İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 14, No. 1, July 2022, pp. 17–26
ISSN 1300-4077 | 22 | 1 | 17 | 26

APPLICATION OF TIME SERIES ANALYSIS TO CLINICAL DATA (HEART RATE (HR), SYSTOLIC BLOOD PRESSURE (SBP), AND DIASTOLIC BLOOD PRESSURE (DBP))

Kamber Kaşali*
Department of Biostatistics,
Ataturk University,
25240, Erzurum, Turkey

Ahmet Dirican
Department of Biostatistics,
Istanbul University,
34098, Istanbul, Turkey

Abstract: Multivariable analysis methods are frequently used in studies in the field of health carried out through the variables such as heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP), etc. In this respect, the basic purpose of this study is to demonstrate that it is more appropriate to analyze the clinical variables that change over time with time series analysis. Data used in the study were obtained from twenty-four-hour rhythm and blood pressure results of holter monitor worn by the patients who have consulted cardiology policlinic with the complaint of blood pressure and heart attack. Heart rate rates (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP) variables were obtained from the appropriate 250 files. According to the results, there is a causal relationship between HR with SBP and DBP for male and female patients. The p values are 0.0017 and 0.0084 for males and 0.0056 and 0.0001 for females, respectively. This result shows that SBP and DBP can be used to predict HR. According to the results of the time series analysis, it is shown that HR and SBP and DBP variables are correlated but correlations are immediate, and stabilized over time. In our study, it has been shown that applying time series analysis for the time-varying data will give more detailed results.

Key words: Time series, Cointegration, Granger causality.

1. Introduction

Time series is called as a dataset of consecutive observations of an event in a given time period (hours, days, weeks, months, years, etc.). Changes in observations in time series arise from trends, seasonal movements, cyclical movements, and irregular fluctuations [3]. Analysis approaches in time series depends on whether the series is stationary or not. Whether initially non-stationary series act in the same way in later period is examined by cointegration analysis [7]. The time series consist of four components [10]:

a) Trend (T).
b) Seasonal Variations (S).
c) Cyclic Variations (C).
d) Random or Irregular movements (R).

* Corresponding author. E-mail address:kmbrkasali76@gmail.com

A time series can contain one or a few of the above components. Between the actual observation values of the time series Y and the above components

$$Y = T + S + C + R, Y = T * S * C * R, \tag{1.1}$$

there is such a relationship given in Eq. (1.1). Many clinical variables in the field of health are likely to change over time (HR, SBP and DBP, etc.). Use of time series in examining these variables is believed to help reveal important findings of clinical facts [4].

2. Material and methods

2.1. Methods

2.1.1. The stationarity in time series

In time series, if the first mean and variance of the series as well as the high-order moments do not show a change with respect to time, the series is expressed as a stationary time series if it is free from periodic fluctuations in time [9]. The conditions required for any Y_t series to be stationary can be listed as follows:

$$E(Y_t) = \mu \text{ (for all } t\text{'s, Constant average)}, \tag{2.1}$$

$$Var(Y_t) = E(Y_t - \mu) = \sigma^2 \text{ (for all } t\text{'s, Constant variance)}, \tag{2.2}$$

$$\gamma_t = E[(Y_t - \mu)(Y_{t+k}) - \mu \text{ (for all } t \text{ and all } k \neq 0\text{, Constant covariance based on delay distance)}. \tag{2.3}$$

Here k is the lag distance. γ_t is the covariance between two values with k period difference between them. In addition, if the joint and conditional probability distribution process does not change over time, the series is expressed as strongly stationary [13].

2.1.2. Model selection criteria

In time series analyzes, criteria such as R^2 are used to select the most suitable model

$$Y_t = \delta + \phi_1 Y_{t-1} + e_t; t = 1, 2, 3, ..T \tag{2.4}$$

R^2 value in above AR (1) model,

$$R^2 = 1 - \sigma^2/[\sigma^2/(1 - \phi_1^2)] = \phi_1^2. \tag{2.5}$$

The R^2 value depends on ϕ_1 and as the ϕ_1 value gets bigger, the R^2 value will also increase. For this reason, R^2 value in time series is not used much as a selection criterion. There are many selection criteria in time series models. The most commonly used of these are the information criteria put forward by [1] and [15].

$$AIC = ln(ESS/n) + 2k/n, \tag{2.6}$$

$$SIC = ln(ESS/n) + (k.lnn)/n. \tag{2.7}$$

Here n is the number of observations, k is the number of estimated parameters, the ESS is the sum of squares of error terms, and it is expressed as follows:

$$ESS = \sum (Y - Y')^2. \tag{2.8}$$

AIC and SIC criteria are also required to have small values. The delay order with the smallest values is accepted as the most appropriate delay order [7].

2.1.3. The stationarity tests

Many statistical methods are used to determine whether the series is stationary or not. These are generally: Graphical Analysis, Correlogram Analysis and Unit Root Analysis. The Unit Root Analysis: One of the most common methods used to determine stationarity in time series is the "Unit Root" analysis. This analysis is tested with different methods that take into account the breakage that may occur in the series. Dickey-Fuller (DF) Unit Root Test: Dickey and Fuller have revealed whether time series models have a unit [5]. If the following AR (1) process is considered;

$$Y_t = \alpha_1 Y_{t-1} + u_t,$$

in this process, there are 3 different situations for α_1.
1. $|\alpha_1| < 1$ if so, there is a stable root and the series is stationary.
2. $|\alpha_1| = 1$ if so, the series is not stationary, that is, it is unit rooted.
3. $|\alpha_1| > 1$ if so, it is unstable and there is no unit root.

Augmented Dickey-Fuller (ADF) Unit Root Test: Correlation can occur between variables in analysis in time series. In cases with such problem, the Augmented Dickey-Fuller (ADF) test, the extended version of Dickey-Fuller test, is used [8].

Phillips-Perron (PP) Unit Root Test: When the assumptions of DF and ADF tests were not followed, Phillips and Perron [11] assert the Phillips and Perron test. Phillips and Perron is a non-parametric test that predicts correcting error terms. The Phillips and Perron test's model is given below:

$$Y_t = \mu + \phi_1 Y_{t-1} + u_1, \tag{2.9}$$

and

$$(1 - \phi_1 L) Y_t = \mu + u_1. \tag{2.10}$$

Here t= 1,2,..,T and the unit root for this model are calculated with $1/\phi_1$. If $\phi_1 = 1$ is in the model, the serial is unit root.

Cointegration Analysis: In case the time series is not stationary, whether the series act together in the long term is investigated by cointegration analysis. Engle-Granger Cointegration Test: Engle-Granger [12] was the first to mention the cointegration relationship between series. With this method, the long-term balance relationship between two variables is investigated [2].

2.1.4. Vector error correction model (VECM)

If there is cointegration between time series variables, it is more appropriate to make the causality between variables with error correction (VECM) model. The VECM model is used to distinguish between the long-term balance of variables and short-run dynamics between variables. The VECM model is given in (2.11):

$$\Delta X_t = \alpha + \sum_{t=1}^{m} \beta_i \Delta X_{t-i} + \sum_{t=1}^{n} \gamma_i \Delta Y_{t-i} + \sum_{t=1}^{p} \psi_i \Delta Z_{t-i} + \lambda EC_{t-1} + e_t. \tag{2.11}$$

The x in the model is the error correction value that allows the variables to come to equilibrium in the long term.

2.1.5. Vector autoregression (VAR) model

If there is no cointegration between time series variables, it is expressed with the serial vector autoregressive (VAR) model. The VAR model is given below:

$$Y_t = \alpha_1 + \sum_{i=1}^{m} \alpha_2 Y_{t-i} + \sum_{j=1}^{n} \alpha_3 X_{t-j} + e_{1t}. \tag{2.12}$$

$$X_t = \beta_1 + \sum_{i=1}^{p} \beta_2 Y_{t-i} + \sum_{j=1}^{q} \beta_3 X_{t-j} + e_{2t}. \tag{2.13}$$

In the VAR model, there are dependent and independent variables. Sims [14] said that no distinction should be made between intrinsic and extrinsic variables in the VAR model. Sims proposed the VAR model.

2.1.6. Granger causality test

The testability of the causality of two AR model variables was demonstrated by Granger [6]. The applicability of the test depends on whether both variables are stationary and stochastic.

$$Y_t = \alpha + \sum_{i=1}^{r} b_i Y_{t-i} + \sum_{j=1}^{m} c_j X_{t-j} + e_t; t = 1, 2, ..T \tag{2.14}$$

In the model given in Eq. (2.14), α is constant; b_i, Y_t's previous period coefficient; c_j, X_t's previous period coefficient and is an error term with a white noise process [6].

2.2. Data

The data used in the study were collected from 24-hour rhythm and blood pressure results of holter monitors worn by patients who came to the cardiology outpatient clinic of Haseki Training and Research Hospital with the complaint of blood pressure and heart palpitations. 450 folders of patients were analyzed from hospital records and thus the data were gathered from 250 files as 125 men's and 125 women's files. Those with missing measurements in the data of 450 patients were excluded from the study. The results of 125 men and women were averaged and A single 24-hour data set was obtained for males and females. The variables such as heart rate (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP) were obtained from the files that met the criteria. The definitions for the variables used in the study are summarized below:

1. Heart rate is a fluctuation in the endpoints of the arteries when blood is pumped from the left ventricle to the major arteries,
2. Systolic blood pressure is the pressure in the vein wall of the blood that is excreted from the heart towards the veins when the heart contracts,
3. Diastolic blood pressure is the pressure that is still present in the vessel wall when the heart relaxes.

2.3. Statistical analysis

The results were presented as means, standard deviations, median, minimum and maximum, percentages, and frequencies. The normality distribution of continuous variables was investigated with the Shapiro Wilk test. If there is normality, we used independent samples t-test for two groups comparison. If not, we used Mann Whitney U test for two groups comparison. These analyses were conducted with a statistical analysis program, IBM SPSS 20. The stationarity of HR, SBP and DBP variables were checked using ADF and PP Analyses. Non-stationarity variables were made stationarity by taking their differences. Autocorelation of the model was investigated by the Lagrange Multiplier (LM) test and its heteroskedasticity (varying variance) was checked using White test. When the variables were determined to be an integrated series of the same degree, the cointegration test was conducted using Engle Granger and Johansen methods. Since cointegration was present between the variables, the vector autoregressive model was selected as a candidate model, and thus, the Granger causality test was applied. These analyses were carried out with the EViews 8 statistical analysis program and statistical significance was defined as $p < 0.05$.

3. Results

Findings of this study were summarized below.

TABLE 1. Descriptive statistics

		MALE (N:125 t:24)			FEMALE (N:125 t:24)		
		HR	SBP	DBP	HR	SBP	DBP
Mean		76.14	125.17	77.1	74.65	132.23	80.45
95% confidence interval	Upper limit	73.36	123.16	75.29	71.7	130.17	78,65
	Lower limit	78.93	127.19	78.91	77.6	134.3	82.25
Standard Deviation		6.59	4.77	4.29	6.98	4.9	4.27
Median		78.53	126.72	78.6	76.77	134.34	81.55
Minimum		65.86	116.15	68.5	63.7	122.87	73.16
Maksimum		85.7	131.7	83.81	83.45	137.66	87.76
Range		19.84	15.55	15.31	19.75	14.79	14.6
IQR (İnterquartile Range)		12.68	8.77	6.85	14.15	9.09	7.65
Skewness		-0.27	-0.69	-0.62	-0.38	-0.64	-0.4
Kurtosis		-1.41	-0.85	-0.76	-1.45	-1.18	-1.09

TABLE 2. Gender comparison

	Male (N:125 t:24)		Female (N:125 t:24)			
	Mean ± SD	Median ± IQR	Mean ± SD	Median ± IQR	Z	p
HR	76.14 ± 6.59	78.53 ± 12.68	74.65 ± 6.98	76.77 ± 14.15	-0.969	0.332
SBP	125.17 ± 4.77	126.72 ± 8.77	132.23 ± 4.9	134.34 ± 9.09	-3.794	<0.001
DBP	77.1 ± 4.29	78.6 ± 6.85	80.45 ± 4.27	81.55 ± 7.65	-2.608	0.009
Z: Statistical value for Mann Whitney U Test						

As shown in the Table 2, there was not a statistically significant difference between male and female patients in terms of HR (p=0.332), whereas the SBP (p<0.001) and DBP (p<0.001) variables were statistically significant in terms of gender. In this study, the data of men and women were analyzed as different layers in order to examine the trends of the sexes separately.

3.1. Results of time series analysis

The relationship between HR and SBP and DBP obtained from 125 male patients was examined using both Engle-Granger and Johansen cointegration tests. Before applying the cointegration tests, the series should have become stationary when the differences of the same degree are taken. Because the degree of integration of the series must be the same. ADF and PP unit root tests will be applied to the series to show whether this required condition is fulfilled. Therefore, the graphs will be first examined to see the properties of series. The graphics for the series dealt with are shown separately and together below (Figure 1, Figure 2).

When looking carefully at the graphs (Figure 1, Figure 2), it is seen that the series show non-stationary properties and move parallel together. Before moving on to the analysis, our expectation is that the series are unit-rooted at the level and are related in the long run. Unit root tests are required to indicate this condition. The unit root test results of the series are given below (Table 3).

When looking at the graphs of the series (Figure 1, Figure 2), it is understood that fixed and trending equations should be considered. Since there is a 24-hour-time-series, maximum 5 delays are given and the appropriate delay is determined according to Schwarz. The obtained results from ADF and PP tests are as follows:

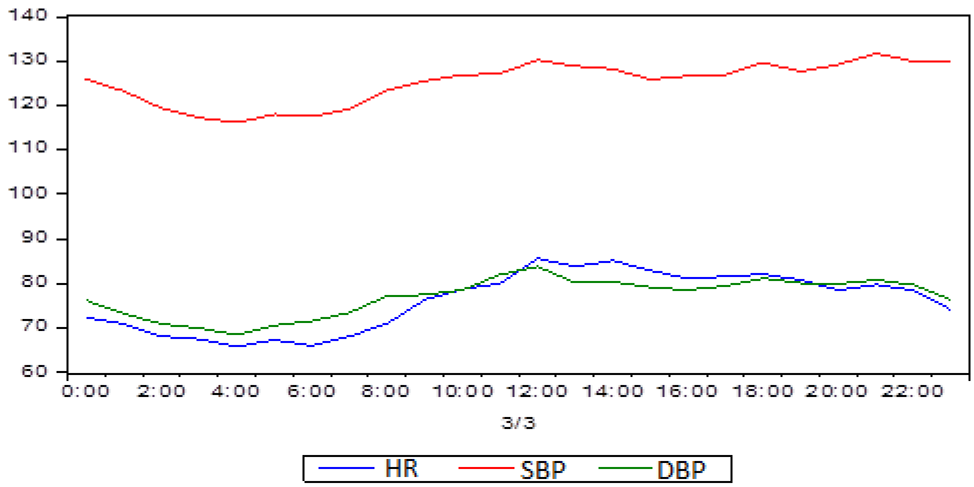

FIGURE 1. HR, SBP and DBP graph of male patients' average.

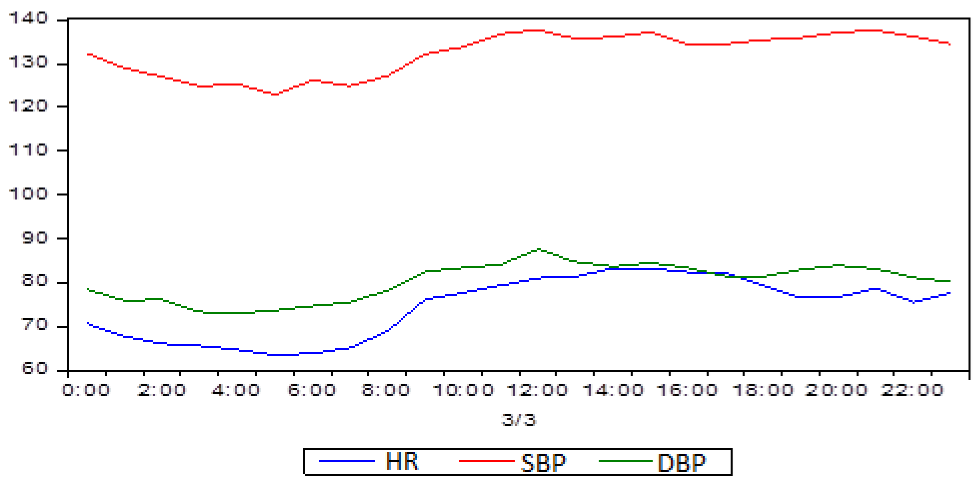

FIGURE 2. HR, SBP and DBP graph of female patients' average.

TABLE 3. ADF and PP tests results

	Male Patients				Female Patients			
	I(0)		I(1)		I(0)		I(1)	
	ADF	PP	ADF	PP	ADF	PP	ADF	PP
HR	0.755	0.677	0.042	0.041	0.794	0.714	0.047	0.047
SBP	0.799	0.702	0.013	0.013	0.746	0.662	0.006	0.006
DBP	0.704	0.61	0.037	0.043	0.706	0.628	0.012	0.013

Looking at the results in the tables, the ADF and PP unit root test results of the HR, SBP and DBP series can be seen. When $p>0.05$, the series are unit-rooted, not stationary. In the original I(0) state of the series, it is seen that they are not stationary ($p>0.05$). For this reason, the test was applied again by taking the first difference I(1) of the series. As the trend effect disappeared for the first difference series, it became stationary ($p<0.05$).

3.2. Examination of autocorrelation between data of patients by LM test

When the LM test results (Table 4) are looked, it is seen that there is autocorrelation since P<0.05 is present even in 1 delayed state. The Newey-West Test was used to eliminate the problem of autocorrelation. The Newey-West test result is shown below:

Table 4. LM tests results

	Male Patients			Female Patients		
SBP, DBP with HR	LM(1)	LM(2)	LM(3)	LM(1)	LM(2)	LM(3)
P	0.001	0.006	0.017	0.009	0.016	0.027

3.3. Heteroskedasticity white correction test between the data of patients

According to the White test result (Table 5), there is no heteroskedasticity problem in our model. The Newey-west test is used to eliminate this problem. The Newey-west Test simultaneously eliminates both the problems of autocorelation and heteroskedasticity. The Newey-west correction version of our model is given below once again The equation obtained from this test will be used to find the cointegration relationship.

Table 5. Heteroskedasticity white correction test results

Heteroskedasticity Test: White					
Male Patients			Female Patients		
Statistical value		p	Statistical value		p
F-statistic	1.521.049	0.2327	F-statistic	1.181.623	0.3566
Obs*R-squared	7.128.456	0.2113	Obs*R-squared	5.930.823	0.3130
Scaled explained SS	4.044.196	0.5431	Scaled explained SS	3.721.332	0.5902

3.4. Newey-west correction test for autocorelation and heteroskedasticity between data of patients

The unit root tests of the residue series obtained from the above model were examined by applying ADF test and PP test (Table 6).

In the Enger-Granger method, the residue series obtained from the regression model is stable with respect to all levels of significance (Table 7). Results were obtained from ADF and PP tests. This result shows that a cointegration relationship exists between HR and SBP and DBP in males.In other words, it shows that these series have acted together in the long term. The series has been found to have no short-term relationship.

Table 6. Heteroskedasticity Newey-west correction test results

Dependent variable: D(HR)									
Male Patients					Female Patients				
Variable	Coefficient	Std. Error	T	p	Variable	Coefficient	Std. Error	t	p
D(SBP)	0.410275	0.33894	1.210.45	0.240	D(SBP)	0.514639	0.19845	2.593.28	0.017
D(DBP)	0.573361	0.31488	1.820.84	0.083	D(DBP)	0.269124	0.29050	0.92639	0.365
C	-0.002745	0.35370	-0.00776	0.993	C	0.235660	0.43854	0.53736	0.596

TABLE 7. ADF and PP test results

	Male Patients		Female Patients	
	ADF	PP	ADF	PP
Residue series	0,004	0,003	0,016	0,016

TABLE 8. Granger causality test results

VEC Granger Causality/Block Exogeneity Wald Tests					
Male Patients			Female Patients		
Dependent variable: D(HR)			Dependent variable: D(HR)		
	Chi-Square	P		Chi-Square	p
D(SBP)	12.78838	0.0017	D(SBP)	7.684185	0.0056
D(DBP)	9.555896	0.0084	D(DBP)	16.21758	0.0001
All	14.34794	0.0063	All	16.58381	0.0003
Dependent variable: D(SBP)			Dependent variable: D(SBP)		
	Chi-Square	P		Chi-Square	p
D(HR)	2.656929	0.2649	D(HR)	0.787467	0.3749
D(DBP)	1.194488	0.5503	D(DBP)	4.477577	0.0343
All	4.238053	0.3747	All	9.673283	0.0079
Dependent variable: D(DBP)			Dependent variable: D(DBP)		
	Chi-Square	P		Chi-Square	p
D(HR)	0.684889	0.7100	D(HR)	0.145798	0.7026
D(SBP)	4.131397	0.1267	D(SBP)	1.533346	0.2156
All	5.729465	0.2203	All	1.539706	0.4631

3.5. Granger causality test

According to the results in the Table 8, there is a causality between SBP and DBP with HR for male patients. Because the p values were 0.0017 and 0.0084 respectively. These p values lead to the rejection of the Ho hypothesis for male patients, which states that "the HR series of SBP and DBP series is not the cause of Granger". This result shows us that SBP and DBP are the causes of Granger of HR, that is, they can be used for estimation. Our results show that DBP is not the cause of Granger SBP (p=0.553) and that SBP is not the cause of Granger DBP (p=0.126). Likewise, it is observed that KAH was not seen to be the cause of Granger neither for the SKB nor for the DKB. (P=0.2649; p=0.7100). Consequently, our study shows that male patients have a one-way causality between SBP and DBP with HR. Likewise, the direction of this causality appears to be from SBP and DBP to HR. According to the results in the table, there is a causality between SBP and DBP with HR for female patients since the p values were 0.0056 and 0.0001 respectively. These p values lead to the rejection of the H_0 hypothesis for female patients which states that "the HR series of SBP and DBP series is not the cause of Granger". This result shows us that SBP and DBP are the cause of Granger of HR, that is, they can be used for estimation. Our results show that DBP is the cause of Granger of SBP (p=0.034) and that SBP is not the cause of Granger of DBP (p=0.215). Likewise, it is observed that HR was not seen to be the cause of Granger neither for SBP nor for DBP (P=0.375; p=0.703). As a result, Our study shows that female patients have a one-way causality between SBP and DBP with HR, and the direction of this causality appears to be from SBP and DBP to HR. It also shows that DBP is the cause of Granger of SBP and that the DBP variable can be used in estimating the SBP variable.

4. Discussion

It is important how the time series coheres with each other, or how they relate to each other. Whether these series act together in the short term or the long term is a question of curiosity. How the changing variables change over time provides information about the patients' condition. This information can be extremely important to save lives. While multiple comparisons are being made for time-varying variables in the field of biosatatistics, time series analysis is not used very often. Nevertheless Studies on time-varying data in the medical field have increased recently. Analyses of these data have been tried to be explained by using multiple comparison methods. These methods do not provide information about the change of the series over time. Under these circumstances, it leads misunderstanding in time series. HR, SBP and DBP variables were used in our study. The relationship between variables was examined by using the time series cointegration method. In 2017, Diego Giulliano Destro Christofaro et al. [4] in their study titled "Relationship between Resting heart Rate, Blood Pressure and Pulse Pressure in Adolescents" analysed the relationship between the same variables by using linear regression multiple comparison method. In the study published by Diego Giulliano Destro Christofaro et al., 24-hour data from 716 female and 515 male adolescent patients aged between 14-17 were collected. HR values were calculated (80.1±11.0 beats/minute) for women and (75.9±12.7 beats/minute) for men and were statistically significant ($p<0.001$). In the same study, Resting HR was associated with SBP in males (Beta=0.15 [0.04-0.26]) and female (Beta=0.24 [0.16-0.33]), with DBP in male (Beta=0.50 [0.37-0.64]) and female (Beta=0.41 [0.30-0.53]). Results were calculated and found to be statistically significant. The relationship between variables was revealed in the study, but no information was given about the change of variables over time and about the relationship between short and long term. In our study, when the results of the time series cointegration analysis were checked, the probability values between HR and SBP and DBP were calculated as 0.0017 and 0.0084 in males and 0.0056 and 0.0001 p in females respectively. These p values show us that the Ho hypothesis stating that "SBP and DBP are not the cause of Granger of HR" is rejected. This shows us that SBP and DBP are the causes of granger of HR, that is, they can be used for HR values estimation. In conclusion, both male and female patients have a one-way causality between SBP and DBP with HR. The direction of causality appears to be from SBP and DBP to HR.

5. Conclusions

Looking at the results, it is seen that time series analysis results put forward more detailed results than multiple comparison methods. According to the time series analysis results, it was shown that SBP and DBP with HR variables are related, relations are instantaneous relations, and they come to equilibrium over the long term.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

[2] Bozkurt, H. (2007). *Zaman Serileri Analizi*. Ekin Kitabevi, Bursa.

[3] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden Day, San Francisco.

[4] Christofaro, D.G.D., Casonatto, J., Vanderlei, L.C.M., Cucato, G.G., Dias, R.M.R. (2017). Relationship between resting heart rate, blood pressure and pulse pressure in adolescents. *Arquivos Brasileiros de Cardiologia*, 108(5), 405-410.

[5] Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, (4), 1057-1072.

[6] Granger, C.W.J. (1969). Investigating causal relations by econometrics models and cross spectral methods. *Econometrica*, 37, 3, 424-438.

[7] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

[8] Kadilar, C. (2000). *Uygulamalı Çok Değişkenli Zaman Serileri Analizi*. Bizim Büro Basımevi, Ankara.

[9] Kayim, H. (1985). *İstatistiksel Ön Tahmin Yöntemleri*. Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Ankara.

[10] Newbold, P., Carlson, W.L., Thorne, B.M. (2000). *Statistics for Business and Economics: Global Edition*. Pearson Education, New York.

[11] Phillips, P.C.B., Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* , 75(2), 335-346.

[12] Engle, R.F., Granger, C.W.J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* , 55, 251-276.

[13] Sevuktekin, M. and Nargelecekler, M. (2007). *Ekonometrik Zaman Serileri Analizi*. Nobel Akademik Yayıncılık, Ankara.

[14] Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.

[15] Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.