukan estimasi parameter regresi runtun waktu dengan variabel *dummy* untuk variasi kalender dan pola musiman secara bersama-sama.
- 4) Melakukan uji non autokorelasi pada sesatan model regresi runtun waktu menggunakan uji Ljung-Box. Jika tidak terdapat autokorelasi pada sesatan, maka model yang diperoleh sudah sesuai. Jika terdapat autokorelasi pada sesatan, maka ditambahkan lag yang signifikan (orde *AR*) yang ditentukan berdasarkan plot *ACF* dan *PACF* sesatan ke dalam model regresi.
- 5) Menghitung nilai *RMSE in-sample* dan *out-sample*.

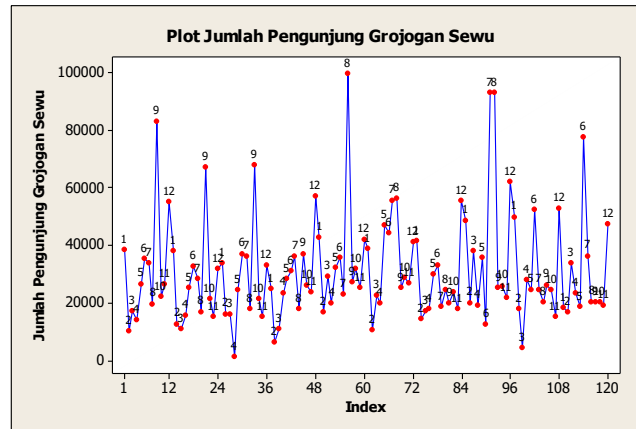
Tahapan untuk model *SARIMA* sebagai berikut.

- 1) Menentukan pola data berdasarkan plot data, plot *ACF* dan *PACF*. Jika data belum stasioner maka dilakukan *differencing*, bisa *differencing* non musiman maupun musiman.
- 2) Menentukan orde *AR* dan *MA* dari data yang sudah stasioner.
- 3) Menghitung estimasi parameter model *SARIMA*.
- 4) Melakukan uji non autokorelasi pada sesatan model *SARIMA* menggunakan uji Ljung-Box.
- 5) Menghitung nilai *RMSE out-sample* dan *in-sample*.

Tahap terakhir adalah membandingkan nilai *RMSE in-sample* dan *out-sample* dari model regresi runtun waktu dan *SARIMA*. Model terbaik ditentukan berdasarkan nilai *RMSE out-sample* terkecil.

4. Hasil dan Pembahasan

4.1 Analisis Regresi Runtun Waktu. Plot data pengunjung tempat wisata Grojogan Sewu terdapat pada Gambar 1.



Gambar 1. Plot data pengunjung Grojogan Sewu

Gambar 1 menunjukkan bahwa pada bulan September 2009, 2010, 2011, 2012, bulan Agustus 2013, bulan Juli dan Agustus 2014, bulan Juli 2015, bulan Juli dan Agustus 2016, bulan Juni 2017 dan bulan Juni 2018 jumlah pengunjung Grojogan Sewu signifikan tinggi. Hal ini menunjukkan adanya variasi kalender pada data walaupun tahun 2012 polanya tidak begitu kelihatan. Pada bulan-bulan tersebut jumlah pengunjung tinggi karena bersamaan dengan adanya hari libur perayaan Idul Fitri seperti yang terdapat pada Tabel 1. Selain itu data pengunjung Grojogan Sewu juga mengalami kenaikan yang signifikan pada setiap bulan Juni, Juli dan Desember. Hal tersebut mengindikasikan data mengandung pola musiman pada data. Berdasarkan Gambar 1, banyaknya variabel *dummy* untuk variasi kalender adalah satu yaitu V_t , sedangkan variabel *dummy* untuk musiman adalah $S_{1,t}, S_{2,t}, \dots, S_{12,t}$.

Dengan menggunakan metode estimasi *Ordinary Least Square (OLS)*, diperoleh model regresi runtun waktu dengan variabel *dummy* variasi kalender dan musiman:

$$y_t = 25959 V_t + 37526 S_{1,t} + 14303 S_{2,t} + 20116 S_{3,t} + 18343 S_{4,t} + 29351 S_{5,t} \\ + 33963 S_{6,t} + 30810 S_{7,t} + 30907 S_{8,t} + 34656 S_{9,t} + 24603 S_{10,t} \\ + 20735 S_{11,t} + 47776 S_{12,t} + \varepsilon_t$$

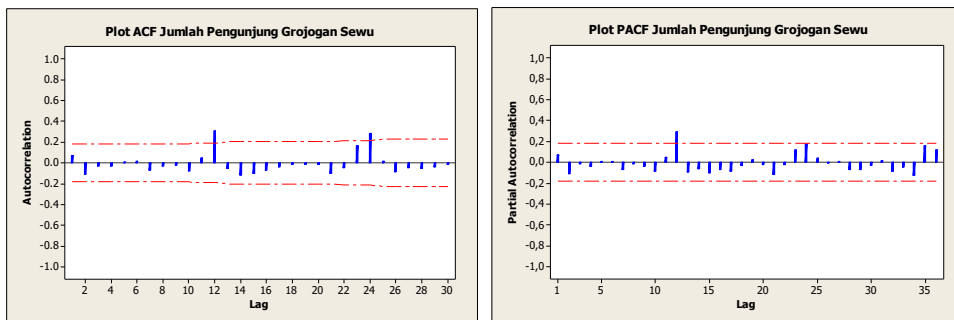
dengan V_t merupakan variabel *dummy* saat bulan perayaan Hari Raya Idul Fitri, $S_{1,t}, S_{2,t}, \dots, S_{12,t}$ merupakan variabel *dummy* untuk bulan Januari sampai dengan Desember. Semua parameter adalah signifikan. Selanjutnya dilakukan uji Ljung-Box

pada 24 lag pertama diperoleh nilai 17,16 yang lebih kecil dari nilai $\chi^2_{11,0.05} = 19,68$, yang berarti bahwa tidak ada autokorelasi diantara sesatan model regresi runtun waktu.

Tabel 1. Libur Hari Raya Idul Fitri Tahun 2009-2019

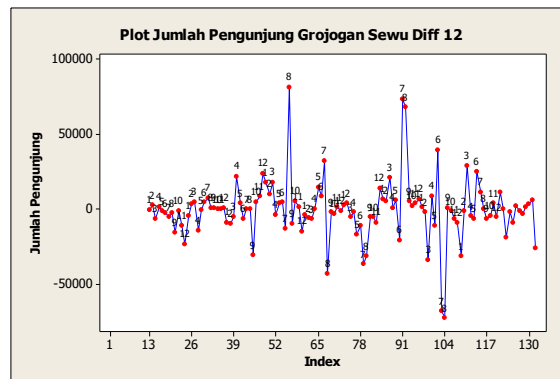
Tahun	Tanggal	Tahun	Tanggal
2009	20-21 September	2015	18-19 Juli
2010	10-11 September	2016	6-7 Juli
2011	31 Agustus-1 September	2017	26- 27 Juni
2012	19-20 Agustus	2018	15-16 Juni
2013	8-9 Agustus	2019	5-6 Juni
2014	29- 30 Juli		

4.2 Analisis SARIMA Gambar 1 menunjukkan bahwa plot data pengunjung Grojogan Sewu tidak stasioner. Plot *ACF* dan *PACF* data pengunjung Grojogan Sewu terdapat pada Gambar 2.



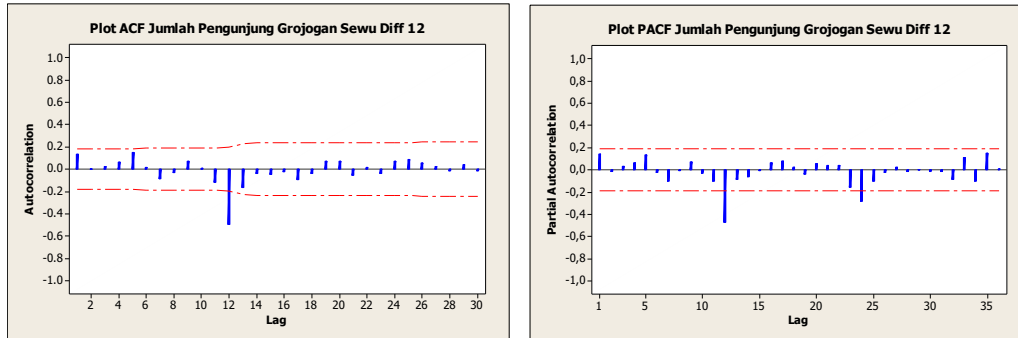
Gambar 2. Plot *ACF* dan *PACF* jumlah pengunjung Grojogan Sewu

Gambar 2 pada Plot *ACF* dapat dilihat bahwa pada lag 12 dan 24 nilainya tinggi, demikian juga pada plot *PACF* nilai lag 12 juga tinggi. Hal tersebut mengindikasikan bahwa pada data terdapat pola musiman. Selanjutnya dilakukan *differencing* 12, dan hasilnya terdapat pada Gambar 3.



Gambar 3. Plot jumlah pengunjung Grojogan Sewu *differencing* 12

Gambar 3 menunjukkan bahwa data yang sudah *didifferencing* 12 sudah stasioner, selanjutnya Gambar 4 merupakan plot ACF dan PACF data yang sudah *didifferencing* 12.



Gambar 4. Plot ACF dan PACF jumlah pengunjung Grojogan Sewu Diff 12

Gambar 4 memperlihatkan nilai lag 12 keluar dari interval konfidensi pada plot ACF, sedangkan pada plot PACF nilai lag keluar pada lag 12 dan 24. Berdasarkan hal tersebut maka nilai orde $P=2$ dan $Q=1$. Model SARIMA yang mungkin adalah $SARIMA(0,0,0)(2,1,1)^{12}$, $SARIMA(0,0,0)(2,1,0)^{12}$, $SARIMA(0,0,0)(1,1,1)^{12}$ dan $SARIMA(0,0,0)0,1,1)^{12}$. Hasil estimasi parameter masing-masing model secara berturut-turut terdapat pada Tabel 2, Tabel 3, Tabel 4 dan Tabel 5.

Tabel 2. Estimasi parameter SARIMA (0,0,0)(2,1,1)¹²

Parameter	Estimasi	Nilai p	Keterangan
ϕ_1	-1,3334	0,000	Signifikan
ϕ_2	-0,8707	0,000	Signifikan
θ_1	-0,5951	0,001	Signifikan

Tabel 3. Estimasi parameter SARIMA (0,0,0)(2,1,0)¹²

Parameter	Estimasi	Nilai p	Keterangan
ϕ_1	-0,8431	0,000	Signifikan
ϕ_2	-0,5735	0,000	Signifikan

Tabel 4. Estimasi parameter SARIMA (0,0,0)(1,1,1)¹²

Parameter	Estimasi	Nilai p	Keterangan
ϕ_1	-0,0086	0,945	Tidak Signifikan
θ_1	0,8706	0,000	Signifikan

Tabel 5. Estimasi parameter SARIMA (0,0,0)(0,1,1)¹²

Parameter	Estimasi	Nilai p	Keterangan
-----------	----------	-----------	------------

θ_1	0,8716	0,000	Signifikan
------------	--------	-------	------------

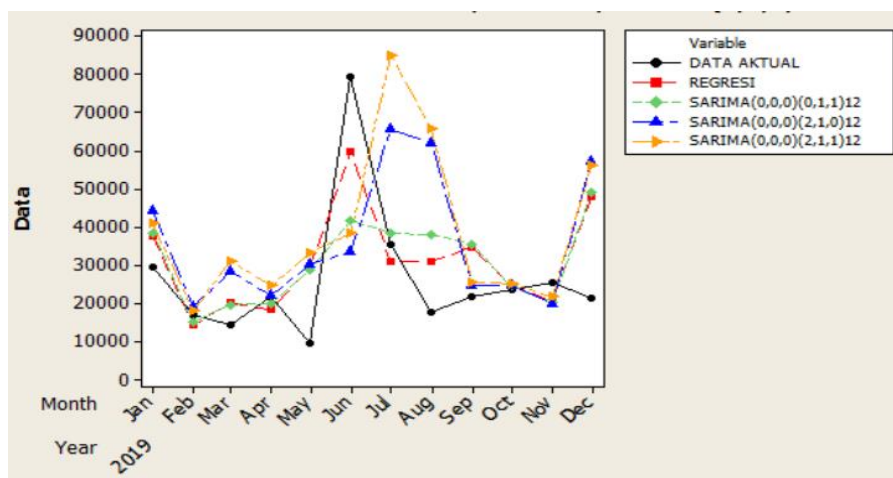
Berdasarkan Tabel 2, 3, 4, dan Tabel 5, model yang semua parameternya signifikan adalah $SARIMA(0,0,0)(2,1,1)^{12}$, $SARIMA(0,0,0)(2,1,0)^{12}$ dan $SARIMA(0,0,0)(0,1,1)^{12}$. Selanjutnya dilakukan uji Ljung-Box dan secara berturut-turut diperoleh nilai p untuk 24 lag pertama adalah 0,834, 0,818 dan 0,982 yang nilainya kurang dari 5%, sehingga H_0 diterima artinya tidak ada autokorelasi diantara sesatan dari ketiga model tersebut.

Untuk menentukan model terbaik dihitung nilai $RMSE$ *out-sample* dan *in-sample* terkecil. Hasil perhitungan nilai $RMSE$ *in-sample* dan *out-sample* dari model regresi runtun waktu dan $SARIMA$ terdapat pada Tabel 6.

Tabel 6. Nilai $RMSE$ *out-sample* dan *in-sample* dan semua model

MODEL	$RMSE$ <i>In-Sample</i>	$RMSE$ <i>Out-Sample</i>
Regresi runtun waktu	13.832,1	12.799,3
$SARIMA(0,0,0)(2,1,1)^{12}$	15.855,1	42.853,7
$SARIMA(0,0,0)(2,1,0)^{12}$	15.940,2	33.023,3
$SARIMAX(0,0,0)(0,1,1)^{12}$	15.668,7	15.027,2

Tabel 6 menunjukkan bahwa model regresi runtun waktu mempunyai nilai $RMSE$ *out-sample* dan *in-sample* terkecil. Berdasarkan hasil perhitungan tersebut maka dapat disimpulkan bahwa model regresi runtun waktu dengan variabel dummy variasi kalender dan musiman mempunyai ukuran kesalahan terkecil daripada model yang lain seperti terlihat pada Gambar 5.



Gambar 5. Perbandingan nilai peramalan model regresi runtun waktu dan *SARIMA*

5. Kesimpulan

Secara umum data runtun waktu yang mengandung variasi kalender perlu penanganan khusus. Model *SARIMA* kurang dapat digunakan untuk memodelkan data yang mempunyai pola variasi kalender karena masih menghasilkan nilai *RMSE* yang lebih besar. Regresi runtun waktu dengan variabel *dummy* variasi kalender dan musiman lebih tepat digunakan daripada model *SARIMA* karena menghasilkan nilai *RMSE* yang lebih kecil.

Daftar Pustaka

- [1] Sunarwan. *Analisis Potensi Obyek Wisata Grojogan Sewu Terhadap Pengembangan Wisata Di kecamatan Tawangmangu Kabupaten Karanganyar*. Publikasi Ilmiah. Program Studi Geografi. Universitas Muhammadiyah Surakarta. 2012.
- [2] Suhartono, Lee, M.H., and Hamzah, N. A. Calender Variation Model Based on Time Series Regression for Sales Forecast The Ramadhan Effects, *Proceedings of the Regional Conference on Statistical Sciences*. 2010.
- [3] Brooks, C., and Persaud, G. Seasonality in Southeast Asian Stock markets: some new evidence on day-of-the week effects. *Applied Economics Letters*. 8, 155-158. 2001.
- [4] Seyyed, F.J., Abraham, A. and Al-Hajji, M. Seasonality in Stock Returns and Volatility: The Ramadhan Effect. *Research in International Business and Finance*, 19, 374-383. 2005.
- [5] Wei, W. W. S. *Time Series Analysis, Univariate and Multivariate Methods*. Addison Wesley Publishing Company. Canada. 2006.
- [6] Hanke, J. E. and Wichern, D. *Business Forecasting*. Ninth Edition. Pearson Education Limited. United Kingdom. 2014.
- [7] Shumway, R. H. and Stoffer, D. S. *Time Series Analysis and Its Applications with R Examples*. Springer Science Business Media LLC. London. 2011.