JOURNAL OF STATISTICS & APPLIED SCIENCES

JOURNAL OF STATİSTİCS & APPLİED SCİENCE, VOLUME - 1, ISSUE – 2
İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ, CİLT - 1, SAYI- 2
E-ISSN 2718-0999

CİLT 1, SAYI 2, 2020

Journal of Statistics & Applied Science, Volume 1, Issue -2
ISSN 2718-0999

İletişim E-posta: jsas.journal@gmail.com : https://dergipark.org.tr/tr/pub/jsas

Journal of Statistics & Applied Science, Volume 1, Issue -2
ISSN 2718-0999

ÖNSÖZ

İstatistik disiplini özellikle uygulama alanında, diğer disiplinlerle birlikte kullanılan ve araştırmaya objektiflik kazandırarak çalışmaları daha bilimsel hale getirmektedir. İstatistik disiplininin uygulama boyutu diğer disiplinlerle yaygın olarak kullanılmasına rağmen, istatistik disiplininin uygulama alanına yönelen yeni yöntemleri bilim insanlarına tanıtan bilimsel dergiler azdır. Derginin kuruluş amacı bu boşluğu doldurarak multidispliner çalışmaları güçlendirmektir. İstatistik ve Uygulamalı Bilimler Dergisi, gerçek yaşamdaki problemler için geliştirilen ve geniş bir bilimsel etkiye sahip yöntemler hakkında net ve erişilebilir makaleler yayınlamaktır. Dergi, orijinal araştırma makalelerine öncelik vermektedir. Genel olarak, orijinal araştırma raporları bir alanda bir ya da iki zorluğu ortaya koymalı, ilgili verileri içermeli, zorlukları çözmek için yeni bir yöntem sunmalı ve önerilen yöntemin ilgili mevcut yöntemlerle daha önce doğru ya da en iyi şekilde cevaplanmayan soruları cevapladığını göstermelidir. Temmuz ve Aralık olmak üzere yılda iki sayı olarak yayınlanmaktadır. Yayımlanmak üzere kabul edilen yayınların her türlü yayın/telif haklarının dergiye ait olduğu yazar tarafından kabul edilir. Dergide yayımlanan makalelerin dil, etik, yasal ve bilimsel sorumluluğu yazara aittir. Makaleler kaynak gösterilmeden kullanılamaz. Tüm hakları saklıdır. İstatistik ve Uygulamalı Bilimler Dergisi'ne yayımlanmak üzere gönderilen çalışmalardan herhangi bir başvuru veya değerlendirme ücreti alınmamaktadır. Sadece DergiPark sistemi üzerinden başvurusu yapılan yayınlar değerlendirmeye alınır.

EDİTÖR

Abdulkadir Keskin

Journal of Statistics & Applied Science, Volume 1, Issue -2
ISSN 2718-0999

İÇİNDEKİLER

1. A stochastic process model for sustainable energy markets of advanced economies Hasan Dinçer, Serhat Yüksel (63-75)
2. A Statistical Analysis of the Relationship Between Meteorological Parameters and the Spread of COVID-19 Cases: Comparison Between Turkey and Italy, Yasemin Ayaz Atalan, Abdulkadir Atalan (76-84)
3. Pandemi Sürecinde Altın Fiyatları ile Kripto Para İlişkisinin Makine Öğrenme Metotları ile İncelenmesi, Ahmet Sel (85-98
4. Does Wind Energy Affect Economic Growth in Developing Countries, Emrah Doğan, Başak Özarslan Doğan (96-105
5 Inference in "One-Way" Random Designs - Discussing Sub-D and ANOVA based Estimators, Adilson Silva (106-118)

Journal of Statistics & Applied Science, Volume 1, Issue -2
ISSN 2718-0999

A stochastic process model for sustainable energy markets of advanced economies

Hasan Dinçer[1] and Serhat Yüksel[2]

[1]*Professor, İstanbul Medipol University, School of Business, İstanbul, Turkey*
[2] *Associate Professor, İstanbul Medipol University, School of Business, İstanbul, Turkey*

* Correspondence Author, e-mail: serhatyuksel@medipol.edu.tr

Abstract

This study aims to evaluate the sustainability in energy markets. For this purpose, oil price volatility is considered with respect to the stability in these markets. On the other side, stock market data and inflation rate are taken into account regarding the financial stability and sustainable macroeconomic performance. Additionally, a stochastic process model is proposed by using VAR analysis for G7 countries so that it is intended to examine this relationship for advanced economies. The findings reveal that the increase in oil prices in G7 countries has no significant effect on stock prices and inflation rate. Considering these results, it is determined that volatility in oil prices does not seriously threaten the financial markets and macroeconomic stability of these countries. This situation shows that G7 countries have a stable financial and economic structure. Therefore, it is understood that in a situation where oil prices increase excessively, these countries will not cause serious problems. These results will also guide the financial and macroeconomic policies that G7 countries will implement. For example, while aiming to control inflation in these countries, it would be appropriate to focus on variables other than oil prices. In addition to the issues mentioned, it is understood that factors other than oil price should be taken into consideration while aiming to increase the efficiency of financial markets.

Keywords: Oil Price Fluctuations; Financial Stability; Macroeconomic Sustainability; VAR Analysis

Özet

Bu çalışma, enerji piyasalarındaki sürdürülebilirliğin değerlendirmesini amaçlamaktadır. Bu amaçla, petrol fiyatlarındaki oynaklık, bu piyasalardaki istikrar açısından değerlendirilmektedir. Öte yandan, finansal istikrar ve sürdürülebilir makroekonomik performans açısından borsa verileri ve enflasyon oranı dikkate alınmaktadır. Buna ek olarak, gelişmiş ekonomiler için bu ilişkinin incelenmesi amacıyla G7 ülkeleri için VAR analizi kullanılarak bir stokastik süreç modeli önerilmiştir. Bulgular, G7 ülkelerindeki petrol fiyatlarındaki artışın hisse senedi fiyatları ve enflasyon oranı üzerinde önemli bir etkisinin olmadığını ortaya koymaktadır. Bu sonuçlar dikkate alındığında, petrol fiyatlarındaki oynaklığın bu ülkelerin finansal piyasalarını ve makroekonomik istikrarını ciddi şekilde tehdit etmediği tespit edilmiştir. Bu durum, G7 ülkelerinin istikrarlı bir mali ve ekonomik yapıya sahip olduğunu göstermektedir. Dolayısıyla, petrol fiyatlarının aşırı yükseldiği bir durumda, ilgili hususun bu ülkelerde ciddi sorunlara yol açmayacağı anlaşılmaktadır. Bu sonuçlar aynı zamanda G7 ülkelerinin uygulayacağı mali ve makroekonomik politikalara da yol gösterecektir. Örneğin, bu ülkelerde enflasyonu kontrol altına almayı hedeflerken, petrol fiyatları dışındaki değişkenlere odaklanmaları yerinde olacaktır. Bahsedilen hususların yanı sıra, finansal piyasaların etkinliğinin artırılması hedeflenirken, petrol fiyatı dışındaki faktörlerin de dikkate alınması gerektiği anlaşılmaktadır.

Anahtar Kelimeler: Petrol Fiyatları; Finansal İstikrar; Makroekonomik Sürdürülebilirlik; VAR

1 INTRODUCTION

Energy is a very important factor in meeting the social needs in the country. The main reason for this is that people need energy from many needs in daily life. For example, people get warm thanks to energy and can lighten their homes [1]. In addition to the issues mentioned, energy also plays a very important role in ensuring the economic development of countries. Energy is one of the important raw materials of industrial production. Therefore, energy is needed to continuously increase the production volume in the country. Thanks to the energy provided, the production volume in the country may increase. This will contribute to the growth of the economy [2]. Furthermore, new investments will create new job opportunities. In this way, it will be possible to reduce the unemployment rate in the country.

Energy can be obtained basically in two different ways. One of the most important sources of energy supply is non-renewable energy types. These energies are those obtained from fossil fuels such as oil and natural gas [3]. The cost of these energies is cheaper than other types of energy. However, the most important negative aspect of the energies mentioned is that they pollute the environment due to carbon emissions. On the other hand, renewable energy types are the second source of energy supply. These types of energy are energy types that take their resources from nature like wind and sun. Therefore, it is accepted that the sources of these energies will never be exhausted. The most important advantage of these energy types is the absence of carbon emissions. In this way, it is accepted that renewable energy sources do not pollute the environment [4]. However, the high initial investment costs in these energy projects are the most important obstacle on these investments.

As can be understood from the above points, energy is an indispensable need for a country and should be provided regardless of its price. The important point in this process is whether the country has sufficient energy reserves. If a country has a sufficient amount of energy, it will be easy to supply. However, if there are not enough energy reserves in this country, this country has to bear some costs in energy supply [5]. There are two different alternatives in front of this country. First, this country can meet its own energy needs with renewable energy investments. The problem here is that the initial cost of these investments is high. In addition, a substantial technical infrastructure is required to make these investments. The second alternative in this process is to supply the required energy from outside [6]. In other words, this country can meet this need by purchasing it from another country that has an energy reserve.

Obtaining the needed energy from the outside has a lower cost initially compared to new renewable energy investments. However, this situation has some risks for the country. First of all, the country will become dependent on energy as a result of the outsourcing of energy. This situation will decrease the energy supply security of the country. This energy obtained from abroad is purchased from the foreign unit [7]. So, if foreign currency becomes more valuable, the energy purchased will also become more expensive. This will cause the country's budget balance to deteriorate. In addition, if there is a political problem with the country where the energy is purchased, there is a risk that this energy will not be available or at a higher cost. Finally, the reduction of oil supply worldwide will also make oil more expensive, which will lead to the country that supplies energy from abroad [8].

When these problems are taken into account, it is understood that the external supply of energy poses a threat to the macroeconomic stability of the country. For example, if energy prices rise, the raw material for industrial production will also increase [9]. Due to these problems, reductions in industrial production will occur. Since this situation will reduce the investments in the country, this situation will lead to the decline in economic growth. In addition to the aforementioned issue, the profitability of companies will also decrease as a result of the decrease in the income in the country. In this case, companies will lay off some of their employees in order to reduce their costs. As a result, the unemployment rate in the country will increase. In other words, the fragility of the country's economy will go up.

On the other hand, as industrial production will become more expensive as a result of the increase in energy prices, the price of products in the market will increase. As a result of the general increase in the prices of the goods in the market, the inflation rate in the country will also increase [10]. Increased inflation will cause uncertainty in the market to increase. Investors, on the other hand, are uneasy in an environment of uncertainty. Therefore, they are reluctant to invest in such an environment. In summary, high inflation will indirectly lead to a decrease in investments in the country. In addition, the interest rate will increase in an environment where inflation is high. Since this situation will increase the cost of investments, this problem will become even bigger.

In this study, the effect of the increase in oil prices on the stock market and inflation was examined. To achieve this goal, Group of Seven (G7) countries are included in the scope of the review. These countries represent the 7 largest economies in the world. Therefore, the problems to be experienced in the financial and macroeconomic systems of these countries can affect the economic system of the world. In the analysis process of this study, the model was established with the help of Vector Auto Regression (VAR) method. This method has several advantages over other similar methods. While other models only look at a one-way relationship, the VAR method takes into account the two-way relationship between variables. Thanks to this situation, it is possible to make a more comprehensive analysis.

The results of this study will guide policy makers. The main reason for this can be understood to what extent oil prices in G7 countries affect financial markets and inflation. Thanks to these results, it will be clearer what kind of policies to control inflation. On the other hand, if there is a relationship between the variables, more explicit strategies can be developed for the financial markets in these countries to be more effective. Analyzing the G7 countries in the study will guide the way to improve the world trade. In addition to the issues mentioned, the use of the VAR method in the analysis process will contribute to achieving more detailed results.

There are four different sections in this study. In the first part of the study, general information about the subject is given. In addition, the second part consists of the literature review. In this section, the variables considered in the study and similar studies for the method used are explained. The third part of the study includes the results of the analysis made with the help of the VAR model. In the last part, there is discussion and conclusion section.

2 LITERATURE REVIEW

Since energy is an indispensable need, this energy need has to be provided regardless of its price. If countries have their own energy reserves, this process can continue more smoothly. However, this requirement is met from other countries since there are not enough energy sources in the country [11]. This situation causes the country to face some risks. For example, when this energy is purchased from abroad, the payment is made in foreign currency. So, if foreign currency is more valuable, the energy purchased will also become more expensive [12]. This overpaid amount will negatively affect the budget balance of the country.

Due to these mentioned negativities, the increase in oil prices is followed by many market experts and academics. It is thought that volatility in these prices may affect many factors in the market [13]. The important point here is that this effect indicated may differ from country to country. For example, in a more fragile market, it can be seen that the increase in oil prices will cause problems in many market factors.

However, in a stronger market, this negative impact is considered to be limited [14]. Therefore, it is thought that it would be appropriate that emerging markets should pay more attention to the increase in oil prices.

Volatility in oil prices is thought to reduce the country's macroeconomic stability. In this process, one of the most important factors is the inflation rate [15]. It is one of the most important raw materials in energy industry production. Therefore, a possible increase in oil prices will increase the costs in industrial production. In this case, the producers will increase their prices due to this cost [16]. Otherwise, the sales prices obtained will be unable to cover the cost of the products. Inflation will occur as a result of the increase in the prices of the majority of the products in a country.

In a country with an inflation rate, investors will be reluctant to make new investments. The main reason for this is that they cannot predict how much the prices will increase. Therefore, they will delay these investment decisions until this problem is resolved [17]. This situation will slow down the country's economic growth. In addition to this mentioned issue, as a result of new investments that cannot be made, the profitability of companies will also decrease. As a result, companies will have to hire workers to lower their costs [18]. This will cause the unemployment rate to increase in the country. As can be seen, the increase in oil prices adversely affects macroeconomic stability, especially in fragile markets.

This has been supported by many studies in literature. Shahrestani and Rafei [19] focused on the oil price shocks on the macroeconomic factors. This analysis is made for Iranian economy. Markov switching vector autoregressive model is used in the examination process. They reached a conclusion that high oil price increase leads to higher inflation rate in the country. This situation threatens the stability of the macroeconomic factors. Also, Al-hajj et al. [20] also examined the effects of increases in oil prices in Malaysia. In this study, unit root and autoregressive distributed lag with structural breaks are considered. According to the analysis results obtained, it was concluded that oil prices cause high inflation and this situation poses a serious threat to the country's economic performance. A similar study was carried out by Nusair [21] for the Gulf countries. Linear and nonlinear ARDL models and panel cointegration models were considered in this study. In this study, it was stated that oil prices had a significant effect on inflation. As a result, it was stated that the increase in oil prices should be paid attention to increase the macroeconomic performance in the country. The idea that volatility in oil prices increases inflation in the country has also been supported by many different researchers in the literature [22-24].

However, some studies in the literature have argued that this argument does not apply to every country. Salisu et al. [25] analyzed the relationship between the increase in oil prices and inflation in their study. For that, dynamic heterogenous panel data models are used in the examination process. It is concluded that this mentioned relationship is not valid for oil exporting countries. Parallel to this study, Bec and De Gaye [26] also studies the impact of oil price on the inflation. In this scope, US, French and UK were evaluated to find this relationship. They stated that in the long run, there is no strong relationship between oil price volatility and inflation rates. On the other side, Choi et al. [27] made a comparative analysis for advanced and developing economies to see if the oil price shocks affect inflation. They reached a conclusion that this relationship is not occurred for advanced countries.

On the other hand, the increase in oil prices may also decrease the efficiency of the financial markets in the country [28]. The excessive increase in oil prices causes problems such as high inflation, threatening the macroeconomic performance of the country [29]. In this case, uneasiness will increase in the country and the financial market investor will not be satisfied with this process. This situation causes portfolio investors to go abroad [30]. As a result, decreases can occur in companies' stocks. It is important that this system operates in a healthy way, as financial markets ensure that the flow of money in the country can be sustained effectively [31].

It is clear that lots of studies supported the view that oil price shocks have an effect on macroeconomic stability and financial market effectiveness. However, some researchers also argued that this relationship is not valid for each type of countries. They indicated that especially for developed economies, oil price has not a strong influence on macroeconomic factors. This situation is mainly occurred for vulnerable economies, such as developing countries. Hence, in this area, new studies should be conducted for different country groups to make a comparative analysis between the results of different evaluations.

3 EVALUATION ON G7 ECONOMIES

In this part of the study, firstly information is given about data set, variable and methodology. After that, VAR analysis results are shared.

3.1 Data Set, Variables and Methodology

In this study, the effects of oil prices on financial development and macroeconomic stability were examined. Within the scope of financial development, the ratio of stock prices to GDP has been taken into consideration. On the other hand, inflation rate was used in relation to macroeconomic stability. Annual data between 1980 and 2018 are used for these variables. The related data has been accessed from World Bank's website. In addition to the issues mentioned, VAR method was used in the analysis process of this study. This method is used to determine the mutual relationship between two or more variables [32,33]. The biggest advantage of this method is that it examines the two-way relationship, not the one-way relationship between variables [34,35]. In this analysis method, effect response graphics and variance decomposition tables can also be accessed [36]. These factors will help to determine the relationship between variables more clearly.

3.2 Analysis Results

In the first stage of the VAR analysis, stationary analysis is conducted. The main reason is that the variables in this analysis should not have unit root. Hence, Levin, Lin and Chu (LLC) and Im, Pesaran and Shin (IPS) panel unit root tests are considered. These unit root tests were preferred in many different studies in the literature [37-42]. The analysis results are given on Table 1.

Table 1. Panel Unit Root Test Results

Variables	LLC Values		IPS Values		Result
	Level Value	First Difference Value	Level Value	First Difference Value	
Oil Price	0.3968	0.0000	0.1497	0.0000	The first difference is used in the analysis.
Stock Values	0.2841	0.0000	0.1297	0.0000	The first difference is used in the analysis.
Inflation	0.0000	-	0.0000	-	It is stationary in the current form.

Table 1 show that the variable of inflation rate is stationary in its current form. On the other hand, the first differences of the variables of oil price and stock values are used in the analysis. The main reason is that the probability values of these variables are higher than 0.05 in their original forms. After that, it is aimed to calculate optimal lag length. In this framework, Akaike (AIC), Schwarz (SC) and Hannan-Quin (HQ) information criteria are accounted. The details of the results are shared on Table 2.

Table 2. Optimal Lag Length

Lag	AIC	SC	HQ
0	20.88	20.92	20.92
1	19.58	19.76*	19.65
2	19.55	19.85	19.67
3	19.60	20.03	19.77
4	19.40*	19.97	19.63*

In Table 2, it is understood that the optimal lag length is 4. For this purpose, three different information criteria are considered. AIC and HQ indicate that optimal lag is 4 whereas it is 1 for SC. Hence, by considering the majority, VAR model is created with the lag of 4. Since there are 3 different variables in the analysis process, 3 different models were established as a result of the VAR analysis. Details of these models are given in Table 3.

Table 3. Details of the Models

Independent Variables		Model 1 Dependent Variable: Inflation	Model 2 Dependent Variable: Oil Price	Model 3 Dependent Variable: Stock Market (SM)
	Coefficient	0.735672	-2.844556	-2.921181
Inflation (-1)	Standard Error	(0.07357)	(1.00407)	(1.54554)
	t-statistics	[9.99974]	[-2.83301]	[-1.89007]
	Coefficient	0.063992	1.806344	0.683050
Inflation (-2)	Standard Error	(0.09269)	(1.26500)	(1.94718)
	t-statistics	[0.69040]	[1.42794]	[0.35079]
	Coefficient	-0.202525	0.088316	1.708924
Inflation (-3)	Standard Error	(0.08835)	(1.20575)	(1.85598)
	t-statistics	[-2.29241]	[0.07325]	[0.92077]
	Coefficient	0.159163	-0.476162	-0.385655
Inflation (-4)	Standard Error	(0.05595)	(0.76366)	(1.17548)
	t-statistics	[2.84455]	[-0.62353]	[-0.32808]
	Coefficient	-0.005807	0.185758	0.003626
Oil Price (-1)	Standard Error	(0.00527)	(0.07189)	(0.11067)
	t-statistics	[-1.10228]	[2.58375]	[0.03277]
	Coefficient	-0.013001	-0.130526	-0.048998
Oil Price (-2)	Standard Error	(0.00529)	(0.07214)	(0.11104)
	t-statistics	[-2.45969]	[-1.80937]	[-0.44126]
	Coefficient	0.004907	-0.075517	0.086585
Oil Price (-3)	Standard Error	(0.00520)	(0.07103)	(0.10934)
	t-statistics	[0.94282]	[-1.06315]	[0.79191]
	Coefficient	-0.015110	-0.447652	-0.124437
Oil Price (-4)	Standard Error	(0.00571)	(0.07793)	(0.11995)
	t-statistics	[-2.64639]	[-5.74446]	[-1.03739]
	Coefficient	0.004998	0.079896	0.022445
Stock Market (-1)	Standard Error	(0.00312)	(0.04259)	(0.06555)
	t-statistics	[1.60176]	[1.87600]	[0.34238]
	Coefficient	-0.007472	-0.146986	-0.121368
Stock Market (-2)	Standard Error	(0.00314)	(0.04281)	(0.06589)
	t-statistics	[-2.38231]	[-3.43370]	[-1.84194]
Stock Market (-3)	Coefficient	0.002464	-0.004380	-0.093196

	Standard Error	(0.00321)	(0.04386)	(0.06751)
	t-statistics	[0.76668]	[-0.09986]	[-1.38046]
Stock Market (-4)	Coefficient	0.005093	0.109791	-0.220496
	Standard Error	(0.00327)	(0.04460)	(0.06866)
	t-statistics	[1.55829]	[2.46142]	[-3.21148]
Constant	Coefficient	0.464575	5.377291	5.607402
	Standard Error	(0.10842)	(1.47974)	(2.27771)
	t-statistics	[4.28491]	[3.63395]	[2.46185]
R-Squared		0.68	0.23	0.10
Adjusted R-Squared		0.67	0.19	0.05
F statistic		0.00	0.00	0.00

Table 3 explains the details of 3 different models created in the VAR analysis. Additionally, mathematical expressions of these models are given on the equations (1)-(3).

Inflation = C(1,1)*Inflation(-1) + C(1,2)*Inflation(-2) + C(1,3)*Inflation(-3) + C(1,4)*Inflation(-4) + C(1,5)*Oil(-1) + C(1,6)*Oil(-2) + C(1,7)*Oil(-3) + C(1,8)*Oil(-4) + C(1,9)*SM(-1) + C(1,10)*SM(-2) + C(1,11)*SM(-3) + C(1,12)*SM(-4) + C(1,13) (1)

Oil = C(2,1)*Inflation(-1) + C(2,2)*Inflation(-2) + C(2,3)*Inflation(-3) + C(2,4)*Inflation(-4) + C(2,5)*Oil(-1) + C(2,6)*Oil(-2) + C(2,7)*Oil(-3) + C(2,8)*Oil(-4) + C(2,9)*SM(-1) + C(2,10)*SM(-2) + C(2,11)*SM(-3) + C(2,12)*SM(-4) + C(2,13) (2)

SM = C(3,1)*Inflation(-1) + C(3,2)*Inflation(-2) + C(3,3)*Inflation(-3) + C(3,4)*Inflation(-4) + C(3,5)*Oil(-1) + C(3,6)*Oil(-2) + C(3,7)*Oil(-3) + C(3,8)*Oil(-4) + C(3,9)*SM(-1) + C(3,10)*SM(-2) + C(3,11)*SM(-3) + C(3,12)*SM(-4) + C(3,13) (3)

On the other side, the values of the coefficients are also calculated. The details of these factors are given on Table 4-6.

Table 4. The Details of Coefficients for Model 1

Dependent Variable	Independent Variables (IV)	Symbols of IV	Coefficients	Probability Values
	Inflation (-1)	C(1,1)	0.735672	0.0000
	Inflation (-2)	C(1,2)	0.063992	0.4902
	Inflation (-3)	C(1,3)	-0.202525	0.0222
	Inflation (-4)	C(1,4)	0.159163	0.0046
	Oil Price (-1)	C(1,5)	-0.005807	0.2707
	Oil Price (-2)	C(1,6)	-0.013001	0.0142
Inflation	Oil Price (-3)	C(1,7)	0.004907	0.3461
	Oil Price (-4)	C(1,8)	-0.015110	0.0083
	Stock Market (-1)	C(1,9)	0.004998	0.1097
	Stock Market (-2)	C(1,10)	-0.007472	0.0175
	Stock Market (-3)	C(1,11)	0.002464	0.4435
	Stock Market (-4)	C(1,12)	0.005093	0.1196
	Constant Term	C(1,13)	0.464575	0.0000

Table 4 gives information about the model in which inflation is the dependent variable. The first hypothesis in this study is that oil price has an effect on the inflation rate. Therefore, the coefficients of the oil price variable are considered. The probability values of C(1,6) and C(1,8) are smaller than 0.05, so it is clear that these variables are significant. However, the coefficients of them are negative. This means that oil price increase has a decreasing but small effect on the inflation in the future periods.

Table 5. The Details of Coefficients for Model 2

Dependent Variable	Independent Variables (IV)	Symbols of IV	Coefficients	Probability Values
	Inflation (-1)	C(2,1)	-2.844556	0.0047
	Inflation (-2)	C(2,2)	1.806344	0.1538
	Inflation (-3)	C(2,3)	0.088316	0.9416
	Inflation (-4)	C(2,4)	-0.476162	0.5331
	Oil Price (-1)	C(2,5)	0.185758	0.0100
	Oil Price (-2)	C(2,6)	-0.130526	0.0708
Oil Price	Oil Price (-3)	C(2,7)	-0.075517	0.2881
	Oil Price (-4)	C(2,8)	-0.447652	0.0000
	Stock Market (-1)	C(2,9)	0.079896	0.0611
	Stock Market (-2)	C(2,10)	-0.146986	0.0006
	Stock Market (-3)	C(2,11)	-0.004380	0.9205
	Stock Market (-4)	C(2,12)	0.109791	0.0141
	Constant Term	C(2,13)	5.377291	0.0003

Table 5 explains the situation where dependent variable is the oil price. That is to say, the influencing factors of oil price are considered in this model. This issue is not related to the hypotheses in this study. It is seen that C(2,5), C(2,8) and C(2,12) are significant. It means that oil price increases in the previous periods lead to decrease in this price in current period.

Table 6. The Details of Coefficients for Model 3

Dependent Variable	Independent Variables (IV)	Symbols of IV	Coefficients	Probability Values
	Inflation (-1)	C(3,1)	-2.921181	0.0592
	Inflation (-2)	C(3,2)	0.683050	0.7259
	Inflation (-3)	C(3,3)	1.708924	0.3575
	Inflation (-4)	C(3,4)	-0.385655	0.7430
	Oil Price (-1)	C(3,5)	0.003626	0.9739
	Oil Price (-2)	C(3,6)	-0.048998	0.6592
	Oil Price (-3)	C(3,7)	0.086585	0.4287
Stock Market	Oil Price (-4)	C(3,8)	-0.124437	0.2999
	Stock Market (-1)	C(3,9)	0.022445	0.7322
	Stock Market (-2)	C(3,10)	-0.121368	0.0659
	Stock Market (-3)	C(3,11)	-0.093196	0.1679
	Stock Market (-4)	C(3,12)	-0.220496	0.0014
	Constant Term	C(3,13)	5.607402	0.0141
	Constant Term	C(3,13)	5.607402	0.0141

In the model of Table 6, stock market is the dependent variable. The second hypothesis in this study is that oil price affects stock market. Therefore, the coefficients of C(3,5), C(3,6), C(3,7) and C(3,8) are important for this situation. It is defined that probability values of all these variables are higher than 0.05. It is concluded that oil price does not have important impact on the stock market for G7 economies. In addition, the stationarity of these models is evaluated to see the appropriateness. Figure 1 indicates this situation.

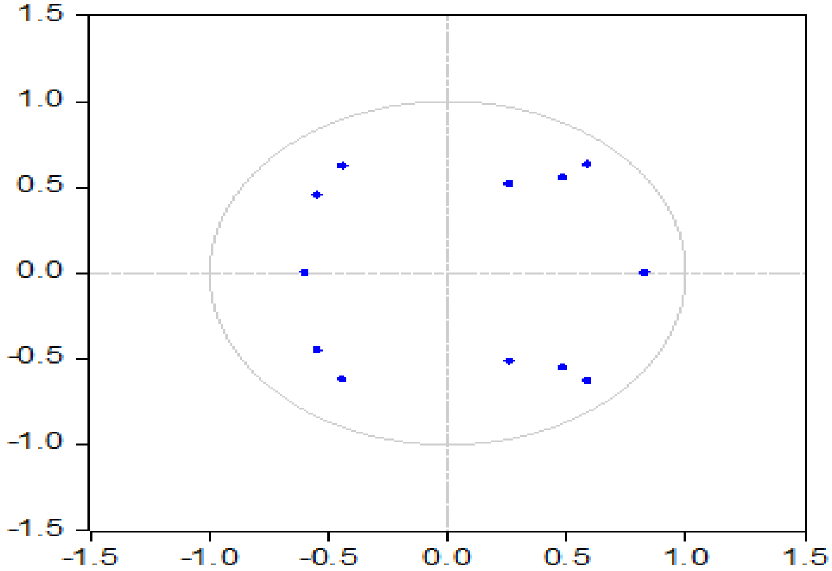


Figure 1. AR Root Graph

Because all points are on the boundary of the circle, it is concluded that the models are appropriate. In the next step, impulse responses are examined. This situation is illustrated on Figure 2.

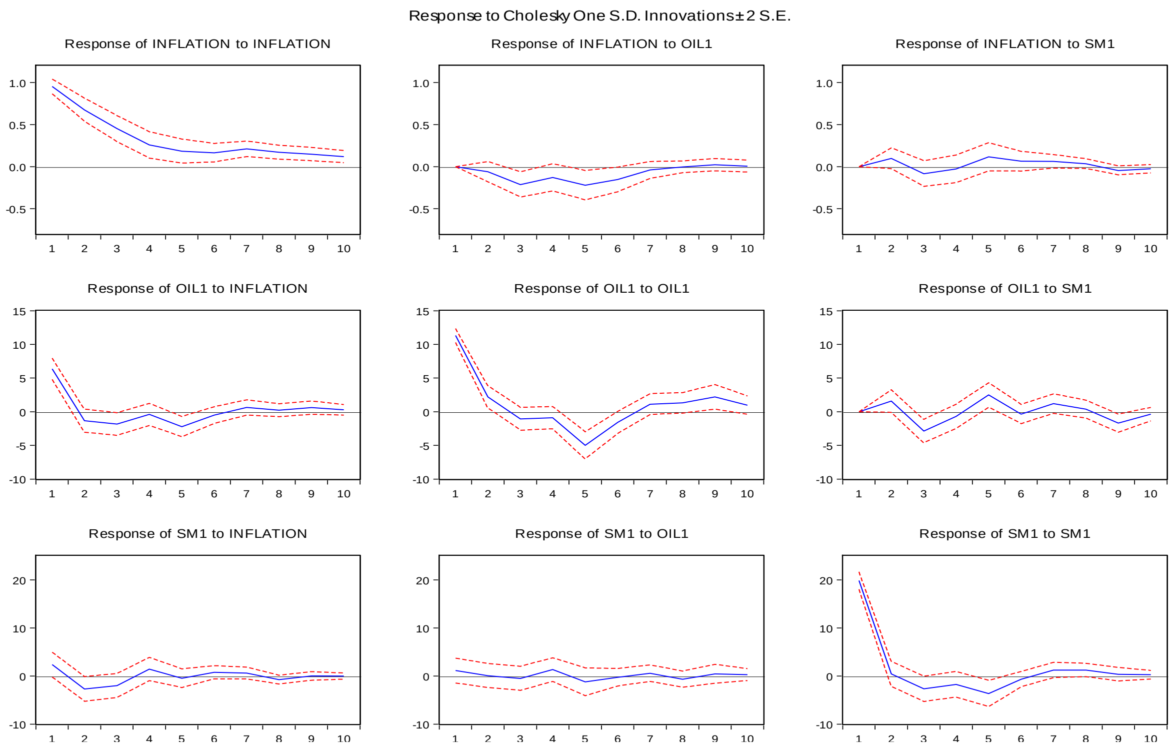


Figure 2. Impulse Responses of Variables

There are 9 different graphs in Figure 2. By considering the hypotheses of the study, the middle graphs on the first and third rows should be evaluated. These figures state that oil price changes do not have any effect on inflation rate and stock market. Furthermore, variance decomposition tables are created for three different variables. The results are given on Table 7-9.

Table 7: Variance Decomposition Table of Inflation

Period	Standard Error	Inflation	Oil Price	Stock Market
1	0.952504	100.0000	0.000000	0.000000
2	1.173499	99.02279	0.263223	0.713986
3	1.277889	95.99909	2.984391	1.016523
4	1.310134	95.21250	3.781071	1.006430
5	1.346437	92.02490	6.253039	1.722065
6	1.366519	90.81021	7.294888	1.894906
7	1.384819	90.76436	7.183994	2.051642
8	1.395869	90.84376	7.070972	2.085272
9	1.404629	90.83173	7.008301	2.159964
10	1.409895	90.86747	6.958613	2.173917

Table 7 states that inflation is mainly explained by itself. In the other hand, it is seen that oil price volatility does not play a significant role in the explanation of the inflation rate.

Table 8: Variance Decomposition Table of Oil Price

Period	Standard Error	Inflation	Oil Price	Stock Market
1	12.99980	24.01410	75.98590	0.000000
2	13.34581	23.78897	74.80054	1.410498
3	13.81527	23.98274	70.39126	5.626006
4	13.86761	23.88570	70.28087	5.833433
5	15.11125	22.26964	70.10022	7.630136
6	15.20679	22.09905	70.31248	7.588468
7	15.30855	21.97464	69.91163	8.113732
8	15.37067	21.81691	70.07544	8.107651
9	15.63289	21.24203	69.73915	9.018819
10	15.66969	21.17469	69.79531	9.029999

It can be understood from Table 8 that inflation has more impact on oil price by comparing with the stock market.

Table 9: Variance Decomposition Table of Stock Market

Period	Standard Error	Inflation	Oil Price	Stock Market
1	20.01020	1.396259	0.312679	98.29106
2	20.19739	3.165850	0.307984	96.52617
3	20.47760	4.021838	0.360606	95.61756
4	20.64489	4.445898	0.766984	94.78712
5	21.00789	4.344346	1.072327	94.58333
6	21.03404	4.465176	1.087115	94.44771
7	21.08649	4.520958	1.156687	94.32236
8	21.14710	4.623754	1.255587	94.12066
9	21.15476	4.620410	1.299367	94.08022
10	21.15793	4.619407	1.314446	94.06615

Table 9 explains that oil price has a very small effect on the stock market. Consequently, it is understood that in G7 economies, volatility in oil prices does not have an essential influence on financial development and macroeconomic stability.

4 DISCUSSION AND CONCLUSION

This study analyzes the impact of oil prices on financial markets and macroeconomic stability. The mentioned study was carried out for G7 countries. The ratio of stock prices to GDP representing the development in the financial market has been taken into consideration. On the other hand, the inflation rate variable is used in relation to macroeconomic stability. In this study, 3 different models are established for each variable with the help of VAR analysis. In this process, firstly, the delay length analysis was made. After that, modeler was developed, and coefficient analyzes were made. In addition, it is aimed to reach a detailed result by performing variance decomposition and impulse response analyzes.

As a result, it was determined that the change in oil prices for G7 countries did not have a serious effect on stock prices and inflation. In other words, a possible increase in oil prices does not adversely affect financial markets and macroeconomic stability in these countries. When these issues are taken into consideration, it is seen that the financial and macroeconomic structure in G7 countries is sound. Therefore, volatility in oil prices will not cause great damage to the economies of this country. This situation is a guideline for the managers of this country on behalf of risk management.

The results obtained from this study are guiding the governments of this country in other matters. For example, it will be more accurate to consider variables other than oil prices in policies to be implemented for inflation. In parallel with this mentioned issue, if the financial system is aimed to be developed in these countries or if the issues that affect these markets are analyzed, using variables other than oil prices may yield more meaningful results. Therefore, it is thought that the results obtained from this study can significantly support projects to be carried out on financial and macroeconomic stability in these countries.

In the literature, most of the studies identified that there should be a relationship between oil price volatility and inflation rates. However, in this study, it is concluded that this relationship is not valid for G7 economies. It is obvious that there are also some studies in the literature which supported this view. As an example, Salisu et al. [25] reached a conclusion that this relationship is not valid for oil exporting countries. Similarly, Bec and De Gaye [26] also defined that oil price shocks do not have a strong effect on the inflation in US, French and UK. Moreover, Choi et al. [27] made a comparative analysis for advanced and developing economies and identified that this relationship is not valid for advanced economies.

The biggest constraint in this study is to examine the effect of oil prices on only two variables. Oil prices are an important factor that can affect many different variables. Therefore, new studies can focus on different variables. On the other hand, in this study, the impact of oil prices on financial and macroeconomic stability has been taken into consideration only for G7 countries. The results of the analysis obtained may differ for other country groups. Therefore, it is considered that similar analyzes will be beneficial for E7 countries and energy importing countries.

REFERENCES

[1] Zhong, J., Hu, X., Yüksel, S., Dinçer, H., & Ubay, G. G. (2020). Analyzing the investments strategies for renewable energies based on multi-criteria decision model. *IEEE Access*, *8*, 118818-118840.

[2] Qiu, D., Dinçer, H., Yüksel, S., & Ubay, G. G. (2020). Multi-Faceted Analysis of Systematic Risk-Based Wind Energy Investment Decisions in E7 Economies Using Modified Hybrid Modeling with IT2 Fuzzy Sets. *Energies*, *13*(6), 1423.

[3] Liu, Y., Gong, X., Yüksel, S., Dinçer, H., & Aydın, R. (2021). A multidimensional outlook to energy investments for the countries with continental shelf in East Mediterranean Region with Hybrid Decision Making Model based on IVIF logic. *Energy Reports*, *7*, 158-173.

[4] Du, L., Dinçer, H., Ersin, İ., & Yüksel, S. (2020). IT2 Fuzzy-Based Multidimensional Evaluation of Coal Energy for Sustainable Economic Development. *Energies*, *13*(10), 2453.

[5] Dincer, H., Yüksel, S., & Martinez, L. (2019). Balanced scorecard-based Analysis about European Energy Investment Policies: A hybrid hesitant fuzzy decision-making approach with Quality Function Deployment. *Expert Systems with Applications*, *115*, 152-171.

[6] Zhou, P., Zhou, P., Yüksel, S., Dinçer, H., & Uluer, G. S. (2020). Balanced scorecard-based evaluation of sustainable energy investment projects with it2 fuzzy hybrid decision making approach. *Energies*, *13*(1), 82.

[7] Dinçer, H., & Yüksel, S. (2019). Multidimensional evaluation of global investments on the renewable energy with the integrated fuzzy decision-making model under the hesitancy. *International Journal of Energy Research*, *43*(5), 1775-1784.

[8] Yüksel, S., Ubay, G. G., & Sezer, D. (2020). Determining The Influence of Oil Prices On Economic Growth And Financial Development: An Analysis for Turkey with VAR Methodology. *Ekonomik ve Sosyal Araştırmalar Dergisi*, *1*(1), 1-23.

[9] Qi, W., Huang, Z., Dinçer, H., Korsakienė, R., & Yüksel, S. (2020). Corporate Governance-Based Strategic Approach to Sustainability in Energy Industry of Emerging Economies with a Novel Interval-Valued Intuitionistic Fuzzy Hybrid Decision Making Model. *Sustainability*, *12*(8), 3307.

[10] Mikayilov, J. I., Mukhtarov, S., Dinçer, H., Yüksel, S., & Aydın, R. (2020). Elasticity Analysis of Fossil Energy Sources for Sustainable Economies: A Case of Gasoline Consumption in Turkey. *Energies*, *13*(3), 731.

[11] Jeong, H. Y., Song, B. D., & Lee, S. (2019). Truck-drone hybrid delivery routing: Payload-energy dependency and No-Fly zones. *International Journal of Production Economics*, *214*, 220-233.

[12] Camprubí, L. (2019). Whose self-sufficiency? Energy dependency in Spain from 1939. *Energy policy*, *125*, 227-234.

[13] Jarrett, U., Mohaddes, K., & Mohtadi, H. (2019). Oil price volatility, financial institutions and economic growth. *Energy policy*, *126*, 131-144.

[14] Van Eyden, R., Difeto, M., Gupta, R., & Wohar, M. E. (2019). Oil price volatility and economic growth: Evidence from advanced economies using more than a century's data. *Applied energy*, *233*, 612-621.

[15] Taghizadeh-Hesary, F., Rasoulinezhad, E., & Yoshino, N. (2019). Energy and food security: Linkages through price volatility. *Energy policy*, *128*, 796-806.

[16] Živkov, D., Đurašković, J., & Manić, S. (2019). How do oil price changes affect inflation in Central and Eastern European countries? A wavelet-based Markov switching approach. *Baltic Journal of Economics*, *19*(1), 84-104.

[17] Ftiti, Z., & Jawadi, F. (2019). Forecasting inflation uncertainty in the United States and Euro Area. *Computational Economics*, *54*(1), 455-476.

[18] Conrad, C., & Hartmann, M. (2019). On the determinants of long-run inflation uncertainty: Evidence from a panel of 17 developed economies. *European Journal of Political Economy*, *56*, 233-250.

[19] Shahrestani, P., & Rafei, M. (2020). The impact of oil price shocks on Tehran Stock Exchange returns: Application of the Markov switching vector autoregressive models. *Resources Policy*, *65*, 101579.

[20] Al-hajj, E., Al-Mulali, U., & Solarin, S. A. (2020). Exploring the nexus between oil price shocks and sectoral stock returns: a new evidence from stock exchange in Malaysia. *Economic Change and Restructuring*, 1-19.

[21] Nusair, S. A. (2019). Oil price and inflation dynamics in the Gulf Cooperation Council countries. *Energy*, *181*, 997-1011.

[22] Albulescu, C. T., Oros, C., & Tiwari, A. K. (2017). Oil price–inflation pass-through in Romania during the inflation targeting regime. *Applied Economics*, *49*(15), 1527-1542.

[23] Hegerty, S. W. (2016). Commodity-price volatility and macroeconomic spillovers: Evidence from nine emerging markets. *The North American Journal of Economics and Finance*, *35*, 23-37.

[24] Degirmen, S., & Saltik, O. (2017). Impacts of realized volatility of oil price over foreign trade related activities in Turkey. *Economic Change and Restructuring*, *50*(3), 193-209.

[25] Salisu, A. A., Isah, K. O., Oyewole, O. J., & Akanni, L. O. (2017). Modelling oil price-inflation nexus: The role of asymmetries. *Energy*, *125*, 97-106.

[26] Bec, F., & De Gaye, A. (2016). How do oil price forecast errors impact inflation forecast errors? An empirical analysis from US, French and UK inflation forecasts. *Economic Modelling*, *53*, 75-88.

[27] Choi, S., Furceri, D., Loungani, P., Mishra, S., & Poplawski-Ribeiro, M. (2018). Oil prices and inflation dynamics: Evidence from advanced and developing economies. *Journal of International Money and Finance*, *82*, 71-96.

[28] Luo, X., & Qin, S. (2017). Oil price uncertainty and Chinese stock returns: New evidence from the oil volatility index. *Finance Research Letters*, *20*, 29-34.

[29] Aydoğan, B., Tunç, G., & Yelkenci, T. (2017). The impact of oil price volatility on net-oil exporter and importer countries' stock markets. *Eurasian Economic Review*, *7*(2), 231-253.

[30] Diaz, E. M., Molero, J. C., & de Gracia, F. P. (2016). Oil price volatility and stock returns in the G7 economies. *Energy Economics*, *54*, 417-430.

[31] Kristjanpoller, W., & Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. *Expert Systems with Applications*, *65*, 233-241.

[32] Yu, Z., Liu, W., Chen, L., Eti, S., Dinçer, H., & Yüksel, S. (2019). The effects of electricity production on industrial development and sustainable economic growth: A VAR analysis for BRICS countries. *Sustainability*, *11*(21), 5895.

[33] Kalkavan, H., Eti., S., & Yüksel, S. (2020). Türkiye'deki Bankacılık Sektörü, Sanayi Gelişimi ve Ekonomik Büyüme Arasındaki İlişkinin VAR Analizi İle İncelenmesi. *Akademik Araştırmalar ve Çalışmalar Dergisi (AKAD)*, *12*(22), 56-74.

[34] Gupta, R., Lau, C. K. M., & Wohar, M. E. (2019). The impact of US uncertainty on the Euro area in good and bad times: evidence from a quantile structural vector autoregressive model. *Empirica*, *46*(2), 353-368.

[35] Gupta, R., Lau, C. K. M., Plakandaras, V., & Wong, W. K. (2019). The role of housing sentiment in forecasting US home sales growth: evidence from a Bayesian compressed vector autoregressive model. *Economic research-Ekonomska istraživanja*, *32*(1), 2554-2567.

[36] Koop, G., & Korobilis, D. (2016). Model uncertainty in panel vector autoregressive models. *European Economic Review*, *81*, 115-131.

[37] Khraief, N., Shahbaz, M., Heshmati, A., & Azam, M. (2020). Are unemployment rates in OECD countries stationary? Evidence from univariate and panel unit root tests. *The North American Journal of Economics and Finance*, *51*, 100838.

[38] Dinçer, H., Yüksel, S., Eti, S., & Tula, A. (2019). Effects of demographic characteristics on business success: an evidence from Turkish banking sector. In *Handbook of research on business models in modern competitive scenarios* (pp. 304-324). IGI Global.

[39] Das, P. (2019). Panel Unit Root Test. In *Econometrics in Theory and Practice* (pp. 513-540). Springer, Singapore.

[40] Yüksel, S., Dinçer, H., & Uluer, G. S. (2020). The Role of Technological Development on Renewable Energy Usage: An Econometric Analysis for G7 Countries. In *Handbook of Research on Sustainable Supply Chain Management for the Global Economy* (pp. 136-153). IGI Global.

[41] Dutta, U. P., Gupta, A., & Sengupta, P. P. (2019). Exploring the Nexus Between Mobile Phone Penetration and Economic Growth in 13 Asian Countries: Evidence from Panel Cointegration Analysis. In *Soft Computing and Signal Processing* (pp. 337-346). Springer, Singapore.

[42] Yüksel, S., Dinçer, H., Karakuş, H., & Ubay, G. G. (2020). The Negative Effects of Carbon Emission on FDI: A Comparative Analysis Between E7 and G7 Countries. In *Handbook of Research on Sustainable Supply Chain Management for the Global Economy* (pp. 20-35). IGI Global.

Journal of Statistics & Applied Science, Volume - 1, Issue - 2
İstatistik ve Uygulamalı Bilimler Dergisi, Cilt - 1, Sayı- 2

A Statistical Analysis of the Relationship Between Meteorological Parameters and the Spread of COVID-19 Cases: Comparison Between Turkey and Italy

Yasemin Ayaz Atalan[1], Abdulkadir Atalan[2*]

[1]Bozok University, Faculty of Engineering and Architecture, Yozgat, Turkey
[2]Gaziantep Islam Science and Technology University, Faculty of Engineering and Natural Sciences, Gaziantep, Turkey

*Corresponding Author, e-mail: abdulkadir.atalan@gibtu.edu.tr

Abstract

This study aims to statistically analyze the effects of meteorological parameters on the spread of coronavirus in Turkey and Italy. The multi-factor ANOVA for data analysis was used for the present study. The data of the COVID-19 active cases were handled covering 55 days between March 10, 2020 (the first case incident) and May 3, 2020 for Turkey and covering 69 days between February 25, 2020, and May 3, 2020 for Italy. The parameters of meteorology include average temperature (^{0}C), humidity (%), wind (mph) and pressure (Hg) in this study. The data of meteorological parameters were considered as the average of each factor's data for the highest populated cities of Turkey and the two most populous cities (Rome and Milan) in Italy. The analysis of variance was significantly analyzed with COVID-19 pandemic (R^2= 0.307; F-ratio=5.6581 prob > .0008 with 55 observations for Turkey, and R^2= 0.437; F-ratio=3.6581 prob > .0182 with 69 observations for Italy), with the highest level. Among the parameters of the weather, average temperature (°C) has a significant impact on COVID-19 pandemic (t-ratio=3.12; $p<0.05$) and humidity (%) provisionally affects COVID-19 pandemic (t-ratio=-1.186; $p<0.10$) for Turkey. In addition, both humidity (%) (t-ratio=-1,38; $p<0.0172$) and wind (mph) (t-ratio=-2,57; $p<0.0125$) parameters have been found to play an important role in the COVID-19 outbreak for Italy.

Keywords: COVID-19; Anova; Temperature; Humidity; Wind; Pressure

Özet

Bu çalışma, meteorolojik parametrelerin Türkiye ve İtalya ülkelerinde korona virüsün yayılmasına etkilerini istatistiksel olarak incelemeyi amaçlamaktadır. Veri analizi için çok faktörlü ANOVA testi bu çalışmada kullanılmıştır. Türkiye için COVID-19 aktif vakalarına ait veriler, 10 Mart 2020 (ilk vaka olayı) ile 3 Mayıs 2020 arasındaki 55 günlük ve İtalya için COVID-19 aktif vakalarına ait veriler, 25 Şubat 2020 ile 3 Mayıs 2020 arasındaki 69 günlük verileri kapsamaktadır. Bu çalışmada, meteoroloji parametreleri olarak ortalama sıcaklık (^{0}C), nem (%), rüzgâr (mph) ve basınç (Hg) faktörleri ele alınmıştır. Meteorolojik parametre verileri, Türkiye'nin en kalabalık şehirlerindeki her bir parametreye ait verilerin ortalaması ile İtalya'nın en kalabalık iki şehrine (Roma ve Milan) ait veriler dikkate alınmıştır. Varyans analizi kullanılarak COVID-19 vakalarına ait verileri (Türkiye için R^2 = 0.307; F-oranı = 5.6581 prob> .0008, 55 gözlem ve İtalya için R^2= 0.437; F-ratio=3.6581 prob> .0182, 69 gözlem) en yüksek düzeyde önemli ölçüde analiz edilmiştir. Meteoroloji parametreleri arasında yer alan ortalama sıcaklık (°C) (t-oranı = 3.12; $p<0.05$) önemli ölçüde ve nem (%) faktörü (t-oranı =-1.186; $p<0.10$) şartlı olarak Türkiye'deki COVID-19 salgınını etkilediği gözlemlenmiştir. Ayrıca, İtalya için hem nem (%) (t-ratio=-1,38; $p<0.0172$) hem de rüzgâr (mph) (t-ratio=-2,57; $p<0.0125$) parametrelerinin COVID-19 salgınında önemli rol oynadığı tespit edilmiştir.

Anahtar Kelimeler: COVID-19; Anova; Sıcaklık; Nem; Rüzgar; Basınç

1. INTRODUCTION

Described as a disease similar to pneumonia cases, the C0VID-19 virus first appeared in Wuhan City, Hubei Province, China, in December 2019 (Atalan 2020a; Nghiem et al. 2020; Saglietto et al. 2020; Tobías et al. 2020). With the rapid increase in the number of cases, research showed that there is a new type of coronavirus that has not been identified before. This virus has been named Coronavirus-2019, SARS nCoV-2 or COVID-19 by health organizations with its emergence in December 2019 (Paital, Das, and Parida 2020). Although there are many reasons for the transmission of this virus from person to person, the main reason is transmitted by air (weather) (Ministry of Health 2020). Furthermore, meteorological factors play an important role in the rapid spread of this virus (Atalan 2020b; Chen et al. 2020; Liu et al. 2020; Shi et al. 2020; Tosepu et al. 2020). However, there is no clearly proven (still a controversial situation) study on whether meteorological factors have a direct or indirect effect on the spread of COVID-19.

Factors such as temperature, humidity, pressure, wind speed, the amount of rainfall density that are thought to be effective on coronavirus are widely discussed in the studies. Different statistical methods were used for these factors in the literature (Cássaro and Pires 2020; Liu et al. 2020). Tosepu et al. determined that temperature has an important effect on COVID-19 by using spearman's correlation method by considering temperature (^{0}C), humidity (%) and the amount of rainfall (mm) (Tosepu et al. 2020). Shi et al. Measured the effect of temperature and humidity on COVID-19 using the modified susceptible-exposed-infectious-recovered method (M-SEIR) (Shi et al. 2020). They concluded that only temperature can have an effect on COVID-19 with this method. Wang et al. found that high temperature and high humidity significantly reduced COVID-19 pandemic with the linear regression analysis (Wang et al. 2020). Tobias et al. have found a slowdown in spread of COVID-19 pandemic during rainy days (Tobías et al. 2020). These studies show that meteorological factors have been emphasized to have a direct or indirect effect on COVID-19.

Most studies have focused on the correlation between factors (Atalan 2018; Ayaz Atalan et al. 2020). Calculating the correlation coefficients does not mean that a factor exerts its effect on a response variable. A second analysis in statistics is needed to see the effect of factors on responses (Dönmez and Atalan 2019). In this study, Anova was performed besides the correlation test to measure the effect of temperature (^{0}C), humidity (%), wind (mph) and pressure (Hg) factors on COVID-19 for Turkey and Italy. Turkey has announced the first COVID-19 case on 10 March 2020. The total number of COVID-19 cases is 126,045 until May 3, 2020 (Ministry of Health 2020). The number of people who died due to COVID-19 is 3397 until May 3, 2020 (Ministry of Health 2020). The data of the COVID-19 active cases are handled for 69 days of data covering between February 25, 2020 (The first day of the announcement of the COVID-19 case) and May 3, 2020 in Italy. In this study, deaths for both countries were not included in the statistical analysis.

This study includes four sections. The first section deals with the literature review of studies related to COVID-19 pandemic. The second part gives detailed information about the methodology of the study. The results obtained from the method mentioned in the methodology section were discussed in the third section. In the last section, conclusion about the study has been provided.

2. METHODOLOGY

In this study, the most populous in terms of density cities in Turkey and Italy of the COVID-19 cases and the meteorological parameters data were discussed. The data used in this study were handled as 55 days of data covering between March 10, 2020 (the first case incident) and May 3, 2020 for Turkey. The number of data collected is a sufficient rate for statistical analysis. In this study, the correlation test was employed with the multi-factor Anova to measure the effect of factors on the number of the COVID-19 cases. COVID-19 cases were defined as the output (or response) factor in the Anova analysis. Data of the regions covered Turkey has a population of 83.15 million (Turkish Statitical Institute 2020). Turkey consists of 81 cities and 7 geographic regions which are Marmara, Aegean, Central Anatolia, Black Sea, Mediterranean, Southeast Anatolia, and Eastern Anatolia. Turkey is also perceived as a bridge between Asia and Europe so the rate of population permeability (being a transfer point especially in air transportation change) is too much. Therefore, Turkey is considered as an autonomous region for COVID-19 studies.

The 55-day COVID-19 case data are shown in figure 1 (Worldometer 2020). According to **Fig.1**., the number of COVID-19 cases should be examined in two parts. While the number of COVID-19 cases increased daily before the 33rd day, the number of COVID-19 cases decreased after that date. The peak point of COVID-19 cases was recorded as 5138 on 11 April 2020 (temperature, 11.83 ^{0}C; humidity, 61.10%; wind, 7.07mph; pressure, 28.30Hg in Turkey. The average number of cases was calculated as 2292.

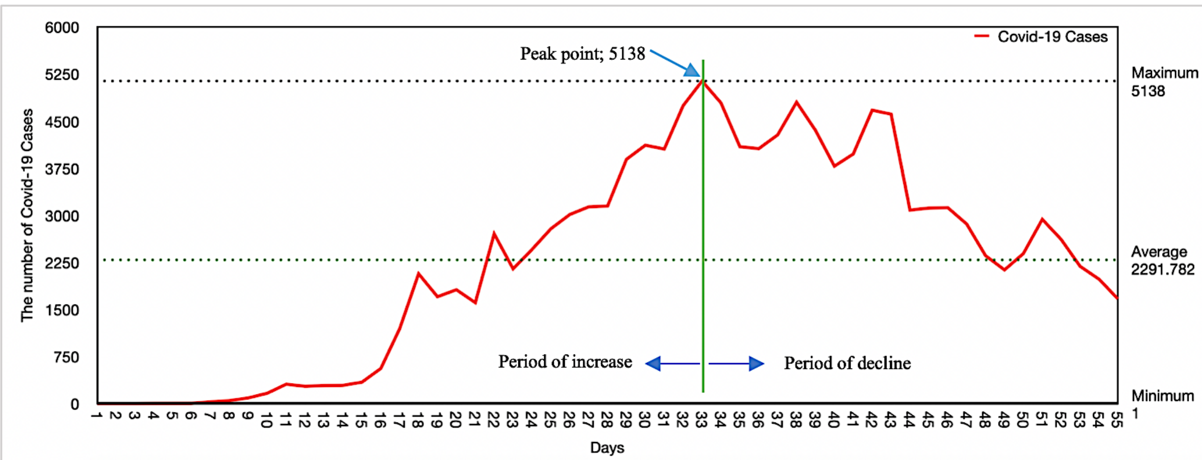


Fig. 1. COVID-19 case numbers by days

The cases of COVID-19 pandemic data are used as 69 days of data covering between February 25, 2020, and May 3, 2020, in Milan and Rome, Italy (Lab24 2020; Worldometer 2020) The daily cases of COVID-19 data for Rome and Milan are shown in **fig. 2.** (Worldometer 2020). The city of Rome is located in the Lazio region in Italy. The most populated city in Italy is Rome. Rome has an area of 5,363 km^2 and a population of 4.35 million (World Population Review 2020b). The city of Milan is located in the Lombardy region in the north of Italy. The second most populated city in Italy is Milan. Milan has an area of 1,575 km^2 and a population of 3.25 million (World Population Review 2020a) (See **fig.3**.).

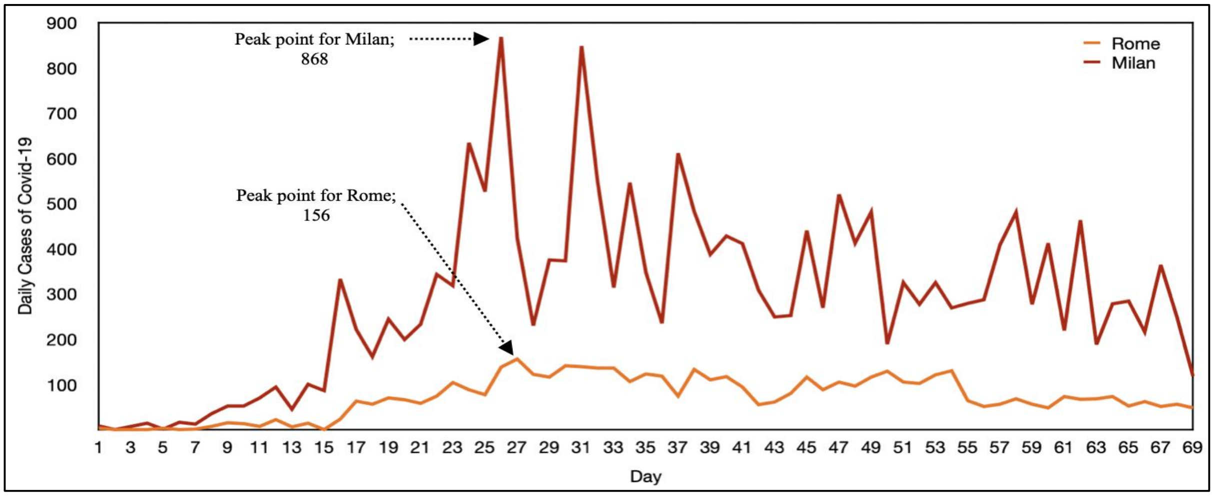


Fig. 2. Daily COVID-19 cases in Italy's most crowded cities Rome and Milan

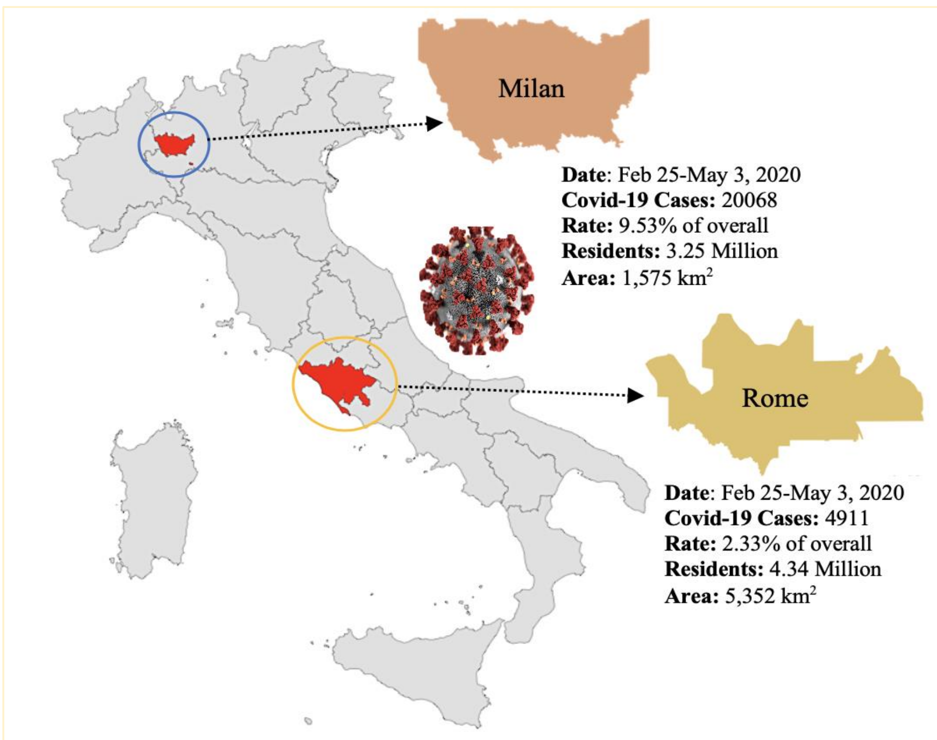
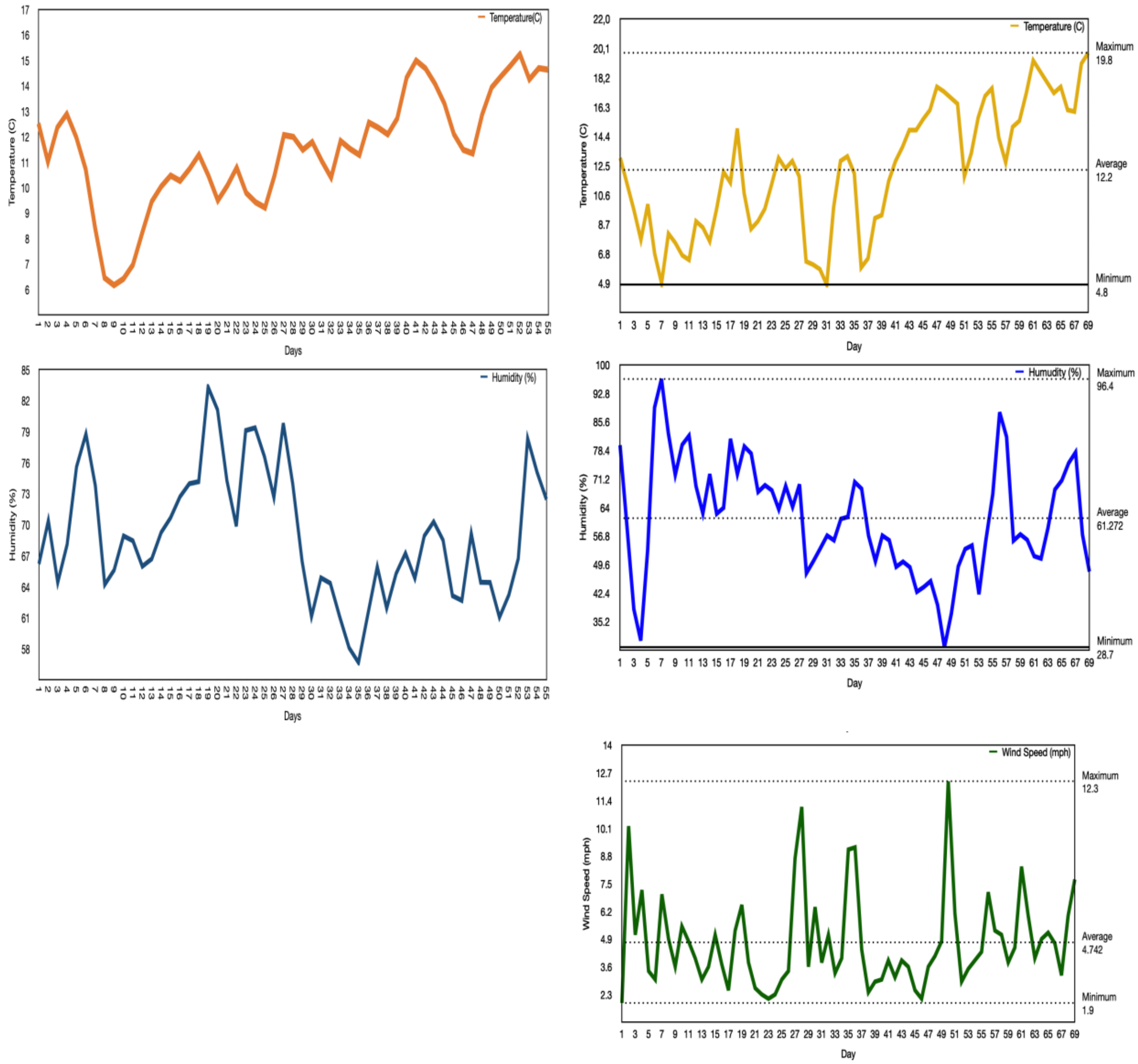
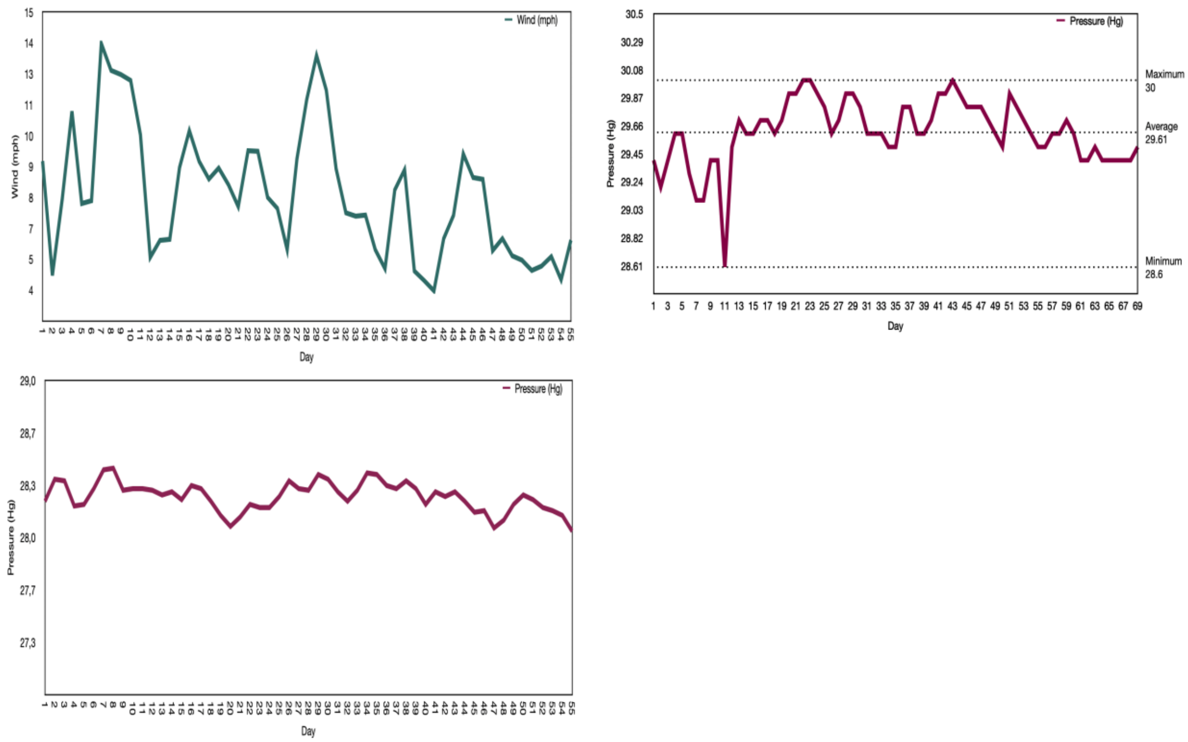


Fig. 3. The descriptive data of the cities in Italy

Meteorological parameters affecting the environmental balance can also be effective in the spread of the COVID-19 virus. Four different weather factors, namely temperature (^{0}C) humidity (%), wind (mph), and pressure (Hg*)* were considered that we think are effective on COVID-19 case data. The data of the weather parameters were calculated as the average value between the dates determined for the study. The average value of daily weather data on the largest cities located in seven regions of Turkey were collected to represent the entire region of Turkey. The reason for using this method is that countries with large borders do not have daily average weather data. The cities of Istanbul for the Marmara Region, Ankara for the

Central Anatolia region, Izmir for Aegean region, Adana for Mediterranean region, Samsun for Black Sea region, Van for Eastern Anatolia region and Diyarbakir for Southeast Anatolia region were determined to calculate the weather data for Turkey. The total population in these cities account for 38.48% of Turkey's population. Istanbul (15.52 million) is the most crowded city in Turkey and approximately 60% of COVID-19 cases is observed in this city. **Fig. 4.** shows the average value of the weather parameters of Turkey. Temperature(^{0}C) and humidity (%) rates are increased day by day, but wind speed (mph) and pressure (Hg) values are stable.





(a) Turkey (b) Italy

Fig. 4. Average values of meteorological parameters for Turkey (a) and Italy (b)

The peak of COVID-19 cases for Milan was observed as 868 COVID-19 cases on March 21, 2020 (with temperature, 12.8 ^{0}C; humidity, 64.1 %; wind, 3.4 mph; pressure, 29.6 Hg on the date of the peak point in Milan). The number of cases in Italy was recorded as 6557 daily new cases and 53578 total cases at the peak date for Milan. The average number of daily new cases was calculated as 291. The peak of COVID-19 cases for Milan was observed as 868 COVID-19 cases on March 22, 2020 (with temperature, 14.6 ^{0}C; humidity, 58.5 %; wind, 6.1 mph; pressure, 29.5 Hg on the date of the peak point in Rome). The number of cases was recorded as 5560 daily new cases and 46638 total cases at the peak date for Rome in Italy. The average number of daily new COVID-19 cases was calculated as 71 in Rome. Although Rome has a crowded population, the number of COVID-19 cases is less than the number of COVID-19 cases occurring in Milan.

The data of weather for Italy showed temperature minimum of 6.16 °C (with the cases of COVID-19, 93), the maximum temperature of 15.24 °C (with the cases of COVID-19, 2615), the lowest humidity of 56.67% (with the cases of COVID-19, 4093), the maximum humidity of 83.30% (with the cases of COVID-19, 1704), the lowest wind of 4.17 mph (with the cases of COVID-19, 3977), the maximum wind speed of 13.77 mph (with the cases of COVID-19, 31), and the minimum pressure of 28.04 Hg (with the cases of COVID-19, 1670), the maximum pressure of 28.44 Hg (with the cases of COVID-19, 49).

3. RESULTS AND DISCUSSIONS

The data set used is not suitable for normal distribution according to Anderson-Darling (0.9160) and Shapiro-Wilk (1.3412) tests. Therefore, ANOVA test was performed for this study, since many data sets that were not significantly normal would yield perfectly suitable results for an ANOVA. The lognormal and Weibull distributions provide a good fit for the cases of COVID-19 data for Turkey and Italy.

Table 1 shows the correlation values of weather parameters with each other. The correlation value varies between -1.00 and +1.00. The correlation value of a factor indicates that it has a negative relationship as it approaches -1.00, and a positive relationship as it approaches +1.00. Anova model has a positive relationship only with the wind (0.2125), while the model has a negative relationship with temperature (-0.2509), humidity (-0.4969), and pressure (-0.9992) for Turkey and a positive relationship only with the temperature (0.2430), while the model has a negative relationship with humidity (-0.5783), wind (-0.3973), and pressure (-0.9994) for Italy. The strongest relationship with cases of COVID-19 number negatively is with pressure for Turkey and Italy.

Table 1. Correlation values of meteorological parameters for Turkey and Italy

Country	**Correlation**	Model	Temp (^{0}C)	Humidity (%)	Wind (mph)	Pressure (Hg)
	Model	1,0000	-0,2509	-0,4969	0,2125	-0,9992
	Temp (C)	-0,2509	1,0000	0,2061	0,4816	0,2256
Turkey	Humidity (%)	-0,4969	0,2061	1,0000	-0,1423	0,4691
	Wind (mph)	0,2125	0,4816	-0,1423	1,0000	-0,2308
	Pressure (Hg)	-0,9992	0,2256	0,4691	-0,2308	1,0000
	Model	1,0000	0,2430	-0,5783	-0,3973	-0,9994
	Temp (C)	0,2430	1,0000	-0,2788	0,1116	-0,2632
Italy	Humidity (%)	-0,5783	-0,2788	1,0000	0,3058	0,5600
	Wind (mph)	-0,3973	0,1116	0,3058	1,0000	0,3791
	Pressure (Hg)	-0,9994	-0,2632	0,5600	0,3791	1,0000

The correlation between parameters does not mean that these parameters have an effect on COVID-19. Therefore, more than one analysis should be made in statistics. **Table 2** shows that, the analysis of variance was significantly analyzed with COVID-19 pandemic (R^2= 0.307; F-ratio=5.6581 prob > .0008 with 55 observations for Turkey, and R^2= 0.437; F-ratio=3.6581 prob > .0182 with 69 observations for Italy), with high level.

Among the parameters, temperature (^{0}C) has the most important effect on COVID-19 cases based on F-ratio and t-ratio (p=0.0037) for Turkey. We also can say that the humidity rate has a significant effect in terms of significance for Turkey. Since significance of humidity (%) is less than the value of p=0.1, its effect on COVID-19 is high. Tosepu et al. (2020) expressed the humidity (%) was insignificant correlated (Tosepu et al. 2020). This is because the data they use in their studies are statistically insufficient. Wind (mph) and pressure (Hg) parameters did not show a significant effect on COVID-19. Another study shows that temperature and humidity have an important and consistent distribution in the spread of the virus (Chen et al. 2020). Both Humidity (%) (t-ratio=-1,38; p<0.0172) and Wind (mph) (t-ratio=-2,57; p<0.0125) parameters have been found to play an important role in the COVID-19 outbreak for Italy.

Table 2. ANOVA data of the parameters for COVID-19

Country	Source	Std Error	Sum of Squares	t-Ratio	F-Ratio	Prob > F Prob > |t|
Turkey	Temperature (C)	72899.42	19001496	3.04	9.2659	0.0037
	Humidity (%)	109.5994	6933250	-1.84	3.3809	0.0719
	Wind (mph)	36.26466	6016.00	-0.05	0.0029	0.9570
	Pressure (Hg)	98.15116	735691.0	0.60	0.3588	0.5519

Italy	Temperature (C)	1.882879	0.182000	-0,01	0,0001	0.9920
	Humidity (%)	0.515344	3444.845	-1,38	1,9026	0.0172
	Wind (mph)	2.055759	11974.693	-2,57	6,6137	0.0125
	Pressure (Hg)	37.80559	412.66600	0,48	0,2279	0.6347

According to Prediction Profiler, the number of COVID-19 cases in the coming days will range from 2.2 to 4373.4. Temperature, humidity, wind, pressure values should be 15.24, 83.3, 13.77 and 28.001, respectively, within these limits.

4. CONCLUSION

In this study, the effects of meteorological parameters on the spread of coronavirus have been analyzed statistically in Turkey and Italy. The multi-factor ANOVA for data analysis was used for the present study. The data of the COVID-19 active cases were handled covering 55 days between March 10, 2020 (the first case incident) and May 3, 2020 for Turkey and covering 69 days between February 25, 2020, and May 3, 2020 for Italy.

The parameters of meteorology include average temperature (^{0}C), humidity (%), wind (mph) and pressure (Hg) in this study. The data of meteorological parameters were considered as the average of each factor's data for the highest populated cities of Turkey and the two most populous cities (Rome and Milan) in Italy. The analysis of variance was significantly analyzed with COVID-19 pandemic (R^2= 0.307; F-ratio=5.6581 prob > .0008 with 55 observations for Turkey, and R^2= 0.437; F-ratio=3.6581 prob > .0182 with 69 observations for Italy), with the highest level.
Among the parameters of the weather, average temperature (°C) has a significant impact on COVID-19 pandemic (t-ratio=3.12; $p<0.05$) and humidity (%) provisionally affects COVID-19 pandemic (t-ratio=-1.186; $p<0.10$) for Turkey. In addition, both humidity (%) (t-ratio=-1,38; $p<0.0172$) and wind (mph) (t-ratio=-2,57; $p<0.0125$) parameters have been found to play an important role in the COVID-19 outbreak for Italy.

Acknowledgements
The authors would like to thank the Lab24 weatherunderground.com, the Ministry of Health, Turkey and www.worldometer.com for sharing the dataset of weather parameters and the data of COVID-19.

Reference

Atalan, Abdulkadir. 2018. "Türkiye Sağlık Ekonomisi Için İstatistiksel Çok Amaçlı Optimizasyon Modelinin Uygulanması." *İşletme Ekonomi ve Yönetim Araştırmaları Dergisi* 1(1):34–51.

Atalan, Abdulkadir. 2020a. "Is the Lockdown Important to Prevent the COVID-19 Pandemic? Effects on Psychology, Environment and Economy-Perspective." *Annals of Medicine and Surgery* 56(2020):38–42.

Atalan, Abdulkadir. 2020b. "Is the Lockdown Important to Prevent the COVID-19 Pandemic? Effects on Psychology, Environment and Economy-Perspective." *Annals of Medicine and Surgery*.

Ayaz Atalan, Yasemin, Mete Tayanç, Kamil Erkan, and Abdulkadir Atalan. 2020. "Development of Nonlinear Optimization Models for Wind Power Plants Using Box-Behnken Design of Experiment: A Case Study for Turkey." *Sustainability* 12(15):6017.

Cássaro, Fábio A. M., and Luiz F. Pires. 2020. "Can We Predict the Occurrence of COVID-19 Cases? Considerations Using a Simple Model of Growth." *Science of The Total Environment* 728:138834.

Chen, Biqing, Hao Liang, Xiaomin Yuan, Yingying Hu, Miao Xu, Yating Zhao, Binfen Zhang, Fang Tian, and Xuejun Zhu. 2020. "Roles of Meteorological Conditions in COVID-19 Transmission on a Worldwide Scale." *MedRxiv*.

Dönmez, Cem Çağrı, and Abdulkadir Atalan. 2019. "Developing Statistical Optimization Models for Urban Competitiveness Index: Under the Boundaries of Econophysics Approach." *Complexity* 2019:1–11.

Lab24. 2020. "Coronavirus in Italia." Retrieved (https://lab24.ilsole24ore.com/coronavirus/).

Liu, Jiangtao, Ji Zhou, Jinxi Yao, Xiuxia Zhang, Lanyu Li, Xiaocheng Xu, Xiaotao He, Bo Wang, Shihua Fu, Tingting Niu, Jun Yan, Yanjun Shi, Xiaowei Ren, Jingping Niu, Weihao Zhu, Sheng Li, Bin Luo, and Kai Zhang. 2020. "Impact of Meteorological Factors on the COVID-19 Transmission: A Multi-City Study in China." *Science of The Total Environment* 726:138513.

Ministry of Health. 2020. "Covid 19." Retrieved (https://covid19.saglik.gov.tr/).

Nghiem, Long D., Branwen Morgan, Erica Donner, and Michael D. Short. 2020. "The COVID-19 Pandemic: Considerations for the Waste and Wastewater Services Sector." *Case Studies in Chemical and Environmental Engineering* 100006.

Paital, Biswaranjan, Kabita Das, and Sarat Kumar Parida. 2020. "Inter Nation Social Lockdown versus Medical Care against COVID-19, a Mild Environmental Insight with Special Reference to India." *Science of The Total Environment* 728:138914.

Saglietto, Andrea, Fabrizio D'Ascenzo, Giuseppe Biondi Zoccai, and Gaetano Maria De Ferrari. 2020. "COVID-19 in Europe: The Italian Lesson." *The Lancet* 395(10230):1110–11.

Shi, Peng, Yinqiao Dong, Huanchang Yan, Xiaoyang Li, Chenkai Zhao, Wei Liu, Miao He, Shixing Tang, and Shuhua Xi. 2020. "The Impact of Temperature and Absolute Humidity on the Coronavirus Disease 2019 (COVID-19) Outbreak - Evidence from China." *MedRxiv*.

Tobías, Aurelio, Cristina Carnerero, Cristina Reche, Jordi Massagué, Marta Via, María Cruz Minguillón, Andrés Alastuey, and Xavier Querol. 2020. "Changes in Air Quality during the Lockdown in Barcelona (Spain) One Month into the SARS-CoV-2 Epidemic." *Science of The Total Environment* 726:138540.

Tosepu, Ramadhan, Joko Gunawan, Devi Savitri Effendy, La Ode Ali Imran Ahmad, Hariati Lestari, Hartati Bahar, and Pitrah Asfian. 2020. "Correlation between Weather and Covid-19 Pandemic in Jakarta, Indonesia." *Science of The Total Environment* 725:138436.

Turkish Statitical Institute. 2020. "Adrese Dayalı Nüfus Kayıt Sistemi." Retrieved (https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr).

Wang, Jingyuan, Ke Tang, Kai Feng, and Weifeng Lv. 2020. "High Temperature and High Humidity Reduce the Transmission of COVID-19." *SSRN Electronic Journal*.

World Population Review. 2020a. "Milan Population 2020." Retrieved (https://worldpopulationreview.com/world-cities/milan-population/).

World Population Review. 2020b. "Rome Population 2020." Retrieved (https://worldpopulationreview.com/world-cities/rome-population/).

Worldometer. 2020. "Coronavirus." Retrieved (https://www.worldometers.info/coronavirus/country/turkey/).

Journal of Statistics & Applied Science, Volume - 1, Issue - 2
İstatistik ve Uygulamalı Bilimler Dergisi, Cilt - 1, Sayı- 2

Pandemi Sürecinde Altın Fiyatları ile Kripto Para İlişkisinin Makine Öğrenme Metotları ile İncelenmesi

MEB, Dr. Ahmet SEL[1]

*Sorumlu Yazar, selahmet43@gmail.com

Özet

Yatırımcılar için borsa endekslerinin tahmini endeksin etkilendiği çok fazla değişken olmasından dolayı zor olduğu kadar ihtiyaç duyulan bir konudur. Doğru tahminler yatırımcıların elde ettikleri faydayı en üst düzeye taşır. Pandemi sürecinde dalgalanmalar tüm piyasa ve yatırım araçlarında fazlasıyla gözlenmiştir. Bu çalışmada, pandemi sürecinde seçili kripto para türlerinin altın fiyatları üzerindeki etkisi incelenmiştir. Uygulamalar Python programlama dili kullanılarak yapılmıştır. Bağımsız değişkenler Bitcoin, EOS, Tether, TRON ve Ripple olmak üzere altın fiyatları üzerinde ilişkisi eğitim ve test kümeleri üzerinden makine öğrenmesi gerçekleştirilerek incelenmiştir. Makine öğrenme modellerinden çok değişkenli, karar ağacı, destek vektör makineleri ve rasgele orman regresyon modelleri bir arada kullanılmıştır. Sonuç olarak; en yüksek açıklama oranına (R2 =0,91) sahip olan destek vektör makineleri yardımıyla altın fiyatları üzerinde Bitcoin +0,755, EOS -0,596, Tether -0,122 ve Tron + 0,220 etkisi olduğu belirlenmiştir. Elde edilen sonuçlar piyasada oynaklığın fazla olduğu dönemlerde yatırımların yönlendirilmesinde karar vericilere yardımcı olacağı düşünülmektedir.

Anahtar Kelimeler: Altın Fiyatları; Kripto Para; Pandemi; Yapay Zeka; Makine Öğrenmesi

Abstract

For investors, forecasting stock market indices is a problem that is as important as it is challenging since the index is affected by so many variables. The benefit that investors get is maximized by accurate forecasts. Fluctuations were observed in all market instruments and investment instruments during the pandemic process. The influence of selected cryptocurrency forms on gold prices during the pandemic process is investigated in this research. The Python programming language is used to render applications. Via training and test sets, the relation between the independent variables Bitcoin, EOS, Tether, TRON and Ripple on gold prices was investigated by machine learning. The multivariate decision tree, help vector machines and random forest regression models were used together for machine learning models As a result; with the help of support vector machines with the highest explanation rate (R2 = 0.91), it has been determined that there are Bitcoin +0.755, EOS -0.596, Tether -0.122 and Tron + 0.220 effects on gold prices. The results obtained will assist decision-makers in directing their investments in times of high market volatility.

Keywords: Gold Price; Crypto Money; Pandemic; Artificial intelligence; Machine Learning

[1] Bu çalışma 1. Uluslararası Uygulamalı İstatistik Kongresinde özet bildiri olarak sunulmuştur.

1. GİRİŞ

Kripto para, yeterli altyapı ve teknolojinin gelişmesiyle birlikte 2010'lu yıllarda uygulamaya girmeye başlamıştır. Matematiğin bir dalı olan şifreleme yani kriptoloji bilimi temeline dayanan bir para birimidir. Kripto para veya diğer adıyla sanal para, para gibi işlev gören ama paranın aksine ulusal sınırlar ve merkez bankalarından bağımsız bir değişim aracıdır (Maese, vd., 2016, s. 133). Kripto para birimleri olarak bilinen bu şifreli paralar; kriptografik para, sanal para, dijital para ya da elektronik para olarak da adlandırılmaktadır. Kripto paralar hiçbir hükümet, organizasyon veya banka tarafından yönetilmediği, daha hızlı ve güvenir olduğu için kullanıcılar tarafından maden veya madenci terimleri ile birlikte de kullanılmaktadır (Ateş, 2016, s. 349). Özellikle bazı kripto paralarının değerinin fazlasıyla artması insanları kripto para madenciliğine yöneltmiştir. Ancak sanılanın aksine kripto para üretimi için gereken ekipman ve teknolojiler zahmetli ve masraflı olabileceğini göstermektedir.

Blok zinciri teknolojisiyle geliştirilmiş ilk kripto para birimi Bitcoin (BTC)'dir. BTC, kullanıcılara ait tüm tasarrufların korunduğu, yapılan her işlemin kayıt altına alındığı, halka açık ve uçtan uca çalışan bir elektronik ödeme sistemidir (Nakamoto, 2019). Özellikle uluslararası ticarette güvenilir ve yenilikçi ödeme yöntemi olarak değerlendirilmiştir. Devletlerin bu ticari işlemlerin içeriğini ve transferleri kontrol edememesi eksi bir yön olarak kalmıştır. Ancak diğer para türleri gibi yatırım ve tasarrufların yönlendirilmesiyle tercih edilmeye başlanmıştır. Kripto paralarda giderek artan talep ve işlem hacmi yeni kripto paraların piyasaya girmesine neden olmuştur. Yeni çıkan kripto paralar için işlem hacimleri yüksek olmamasına rağmen spekülatif olaylara fazlasıyla açık oldukları ve kısa sürede yüksek oranda artış ve azalışlar gösterdiği bilinmektedir.

Yatırım araçlarından başında gelen altın, özellikle Türkiye'de "yastık altı" yatırım tabiriyle birikimi oldukça fazladır. Teknolojinin ilerlemesi ile daha çok fiziksel olarak biriktirilen altın yerini dijital hesaplara bırakmıştır. Yatırımcılar için portföy sepetinin gözde elamanı olan altının yanında risklerin azaltılması adına çeşitlendirme yoluna gitmektedirler. Bundan dolayı sepet oluşumunda yatırım araçlarının tamamı için fiyat tahminleri yatırımların yönlendirilmesi adına büyük önem taşımaktadır.

Aralık 2019'da ortaya çıkan koronavirüs (Covid-19) pandemisi piyasalarda olduğu gibi yatırım araçlarının çoğunda da olağan dışı dalgalanmalara sebep olmuştur. Özellikle piyasa ve yatırım araçlarında ani düşüş ve yükselişlerin görülmesi yatırımcılar için tam bir kafa karışıklığı yaratmıştır. Fiyatların tahmin teknikleri ve modelleri ile belirlenmesi gerçekçi karar alma adına yatırımcılar için son derece önemlidir. Tahminin doğruluğu başarılı kararlar alınmasını sağlar ve yatırımcıların fayda maksimizasyonuna ulaşmasına imkân tanır. Yatırım araçları arasında olağandışı durumların güvenli yatırım aracı genellikle altın olarak tabir edilmektedir.

Bu çalışmada, pandemi sürecinde altın fiyatlarının belirlenmesinde kripto paraların etkisi incelenmiştir. Altın fiyatlarının tahmini ve yatırım araçları ile ilişkisi konusunda literatürde oldukça fazla çalışma olmasına rağmen makine öğrenme modelleri ve kripto paraların etkilerinin araştırıldığı çalışmaya rastlanmamıştır. Ayrıca çalışma pandemi dönemindeki ilişkileri incelediğinden dolayı diğer çalışmalardan farklılaşmaktadır. Çalışmada analizler Python programlama dili kullanılarak makine öğrenme modellerinden çok değişkenli, karar ağacı, destek vektör makineleri ve rasgele orman regresyon modelleri bir arada kullanılmış ve en iyi sonuç aranmıştır. Elde edilen sonuçların pandemi dönemi gibi olağan dışı durumlarda yatırımların yönlendirilmesi ve portföy sepeti oluşturmada yardımcı olması beklenmektedir.

2. LİTERATÜR TARAMASI

Yapılan literatür araştırmaları kapsamında zaman serileri yardımıyla yapılan araştırmalar incelendiğinde; Elmas ve Polat (2014), çalışmalarında 1988-2013 yılları arasında altın fiyatını etkileyen; döviz kuru, Dow Jones Endeksi, faiz oranı, enflasyon oranı, gümüş fiyatı ve petrol fiyatı faktörleri üzerine zaman serileri ile incelenmiştir. Altın fiyatlarını; petrol fiyatları, gümüş fiyatları ve enflasyon oranının pozitif yönde, döviz kuru, Dow Jones Endeksi ve faiz oranının ise negatif yönde etkilediği tespit edilmiştir. Elde edilen sonuca göre faiz oranı dışındaki diğer faktörlerin altın fiyatına etkisi anlamlı bulunmuştur. Gültekin ve Hayat (2016), çalışmalarında altın fiyatlarının Türkiye'de etkilendiği faktörleri zaman serileri için kullanılan Vektör Otoregresif Modeller (VAR) modeli ile araştırmıştır. 2005:01-2015:04 için aylık veriler ile İstanbul Altın Borsası'nda (İAB) altın fiyatları, faiz oranı, TÜFE, döviz kuru ve BİST 100 endeksi, altının ons fiyatı ve petrol fiyatı ile incelenmiştir. Sonuç olarak, İAB altın fiyatı tahmininde en fazla oran ons ve petrol fiyatı olurken en az faiz oranı olarak belirlenmiştir. Güleç (2018), çalışmasında önde gelen kripto paraları incelemiştir. Çalışmada Bitcoin'in döviz, emtia piyasaları, hisse senedi ve faiz ile olan ilişkisi incelenmiştir. Veriler aylık olarak 2012-2018 dönemini ele almaktadır. İnceleme yöntemi olarak Johansen Eşbütünleşme ve Granger Nedensellik analizleri uygulanmıştır. Sonuç olarak; Bitcoin fiyatları için artan bir trende ve yüksek bir volatiliteye sahip olduğu belirlenmiştir. Karasu ve diğ. (2018), çalışmalarında altın fiyatlarına etki eden faktörleri incelemişlerdir. 2003-2014 dönemi aylık veriler kullanılarak altın fiyatlarını etkileyen petrol fiyatları, mevduat faiz oranları, enflasyon oranları, gümüş fiyatları ve reel döviz kurları değişkenler olarak seçilerek eşbütünleşme analizi uygulanmıştır. Analiz sonuçlarında Türkiye'de enflasyon oranı ve gümüş fiyatlarının altın fiyatları üzerinde etkisi olduğu belirlenmiştir. Yıldırım (2018), çalışmasında Bitcoin ile altına arasındaki ilişkiyi ADF Birim Kök Testleri, Hata Düzeltme Modeli, Johansen Koentegrasyon Testi ve Düzeltilmiş En Küçük Kareler Modelleri ile değerlendirmiştir. Analiz sonuçlarında Bitcoin ve altın arasında kısa dönem için ilişkisiz olduğu görülmüştür. Ayrıca Bitcoin fiyatlarının altını etkilemezken, altın fiyatlarının Bitcoin fiyatlarını uzun dönemde etkilediği görülmüştür. Cingöz ve Kendirli (2019), çalışmalarında altın için fiyat hareketinde BİST 100 ve dolar kurundaki değişimlerin ilişkisi araştırılmıştır. Altın fiyatları üzerine BİST 100 ve dolar kurunun uzun vadede anlamlı bir etkisi olabileceği ancak kısa vadede değişkenlerin altın fiyatı üzerinde anlamlı bir etkisinin olmadığı sonucuna ulaşılmıştır. Klein ve diğ. (2018), çalışmalarında Bitcoin ile altın arasındaki ilişki incelenmiştir. Çalışma da koşullu varyans analizi ile Bitcoin ve altının özelliklerini analiz edip karşılaştırılmış ve yapılarındaki farklılıkları bulunmuştur. Sonuç olarak Bitcoin' in ile altının ters ilişkili olduğu ve hisse senedi piyasalarında temelde farklı özelliklere sahip olduğu görülmüştür.

Yapay sinir ağları ile yapılan çalışmalar incelendiğinde ise; Kocatepe ve Yıldız (2016), çalışmasında yapay sinir ağları yardımıyla altın fiyatında meydana gelen değişimler için yön tahmini yapılmıştır. Uygulama da 2007- 2015 dönemi aylık veriler ile Türkiye altın gram fiyatı bağımlı değişken iken bağımsız değişkenler ham petrol fiyatı, dolar kuru, dolar endeksi, BIST100 endeksi, Standard&Poor's 500 endeksi, Türkiye enflasyon, faiz ve tahvil oranları, ABD enflasyon, gümüş ve bakır fiyatları ele alınmıştır. Sonuç olarak bağımsız değişkenlerin altın fiyatındaki değişim yönünü %75,24 oranına başarılı olarak tahmin ettiği belirlenmiştir.

Portföy sepeti oluşturma konusunda yapılan çalışmalara bakıldığında; Okuyan ve Deniz (2019), çalışmalarında kripto paraların portföy sepetine olan etkileri değerlendirilmiştir. Kripto para türlerinden en gözde olan Bitcoin ve Etheryumun başlıca ülkelerin ve Türkiye'nin borsa endeksleri ile altın, gümüş ve platin fiyatları kullanılmıştır. Elde edilen sonuçlarda kripto para getirileri ile kıymetli maden ve hisse senedi endeksi getirileri arasında pozitif ve anlamlı bir ilişki bulunamamıştır. Bu sonuç kripto paraların kıymetli maden ve hisse senedi portföylerinde iyi bir seçim olacağını göstermiştir. Gül (2020), çalışmasında kripto paralar ile portföy çeşitlendirmesini araştırmıştır. Bu amaçla, 2015-2020 dönemi günlük verileri yardımıyla hisse senetleri, döviz kurları, emtialar ve yatırım fonları ile çeşitli portföyler oluşturulmuştur. Daha sonra kripto paralar (Bitcoin, Ethereum, Ripple) portföylere ilave edilerek performansları değerlendirilmiştir. Analiz sonuçlarında, kripto paralar ile diğer varlıkların arasındaki korelasyonların genellikle ters yönlü yani negatif olduğu görülmüştür. Ayrıca, portföylere eklenen kripto paraların çoğunlukla düşük riskler ile daha yüksek getiriler elde edildiği görülmüştür. Bundan dolayı portföy çeşitlendirmesinde kripto paralar iyi bir araç olabilecekleri sonucu çıkmıştır.

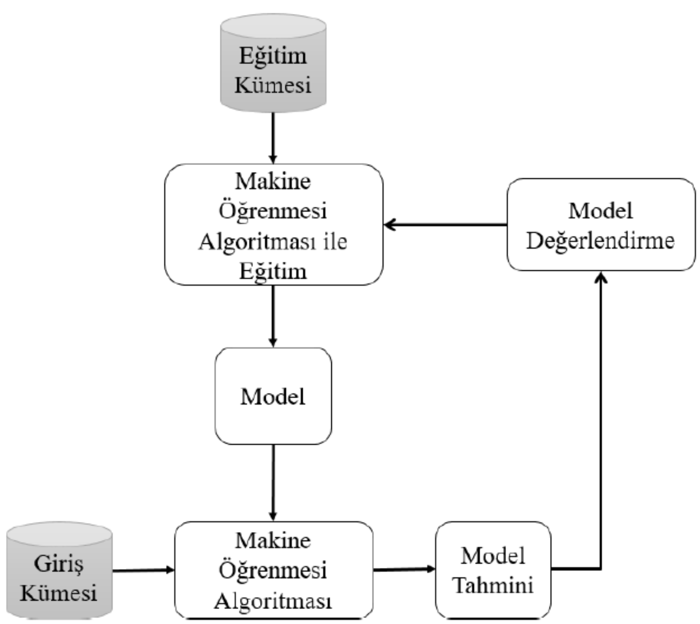
Pandemi döneminde altın fiyatları üzerine yapılan çalışmalar incelendiğinde; Gülhan (2020), çalışmasında Kovid-19 pandemisinin altın fiyatları üzerindeki etkisini ölçmek amacıyla 22.01.2020-08.04.2020 tarihleri arası günlük verilerle ARDL modeli kullanmıştır. Bağımlı değişken olarak altın fiyatlarının ele alındığı çalışmada, Kovid-19 Türkiye vaka sayısı, Kovid-19 Dünya vaka sayısı, US dolar kuru, politika faizi ve akaryakıt fiyatları bağımsız değişkenler olarak analize dâhil edilmiştir. Analiz sonuçlarına göre, kısa dönemde altın fiyatları ile dolar kuru ve politika faizi arasında negatif, akaryakıt fiyatları, Kovid-19 Dünya vaka sayısı, Kovid-19 Türkiye vaka sayısı gecikme değerleri, altın fiyatı gecikme değerleri arasında ise pozitif yönlü ilişki tespit edilmiştir.

3. YÖNTEM

Analiz sonuçlarının anlamlı çıkması adına veriler öncelikle normalleştirme işlemine tabi tutulmuştur. X' normalleştirilmiş değer olmak üzere her bir veri olan X için μ: ortalama ve σ: standart sapma değeri ile Denklem (1) uygulanarak normalleştirme işlemine tabi tutulmuştur.

$$X' = \frac{X-\mu}{\sigma} \quad (1)$$

Elde edilen normalleştirilmiş veriler üzerinden eğitim (%66) ve test (%33) kümeleri oluşturularak makine öğrenmesi gerçekleştirilmiştir. Makine öğrenme sürecinin temel işleyiş yapısı Şekil 1.'de gösterilmiştir.



Şekil 1. Temel makine öğrenmesi süreci

Eğitim kümesi ile makine öğrenmesi algoritması eğitildikten sonra giriş kümesinde yer alan test değerleri yardımıyla elde edilen model üzerinden çıktılar alınarak model değerlendirme işlemine tabi tutulur. Değerlendirme sonucunda elde edilen model yeterli performansı göstermesi durumunda model kabul edilmiş olacaktır. Çalışmada makine öğrenme modellerinde ise çok değişkenli, karar ağacı, destek vektör makineleri ve rasgele orman regresyon modellerine ayrı ayrı uygulanarak en iyi performansı veren değer araştırılmıştır. Analiz işlemleri sırasında performans ölçütleri olarak açıklayıcılık (belirtme) katsayısı (R^2), kök ortalama kare hatası (Root Mean Square Error) (RMSE) ve ortalama mutlak hata (Mean absolute error) (MAE) değerleri hesaplanarak ölçülmüştür. Burada;

$$y_i = Gözlem\ değerleri$$

$$\bar{y}_i = Gözlem\ değerlerinin\ ortalaması$$

$$\hat{y}_i = Tahmin\ değerleri$$

$$n = Gözlem\ sayısı$$

olmak üzere;

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n}} \tag{3}$$

$$MAE = \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})}{n} \tag{4}$$

şeklinde hesaplanmaktadır. Burada R^2 için 1'e en yakın değeri alması beklenirken RMSE ve MAE için en düşük değeri veren tahmin başarılı sayılacaktır.

3.1. Çoklu Regresyon Analizi

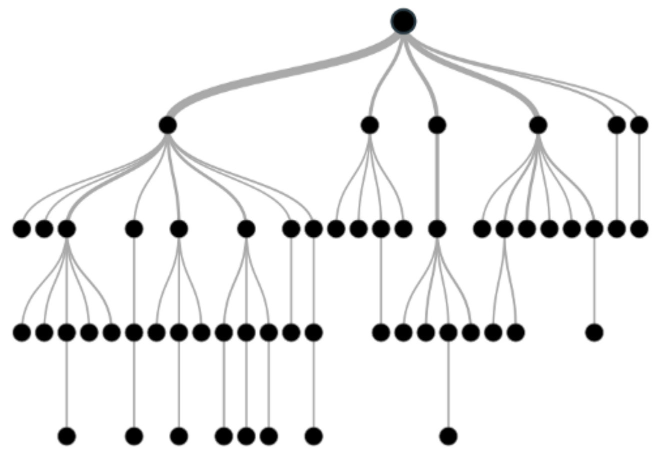
Regresyon analizi, bağımsız değişken/değişkenler ile bağımlı değişken/değişkenler arasındaki ilişkiyi incelemek için kullanılan bir yöntemdir. Bağımsız ve bağımlı değişkenlerin modelde birer tane olması ile tek değişkenli veya basit regresyon modeli, bağımlı değişkenin tek iken birden fazla bağımsız değişkenin olduğu regresyon modeline de çoklu regresyon modeli olarak adlandırılmaktadır. Birden çok bağımlı ve bağımsız değişkeni bulunan modeller ise çok değişkenli regresyon modeli olarak bilinmektedir. Bir tane bağımlı değişkenin için bağımsız değişkenin sayısının ise iki ve daha fazla olduğu regresyon modeline çoklu regresyon modeli denir (Özdamar, 2004:189).

Bağımlı değişkenin bir tane ve bağımlı değişken üzerindeki etkileri aranan bağımsız değişkenin birden fazla olduğu regresyon modeline çoklu regresyon modeli denir. Çoklu regresyon modeli için Y: bağımlı değişken olmak üzere i=1..n tanımlanan X_i'ler bağımsız değişkenler için β_0: sabit iken β_i'ler katsayılardır. n değişkenli model için Denklem (5) gibi yazılabilir.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{5}$$

3.2. Karar Ağacı Modelleri

Veri madenciliği çalışmalarında kullanılan karar ağacı yöntemi, tahmin ve sınıflandırmada başvurulan önemli yöntemlerden biridir. Karar ağacı, girdisi bulunmayan bir kök düğüm ve her biri ayrı birer girdi olarak değerlendirilen iç düğümlerden meydana gelen yönlü bir ağaç şeklindedir. Düğümler karar ağacı modeli için çıktıları bir başka düğüm için girdi ise test ya da iç düğümü, girdi değilse de yaprak düğümler şeklinde tanımlanır (Maimon ve Rokach, 2010, s. 150). Karar ağacı modellerinin görsel yapısı Şekil 2.'de gösterilmiştir.



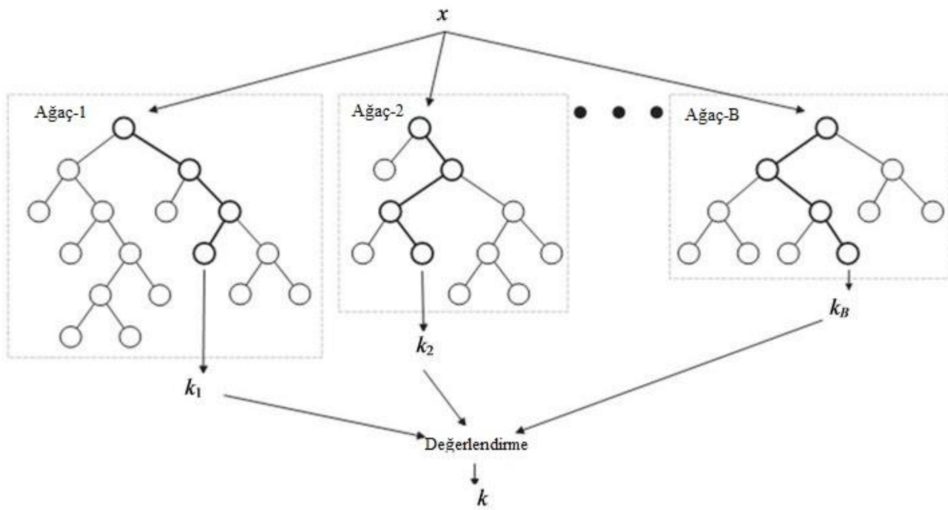
Şekil 2. Karar Ağacı Modeli

Karar ağacı modelleri gözetimli öğrenme modeli olmakla birlikte sınıflandırma ve regresyon modellerinde kullanılmaktadır. Karar ağacının her bir dalı programlama da dilinde yer alan "eğer" yapısına uygun olarak ayrılmaktadır. Meydana gelen dallar yardımıyla analize konu olan veri tüm modele işlenmiş olur. Karar ağacı modeli eğitim verisi sonucunda eğitilen modelin test verisi kullanılarak istatiksel olarak etkinliğinin belirlenmesinden sonra tahmin yöntemleri için kullanılmaktadır. Modelin yapısı özellikle sınıflandırma sonuçlarının görsel olarak gösterimini kolaylaştırmaktadır. Diğer yöntemlere nazaran karar ağıcı modellerinin hızlı sonuç vermesi ve veriler arası ilişkilerin yapısal olarak gösterilmesi tercih edilmesinde önemli bir rol oynamaktadır.

Karar ağacı algoritmasının temel amacı karmaşık yapılardan ziyade küçük boyutlu ve az derinlikli ağaçların oluşturulmasıdır. Karar ağacı algoritmaları kullanılarak elde edilen karmaşık ve büyük yapılar genellikle düşük başarı düzeyine sahiptir. Bu sebeple, küçük boyutlu karar ağaçları modelleri oluşturmak amacıyla birçok farklı yaklaşım önerilmiştir. Bu yaklaşımlardan biri düğümlerin oluşturulması ve ayrılmasında ölçütlerin kullanılmasıdır. GINI indeksi, bilgi kazancı, ki-kare istatistiği gibi ölçütler, başlıca kullanılan düğüm ayırma ölçütleri arasındadır (Kothari ve Dong, 2001, s. 172). Karar ağacının oluşturulması ve analiz kısmında farklı ölçütlerin kullanılması modelin başarı düzeyini de etkilemektedir. Bundan dolayı tek bir ölçüt ile modelin başarısının sınanması yerine farklı ölçütler kullanılarak başarı düzeyenin karşılaştırılması önemlidir.

3.3. Rasgele Orman Modelleri

Rasgele orman algoritması topluluk öğrenme yöntemlerinden biridir. Topluluk öğrenme yöntemleri, farklı modelleri birleştirerek sonuçları iyileştirmeyi amaçlar. Rasgele orman algoritması birden fazla karar ağaçlarının topluluğundan elde edilmektedir. Rasgele orman algoritması avantajlarından biri hem sürekli hem kesikli değişkenlerin birlikte kullanabilmektedir. Ayrıca büyük küçük boyutlu veri setlerinde kullanılabilmektedir. Rasgele orman ise doğrululuk oranı diğer algoritmaya göre yüksek çıkmaktadır. Dezavantaj olarak ise algoritma siyah kutu yani ağaç yapısı görülmemektedir (Breiman, 2001, s. 20). Rasgele orman yönteminin oluşturduğu yapı Karar ağaçları (KA) yapısıyla meydana gelmektedir. Şekil 3.'te rasgele orman modelinin yapısı görülebilir.



Şekil 3. Rasgele Orman Modeli Yapısı

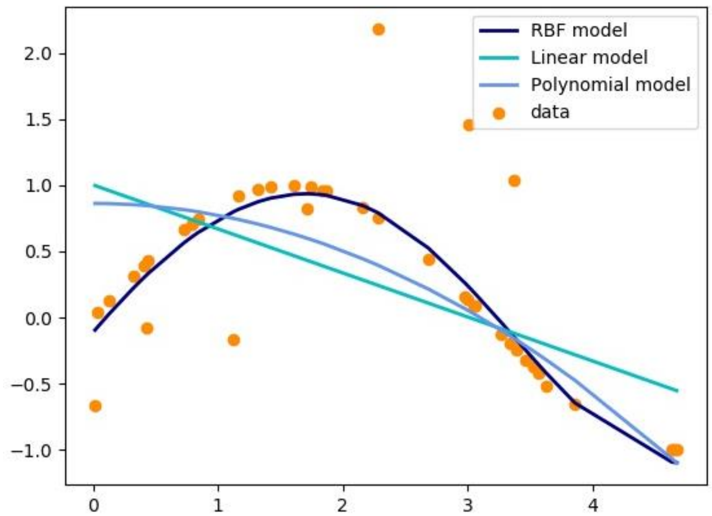
Karar ağaçlarının her birinden elde edilen sonuçların ortalamaları tahmin değerini vermektedir. Ancak karar ağaçlarında olduğu gibi elde edilen modelin görselleştirilmesi çoğu zaman sorun olabilmektedir. Ağaç yapısını oluşturan dalların oluşumunda en önemli adım meydana gelen dallanmaların hangi kriter ya da öznitelik değerinin dikkate alınacağının belirlenmesidir. Kaynaklar incelendiğinde bu problemin çözümünde kullanılan çeşitli yaklaşımlar olduğu görülmektedir. Bunlardan en önemli olarak kabul edilenler ise Gini indeksi, Twoing kuralı, bilgi kazancı ve bilgi kazanç oranı ile Ki–Kare olasılık tablo istatistiği yaklaşımlarıdır. ID3 algoritması tek değişkenli karar ağaçlarından olmak üzere bilgi kazancı yaklaşımını kullanırken, C4.5 algoritması bilgi kazancı ve bölünme bilgisi kavramından yararlanmaktadır. Regresyon ağacı ve sınıflandırma yöntemi olarak bilinen CART algoritması ise Twoing kuralını kullanmaktadır (Breiman, Friedman, Olshen ve Stone, 1984:28).

Bagging yöntemi, birçok sınıflama modeline uygulanabilmekle birlikte daha çok karar ağaçları için kullanılmaktadır. Bagging yöntemi veri setinden sınıf yapısını bozmayacak şekilde rastgele örnekler seçilerek (bootstrap) oluşturulan çok sayıdaki karar ağacının yaptığı sınıf tahminleri oylanarak en çok oy toplayan sınıfı asıl sınıf tahmini olacak şekilde belirleyen öğrenme yöntemidir. Bagging yönteminde art arda oluşturulan ağaçlar önceden oluşturulan ağaçlara bağımlı değildirler ve ağaçlar orijinal veri setinden bootstrap örnekleme yapılarak oluşturulmaktadır.

Rasgele Orman (Random Forests) yönteminde sonradan gelen veriye ait tahmin yapılmasının yanında, değişkenlerin önem derecesi de hesaplanmaktadır. Veri setinde çok sayıda değişken varsa, değişken önem derecesinin hesaplanması model indirgemesi açısından oldukça kullanışlıdır. Örneğin binlerce değişkenin bulunduğu veri setinde, Rasgele Orman yöntemiyle elde edilen önem derecesine göre, kurulacak yeni modelde önem derecesi yüksek değişkenler kullanılarak daha etkin tahminlerin yapılması sağlanabilir.

3.4. Destek Vektör Makineleri

1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından temelleri atılan Destek Vektör Makineleri (Support Vector Machine) (SVM) istatiksel öğrenme teorisine dayalı bir gözetimli öğrenme algoritmasıdır. Her ne kadar temelleri 60'lı yıllara dayansada 1995 yılında Vladir Vapnik, Berhard Boser ve Isabelle Guyon tarafından geliştirilmiştir (Akpınar, 2017). SVM modelleri aşırı öğrenmeyi azaltan yapısı ile sınıflandırma ve regresyon problemlerinin çözümünde uygulanan ve daha iyi sonuçlar veren güdümlü bir öğrenme algoritmasıdır (Panigrahi ve Mantri, 2015, s. 763). SVM algoritmasında, regresyon için Destek Vektör Bağlayıcısı (Support Vector Regressor-SVR) adı verilen bir yapı bulunmaktadır. SVR, regresyon hatalarını en aza indirmek için deneysel riskleri ölçer ve bunun için de bir maliyet fonksiyonu kullanır (Yu, Chen ve Chang 2016, s. 705). SVM algoritması, doğrusal olan ve doğrusal olmayan türlere sahiptir. Hiper düzlemlerdeki en optimal noktayı bulabilmek için kernel adı verilen çekirdek yapılar kullanılır (Lin ve Wang, 2002, s. 465). Şekil 4.'te SVR modelleri için Radial Basis Function (RBF) metodu kernel yapısına göre doğrusal ve polinomal eğrilere göre daha başarılı sonuç verdiği görülmektedir.



Şekil 4. SVR (rbf)-Doğrusal-Polinomal Tahminler

SVM yönteminde doğrusal sınıflanabilen veriler için birbirinden ayırt edilmesi amacıyla doğrusal fonksiyonlar arasından, en büyük marjini olan fonksiyon seçilmektedir. Eğer sınıflamada örneklerin doğrusal bir düzlem ile ayrıştırılabilecek düzeyde değilse, yöntemde kullanılan Kernel fonksiyonu yardımıyla daha yüksek boyutlu bir uzaya aktarılması mümkün olmaktadır. Bu şekilde marjini en yüksek olan hiper düzlemler bulunur. Sonuç olarak veriler bu ayırt edici hiper düzleme göre sınıflara atanır (Coşgun ve Karaağaoğlu, 2011, s. 185). Python programlama dili altında bulunan SVR modeli kernel yapıları {'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'} seçenek olarak verilmektedir. Verinin yapısı ve model başarısının yüksekliğine göre bu yapılardan uygun olanı modele uygulanır.

4. UYGULAMA

Uygulamada Python programlama dilinde numpy, matplotlib, pandas, sklearn.metrics ve statsmodels.api kütüphaneleri kullanılmıştır. Bu kütüphanler yardımıyla makine öğrenme modellerinden çok değişkenli, karar ağacı, destek vektör makineleri ve rasgele orman regresyon modelleri bir arada kullanılmış ve en iyi sonuç aranmıştır.

4.1. Veri ve Değişkenler

Analize konu olan değişkenlerin değerleri dolar cinsinden ele alınarak açık erişim adresi olan (investing.com) sitesinden elde edilmiştir. İnceleme sırasında piyasa da aktif olarak işlem gören 3855 kripto para olduğu görülmüştür. Araştırmanın pandemi sürecini kapsamasından dolayı 1/12/2019 ile 8/09/2020 tarihleri arasında veriler incelenmiştir. Analize veriler değerlendirmeye alınırken tarih aralığında ortak olarak işlem gördükleri günler dikkate alınmıştır. Uygulamada kripto paralar arasından en yaygın olarak kullanılan ve piyasa hacmi en az %0,5 yoğunluğa sahip olanlar kullanılmıştır. Bağımsız değişken olarak ele alınan 13 tane kripto para ve değerleri Tablo 1.'de gösterilmiştir.

Tablo 1. Piyasa Hacmi %0,5'ten Büyük Kripto Paralar

Kripto Para	Sembol	Fiyat(USD)	Piyasa Hacmi
Tether	USDT	1,00	38,39%
Bitcoin	BTC	10591,9	24,23%
Ethereum	ETH	340,45	14,12%
Chainlink	LINK	9,54	2,16%
EOS	EOS	2,48	1,66%
TRON	TRX	0,03	1,45%
Ripple	XRP	0,23	1,30%
Litecoin	LTC	44,20	1,12%
Bitcoin Cash	BCH	211,49	0,96%
Power Ledger	POWR	0,08	0,93%
Neo	NEO	21,78	0,79%
Bitcoin SV	BSV	152,63	0,56%
Ethereum Classic	ETC	4,91	0,50%

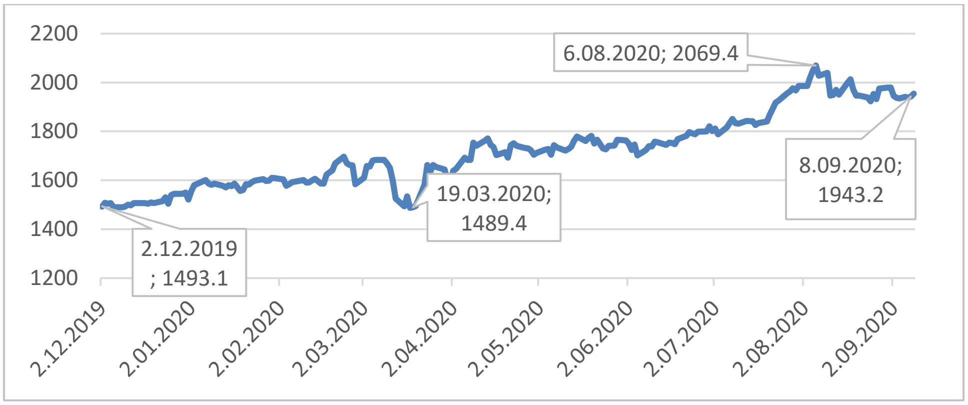
Kaynak: Kripto Para, link: https://tr.investing.com/crypto/

Tablo 1. İncelendiğinde piyasa değerleri incelendiğinde BTC, ETH ve BCH'nin ilk sıralarda yer aldığı görülmektedir. Ancak BTC'nin piyasa değeri USDT'den çok fazla olmasına rağmen piyasa yoğunluğunda ilk sırada yer almaktadır. Bağımsız değişkenler arası korelasyon matrisi Tablo 2.'de gösterilmiştir.

Tablo 2. Bağımsız Değişkenler Arası Korelasyon Matrisi

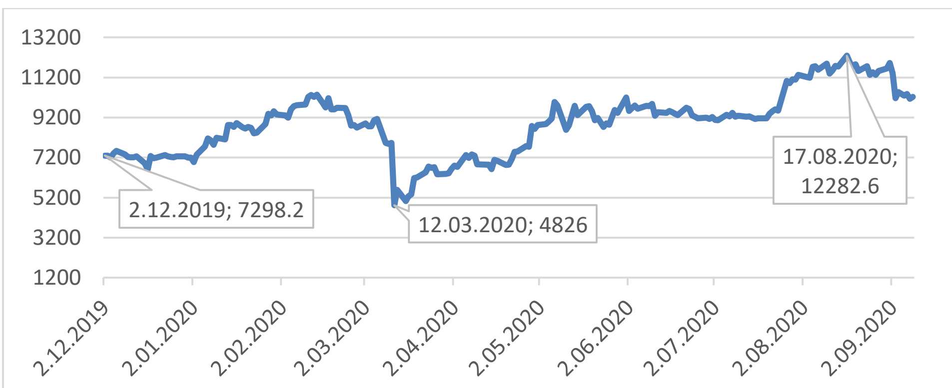
	BCH	BTC	BSV	LINK	EOS	ETC	ETH	LTC	NEO	POWR	USDT	TRX	XRP
BCH	1,00												
BTC	0,43	1,00											
BSV	0,90	0,43	1,00										
LINK	0,06	0,78	0,08	1,00									
EOS	0,97	0,43	0,83	0,09	1,00								
ETC	0,94	0,45	0,90	0,02	0,88	1,00							
ETH	0,24	0,90	0,25	0,94	0,25	0,21	1,00						
LTC	0,94	0,58	0,80	0,27	0,96	0,84	0,44	1,00					
NEO	0,35	0,82	0,29	0,78	0,41	0,29	0,85	0,53	1,00				
POWR	0,11	0,75	0,21	0,70	0,06	0,17	0,79	0,24	0,57	1,00			
USDT	-0,40	-0,61	-0,53	-0,28	-0,33	-0,49	-0,41	-0,35	-0,39	-0,42	1,00		
TRX	0,34	0,77	0,28	0,79	0,41	0,28	0,84	0,52	0,93	0,52	-0,36	1,00	
XRP	0,72	0,76	0,57	0,59	0,77	0,58	0,71	0,87	0,72	0,36	-0,35	0,71	1,00

Bağımsız değişken olarak alınan 13 kripto para içerisinden regresyon analizi sonuçlarında çoklu bağlantı sorunu olmaması adına korelasyon değerleri hesaplanarak mutlak değerce 0,80'den büyük olan değişkenler analizden çıkarılmıştır. Elde edilen sonuçlarda ise bağımsız değişkenler Bitcoin, EOS, Tether, TRON ve Ripple olmak üzere altın fiyatları bağımlı değişken olarak belirlenmiştir. Bağımlı değişken altın için pandemi sürecindeki fiyat değişimi Grafik 1.'de gösterilmiştir.



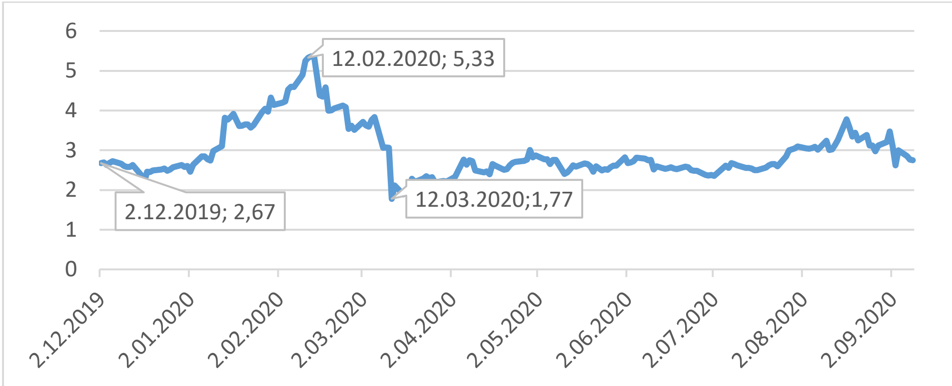
Grafik 1. Pandemi Süreci Altın Fiyatları ($), (Kaynak:investing.com)

Grafik 1. incelendiğinde altın fiyatlarında 2 Aralık 2019 tarihinde başlayan artış trendi 19 Mart 2020 tarihinde dip seviyeye geldikten sonra artışlar devam etmiştir. Ağustos 2020 başında ise en yüksek değerine ulaşmıştır. Bitcoin fiyatlarının pandemi süreci içerisindeki değişimi Grafik 2.'de gösterilmiştir.



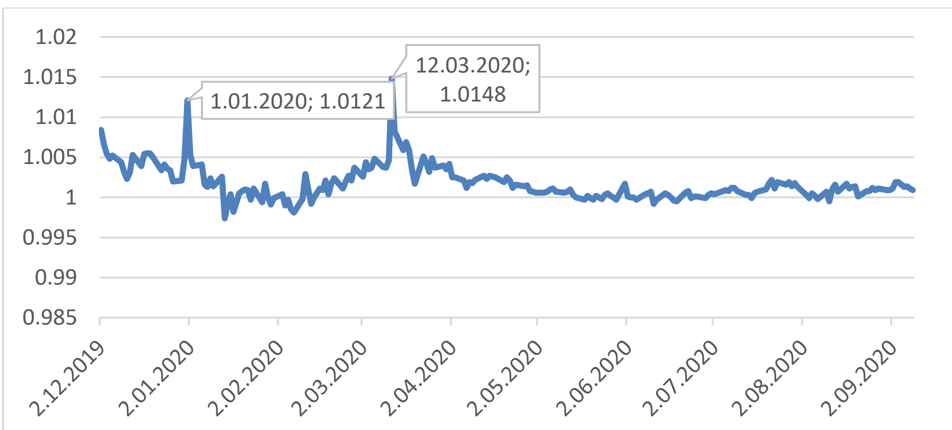
Grafik 2. Pandemi Süreci BTC Fiyatları ($), (Kaynak:investing.com)

Bitcoin fiyatlarının pandemi sürecinde değişimi izlendiğinde 2 Aralık 2019 tarihinde itibaren artış trendi Şubat ortalarında biterek 12 Mart 2020 tarihinde dip noktaya ulaşmıştır. Ancak 17 Ağustos 2020 tarihinde yaklaşık 2,5 katı değer kazandığı görülmektedir. EOS için pandemi süreci fiyat değişimi Grafik 3'te gösterilmiştir.



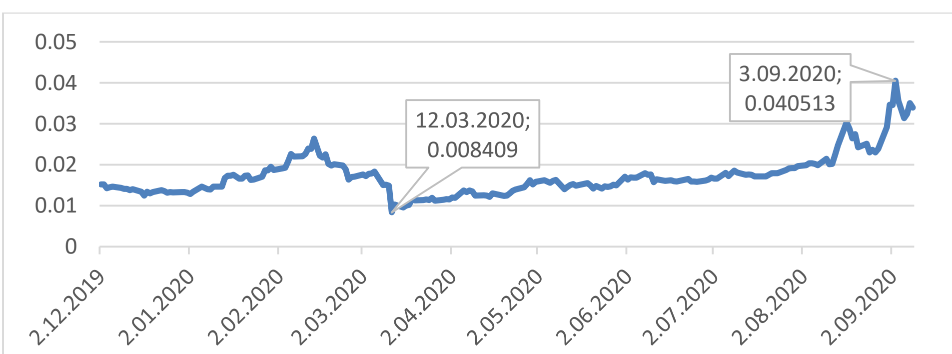
Grafik 3. Pandemi Süreci EOS Fiyatları ($), (Kaynak:investing.com)

EOS için Grafik 3. İncelendiğinde 12 Şubat 2020 tarihinde pandemi başındaki değerinin 2 katına çıktığı görülmektedir. Ancak tam bir ay sonra 3'te 1 fiyatına inan EOS için fiyatının genel anlamda 2-3 bandında gezindiği söylenebilir. Tether için fiyat değişimleri Grafik 4'te gösterilmiştir.



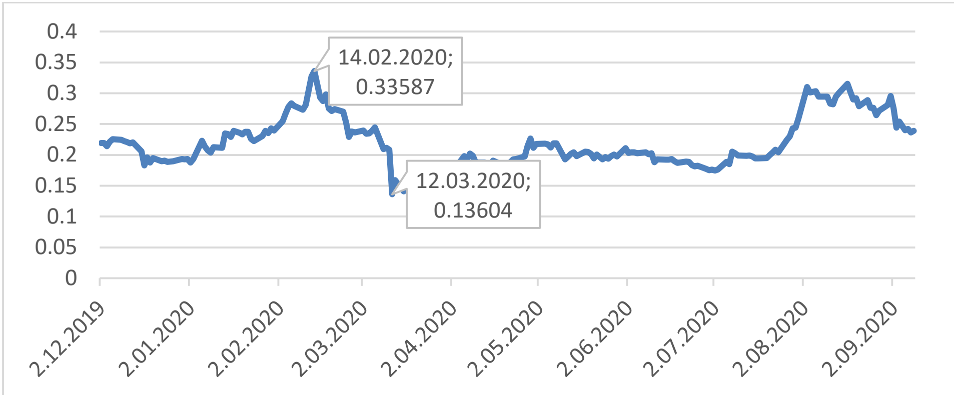
Grafik 4. Pandemi Süreci USDT Fiyatları ($), (Kaynak:investing.com)

Tether fiyat değişimi için Grafik 4. incelendiğinde dikkat çeken nokta 12 Mart 2020 tarihinde en yüksek değerine ulaşmasıdır. Bunun dışında 1$ bandında fiyatının değişkenlik gösterdiği söylenebilir.



Grafik 5. Pandemi Süreci TRX Fiyatları ($), (Kaynak:investing.com)

TRON fiyat değişiminin 12 Mart 2020 tarihinde dip yaptığı ve 3 Eylül 2020 tarihinde ise tavan yaparak yaklaşık 5 katına çıktığı görülmüştür.



Grafik 6. Pandemi Süreci XRP Fiyatları ($), (Kaynak:investing.com)

Ripple fiyatlarının ise 14 Şubat 2020 tarihinde tavan yaptığı görülmüştür. Genel olarak değerlendirdiğimizde bağımsız değişken olarak alınan kripto paralar için 12 Mart 2020 tarihinde Tether tavan yapmasına rağmen diğerleri dip fiyat noktalarına düşmüşlerdir.

4.2. Analiz Sonuçları

Normalleştirme işlemine tabi tutulmuş veriler üzerinden regresyon modelleri makine öğrenmesi sonucunda elde edilen sonuçlar ile kernel yapısı Radial basis function (rbf) olarak seçilen Destek Vektör Makineleri yönteminin daha başarılı olduğu görülmüştür.

Tablo 3. Makine Öğrenmesi Sonucu Modellerin Regresyon Başarıları

Regresyon Yöntemi	**Rkare**	**MAE**	**RMSE**
Çok Değişkenli	0,80	0,38	0,45
Karar Ağacı	0,80	0,28	0,52
Destek Vektör Makineleri	0,91	0,25	0,35
Rasgele Orman	0,83	0,28	0,37

Destek Vektör Makineleri regresyon modeli sonucunda XRP anlamsız olduğu için modelden çıkartılması sonucu elde edilen model aşağıdaki gibidir.

Tablo 4. SVM Regresyon Modeli Katsayı Sonuçları

Bağımsız	**Katsayı**	**Standart Hata**	**p(olasılık)**
BTC	0,755	0,047	0,000
EOS	-0,596	0,029	0,000
USDT	-0,122	0,031	0,000
TRX	0,220	0,043	0,000

Analiz sonucunda elde edilen regresyon modeli aşağıdaki gibidir.

$$Altın = 0{,}755 * BTC - 0{,}596 * EOS - 0{,}122 * USDT + 0{,}220 * TRX$$

Destek Vektör Makineleri regresyon modeli sonucunun Altın fiyatları üzerinde R^2=0,91 olmak üzere Bitcoin pozitif yönlü 0.755, EOS negatif yönlü 0.596, Tether için negatif yönlü 0.122 ve Tron için pozitif yönlü 0,220 etkinin olduğu belirlenmiştir.

Yoğunluk değerleri dikkate alındığında Tether ilk sırada yer alarak negatif etkisi gözlenmiştir. Bitcoin ise ikinci sırada yer alarak pozitif yönde daha güçlü bir etkiye sahiptir. Yoğunluk sıralamasında EOS ve Tron yakın değerler almasına rağmen EOS negatif yönde daha güçlü bir etkiye sahip olduğu görülmüştür.

5. SONUÇ ve ÖNERİLER

Tasarruf ile elde edilen paranın yatırım araçları yardımıyla biriktirilmesi gelir elde etmekten öte günümüzde değer kaybının önlenmesi amacıyla da gerçekleştirilmektedir. Bu yüzden yatırımların yönlendirilmesinde tahminlerin doğru ve yerinde olması yatırımcı için hayati önem arz etmektedir. Ancak pandemi sürecinde piyasalarda meydana gelen büyük dalgalanmalar panik ortamını arttırmıştır. Doğru kararlar alabilen yatırımcılar için büyük kar elde etme fırsatı doğmuş olmasına rağmen bir o kadar da zarara sebep olduğu görülmüştür. Kar ve zarar dengesini gözetebilmek adına portföy sepeti uygulamasında bu çalışma için altın ve kripto paralar incelemeye alınmıştır. Çalışmada pandemi sürecinde seçili kripto para türlerinin altın fiyatları üzerindeki etkisi incelenmiştir.

Bağımsız değişkenler Bitcoin, EOS, Tether, TRON ve Ripple olmak üzere altın fiyatları üzerinde ilişkisi eğitim (%66) ve test (%33) kümeleri üzerinden makine öğrenmesi gerçekleştirilerek incelenmiştir. Bitcoin fiyatlarının tahmin sonuçlarının regresyon performansları R^2, RMSE ve MAE değerleri hesaplanarak ölçülmüştür. Makine öğrenme modellerinden çok değişkenli, karar ağacı, destek vektör makineleri ve rasgele orman regresyon modelleri bir arada kullanılmıştır. Sonuç olarak; destek vektör makineleri için kernel yapısı "rbf" için açıklayıcılık oranı R^2 =0,91 olmak üzere altın fiyatları üzerinde Bitcoin +0,755, EOS -0,596, Tether -0,122 ve Tron + 0,220 şeklinde etkisi olduğu belirlenmiştir.

Klein ve diğ. (2018), çalışmasının sonucu olan Bitcoin ile altın arasındaki ters ilişki ve Yıldırım (2018), çalışmasında Bitcoin fiyat hareketlerinin altın fiyatlarını etkilemezken, altın fiyat hareketlerinin Bitcoin fiyatlarını uzun vadede etkilediğini belirlemiştir. Ancak pandemi sürecinde bu ilişkinin pozitif yönlü olarak gerçekleştiği görülmüştür. Ayrıca portföy çeşitlendirmelerinde Okuyan ve Deniz (2019) ile Gül (2020), çalışmalarında kripto paraların hisse senedi ve kıymetli maden portföyleri için iyi bir çeşitlendirme varlığı olduğunu göstermiştir. Dolayısıyla, kripto paraların portföy çeşitlendirmesi için iyi bir araç olabilecekleri ve portföy performanslarını olumlu etkiledikleri sonucuna varılmıştır.

Kripto paraların fazlasıyla etkin olmasına rağmen fiyat oynaklıklarının yüksek olması yatırım konusunda sınırlandırmaktadır. Ancak yapılan yatırımların piyasa içinde büyüklüğü ve yüksek getirilerinden dolayı tercih edilmeleri giderek artmaktadır. SVM modelinin kripto para türleri ile altın fiyatlarını tahmin etmede diğer makine öğrenme modellerine göre daha tutarlı sonuçlar verdiği görülmüştür. Elde edilen sonuçlar piyasada oynaklığın fazla olduğu dönemlerde yatırımların yönlendirilmesinde karar vericilere yardımcı olacaktır. Bundan sonraki çalışmalarda pandemi süreci sonrası değerlendirilebilir ve dönemsel olarak yatırım araçları da çeşitlendirilerek kripto paralar ile emtia ürünleri arası ilişkiler incelenebilir.

KAYNAKÇA

Aksoy, E., Teker, T., Mazak, M., & Kocabıyık, T. (2020). Kripto paralar ve fiyat ilişkileri üzerine bir analiz: Toda-Yamamoto nedensellik analizi ile bir inceleme. *Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (37), 110-129.

Akpınar, H. (2014). *Data: Veri madenciliği veri analizi*. Papatya Yayıncılık Eğitim.

Ateş, B. A. (2016). Kripto para birimleri, bitcoin ve muhasebesi. *Çankırı Karatekin Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *7*(1), 349-366.

Breiman, L. (2001). *Random forests. Machine learning,* 45(1), 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). Classification and regression trees. Monterey, Calif., USA: Wadsworth.

Cingoz, F., & Kendirli, S. (2019). Altın fiyatları, döviz kuru ve borsa istanbul arasındaki ilişki. *Finans Ekonomi ve Sosyal Araştırmalar Dergisi (FESA)*, *4*(4), 545-554.

Cosgun, E., & Karaağaoğlu, E. (2011). Veri madenciliği yöntemleriyle mikrodizilim gen ifade analizi. *Hacettepe Tıp Dergisi*, *42*, 180-189.

Elmas, B., & Polat, M. (2014). Altın fiyatlarını etkileyen talep yönlü faktörlerin tespiti: 1988-2013 dönemi. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *15*(1), 171-187.

Güleç, Ö. F. (2018). Bitcoin ile finansal göstergeler arasındaki ilişkinin incelenmesi. *Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *7*(2), 18-37.

Gülhan, Ü. (2020). Kovid-19 pandemisinin altın fiyatlarına etkisi: ardl analizi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *34*(3), 1111-1125.

Gültekin, Ö. E., & Hayat, E. A. (2016). Altin fiyatini etkileyen faktörlerin var modeli ile analizi: 2005-2015 dönemi/Analysis of factors affecting the gold prices through var model: 2005-2015 period. *Ege Akademik Bakis*, *16*(4), 611.

Karasu, S., Altan, A., Saraç, Z., & Hacioğlu, R. (2018, May). Prediction of Bitcoin prices with machine learning methods using time series data. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

Klein, T., Thu, H. P., & Walther, T. (2018). Bitcoin is not the new gold–a comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, *59*, 105-116.

Kocatepe, C. İ., & Yıldız, O. (2016). Ekonomik endeksler kullanılarak Türkiye'deki altın fiyatındaki değişim yönünün yapay sinir ağları ile tahmini. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *4*(3).

Kothari, R., & Dong, M. (2001). Decision trees for classification: A review and some new results. In *Pattern recognition: from classical to modern approaches* (pp. 169-184).

Maese, V. A., Avery, A. W., Naftalis, B. A., Wink, S. P., & Valdez, Y. D. (2016). Cryptocurrency: A primer. *Banking LJ*, *133*, 468.

Maimon O., Rokach L. (2010). "Classification Trees". Data Mining and Knowledge Discovery Handbook, Editör: O., Maimon, L. Rokach, Springer, New York, A.B.D., 149-175.

Nakamoto, S. (2019). *Bitcoin: A peer-to-peer electronic cash system*.

Nebil, F. S. (2018). Bitcoin ve kripto paralar. *Pusula Yayıncılık, İstanbul*.

Okuyan, H. A., & Deniz, A. G. D. (2019). Kripto paraların geleneksel finansal varlıklarla ilişkisi. *Scientific Committee*, 108. II. International Conference on Empirical Economics and Social Sciences (ICEESS' 19) June 20-21-22, / Bandirma – Turkey

Özdamar K. (2004). *Paket programlar ile istatistiksel veri analizi*. Eskişehir: Kaan Kitabevi.

Panigrahi, S. S., & Mantri, J. K. (2015, October). Epsilon-SVR and decision tree for stock market forecasting. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)* (pp. 761-766). IEEE.

Lin, C. F., & Wang, S. D. (2002). Fuzzy support vector machines. *IEEE transactions on neural networks*, *13*(2), 464-471.

Yıldırım, H. (2018). Günlük bitcoin ile altın fiyatları arasındaki ilişkinin test edilmesi: 2012-2013 yılları arası johansen eşbütünleşme testi. *Itobiad: Journal of the Human & Social Science Researches*, *7*(4).

Yu, P. S., Chen, S. T., & Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, *328*(3-4), 704-716

Journal of Statistics & Applied Science, Volume -,1 Issue -2
ISSN 2718-0999

Does Wind Energy Affect Economic Growth in Developing Countries?

Emrah DOĞAN[1*], Başak ÖZARSLAN DOĞAN[2],

[1] *İstanbul Gelişim University, International Trade and Finance, İstanbul, Turkey*
[2] *Anadolu University, Economics, Eskişehir, Turkey*
* Correspondence Author, e-mail: emdogan@gelisim.edu.tr

Abstract

Energy is of great importance for the sustainability of social and economic life. Energy consumption in the world has increased gradually depending on the economic development. Fossil fuels, which are frequently used in energy production, cause environmental degradation. Therefore, countries have turned their attention to renewable energy sources such as solar, wind and biomass. In this context, the relationship between economic growth and wind energy use was analyzed for 23 developing countries using the data of 2004-2016 and the Driscoll and Kraay estimator. The findings obtained as a result of the study prove that the increase in wind energy has a positive and significant effect on economic growth in the mentioned countries.

Keywords: Macro Economics, Economic Growth, Renewable Energy, Panel Data Analysis.

Özet

Sosyal ve ekonomik hayatın sürdürülebilmesi için enerji büyük önem taşımaktadır. Dünyada ise enerji tüketimi, ekonomik gelişmeye bağlı olarak giderek artmıştır. Enerji üretiminde sık kullanılan fosil yakıtların çevresel bozulmalara yol açmasından dolayı ise ülkeler dikkatlerini güneş, rüzgar ve biyokütle gibi yenilenebilir enerji kaynaklarına çevirmiştir. Bu kapsamda çalışmada, 23 gelişmekte olan ülke için 2004-2016 yılları arasında rüzgar enerjisinin ekonomik büyüme üzerindeki etkisi Driscoll ve Kraay Tahmincisi aracılığıyla tahmin edilmiştir. Çalışma sonucunda elde edilen bulgular, belirtilen ülkelerde rüzgar enerjisi artışının ekonomik büyüme üzerinde pozitif ve anlamlı bir etkiye sahip olduğunu kanıtlar niteliktedir.

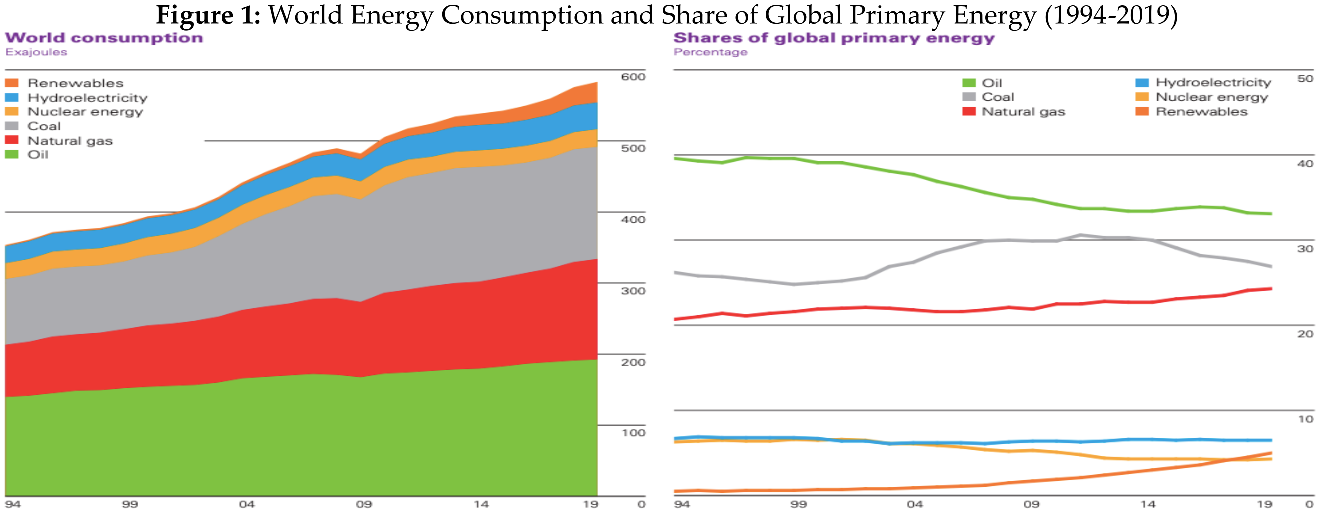
Anahtar Kelimeler: Makro İktisat, Ekonomik Büyüme, Yenilenebilir Enerji, Panel Veri Analizi.

1 INTRODUCTION

Energy is one of the most important factors in building a sustainable future. In this context, there are many issues that are discussed about the future of energy resources. But three of these issues form the basis of energy discussions. The first of these issues is the limited reserves of fossil energy resources and the possibility that they will be exhausted at a later date. The second is that greenhouse gas emissions resulting from the use of fossil fuels increase, causing global warming and climate change. Finally, in economies that are dependent on foreign energy, in a possible energy crisis is the energy supply problem that will occur (Apergis and Danuletiu, 2014, p.578-579; Çalışkan, 2009, p.306; Gültekin and Uğur, 2019: 326).
Fossil energy sources such as coal, oil and natural gas constitute the majority of the energy used in the world today. In In this context, energy consumption according to its source and the shares of these resources in total

energy consumption are expressed in Figure 1. In the light of the data obtained from Figure 1, oil is the most consumed energy source in the world. This is followed by coal, natural gas, hydroelectricity, nuclear energy and other renewable energy sources.

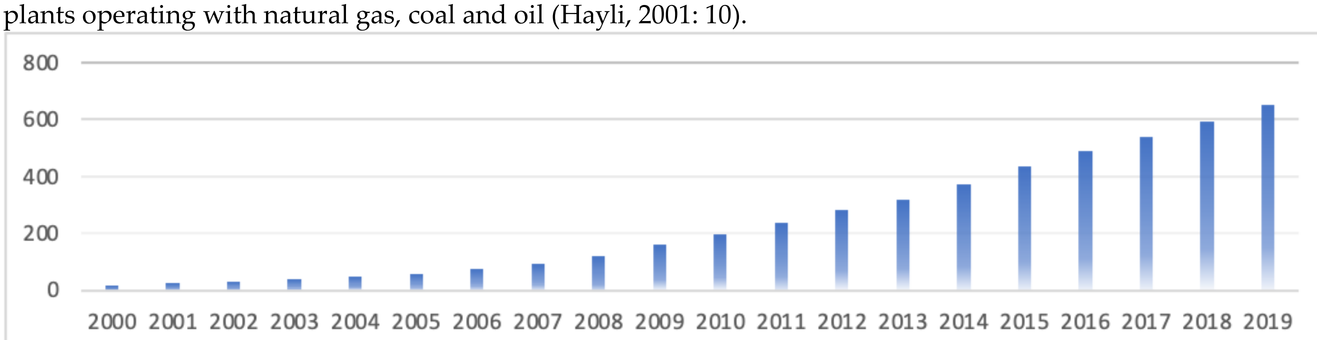
Figure 1: World Energy Consumption and Share of Global Primary Energy (1994-2019)



Source: BP Statistical Review of World Energy Outlook, 2020

In the period between 1994 and 2019, although the use of fossil energy sources is at the top, the consumption of renewable energy resources is increasing every year. So much so that, especially after the 2000s, with the developing technology and the increased awareness of these resources, the interest in renewable energy has increased. Increasing interest in renewable energy has brought with it more investment in this field. In this context, between 2010 and 2019, approximately 2.6 trillion dollars (excluding large-scale hydroelectric) renewable energy investments were made globally. Approximately 41% of this investment belongs to wind energy (Frankfurt School-UNEP Center / BNEF. 2019).

Wind energy is one of the fastest developing completely environmentally friendly energy sources. In addition, it is an energy source with low raw material cost. For example, the cost of electricity generated from wind energy is 50% of the cost of solar and nuclear energy, and approximately 25-30% of electricity produced from thermal power plants operating with natural gas, coal and oil (Hayli, 2001: 10).

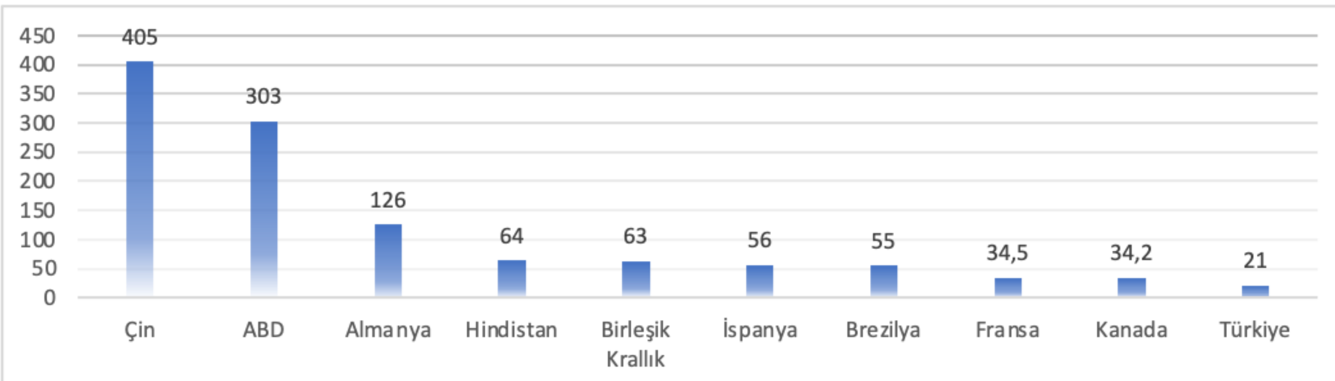


Source: Wind Energy İnternational, 2020 (https://library.wwindea.org/global-statistics/)

Figure 2: Global Installed Wind Energy Capacity (MW) between 2000-2019

Figure 2 shows the global installed power capacity of wind energy between 2000-2019. Accordingly, the installed power capacity, which was approximately 18 thousand MW in 2000, reached approximately 650 thousand MW by 2019. The most important factor in this increase in wind energy installed power capacity is the decrease in wind energy-based electricity generation costs. In wind energy, while the cost reduction in onshore wind has been 39%

since 2010, there has been a 29% reduction in offshore wind (IRENA, 2020).
Figure 3 shows the top ten countries in wind energy generation as of 2019. Accordingly, China ranks first both in terms of the wind energy capacity used in 2015 and the capacity added in 2016. China ranks first in wind energy, as in other renewable energy sources. For many years, China has met its increasing energy need from fossil fuels in parallel with its economic development. In this context, China is trying to eliminate this problem, which gradually turns into energy security due to fossil fuel dependence, by turning to renewable energy sources (Turan, 2019). USA ranks second in wind energy production after China. Wind energy in the country provides more than 20% of the electricity production of 6 states (AWEA, 2020). Germany is third in the world in wind power generation, India fourth and the UK fifth. Turkey ranks tenth in the world.



Source: BP Statistical Review of World Energy Outlook, 2020
Figure 3: Top 10 Countries in Wind Energy Production as of 2019 (TWh)

2 LITERATURE REVIEW

Although there are many studies examining the relationship between renewable energy and economic growth in the economics literature, the number of studies examining the effects of wind energy on economic growth remains limited. At this point, this study is expected to fill the gap in the literature.
The effect of wind energy on economic growth is examined in the literature in two ways. The first of these examines the effect of wind energy on economic growth in theoretical terms. The first of the studies mentioned Koçaslan (2010), for the sustainable development objectives in Turkey conducted a study of the importance of wind energy. Accordingly, in addition to being a completely domestic, renewable and environmentally friendly resource, wind energy is a resource that will meet the total final energy consumption; It has reached the conclusion that it will also contribute to the economic, environmental and social areas.
Pegels and Lütkenhorst (2014) compared wind power and solar energy in five different aspects, such as competitiveness, innovation, job creation, climate change mitigation and cost, within the framework of the German example. Accordingly, they concluded that wind energy meets these five substances better than solar energy.
Khan and Erdoğdu (2012), their study examined the economic benefits of wind energy for Turkey; In case of generating energy from wind instead of fossil fuels, a total of 56.7 million tons less carbon dioxide will be emitted. and that energy will be produced at around 56 million TL cheaper and stated that 112 thousand people will be provided with employment opportunities.
The second study examining the effect of wind energy on economic growth is empirical.For example, Gültekin and Uğur (2019) investigated the macroeconomic determinants of wind energy consumption for OECD countries with the help of panel data analysis for the years 2000-2015. In the study, government final consumption expenditures, government activity, energy use per capita and GDP per capita were used as determinants of wind energy. According to the analysis findings obtained, while public size is not a determinant in wind energy consumption in any country, government efficiency and energy use are the determinants of wind energy in some

countries.

Koç and Apaydın (2020) analyzed the relationship between wind energy and economic growth with the data for the period 1991-2017, using panel data for G-20 countries. According to the results of the analysis, a significant and positive relationship was found between wind energy and economic growth.

Armeanu et al. (2017) measured the impact of renewable energy sources on economic growth in their study including the Panel Data Analysis they conducted for 28 EU Member States in the period between 2003-2014. In this study, they concluded that wind energy, which is one of the renewable energy resources, increased the economic growth by 4%.

Ohler and Fetters (2015) examined the causality relationship between electricity generation from renewable energy sources and GDP in their study using the Panel Error Correction model for the period 1990-2008 and 20 OECD countries. According to the study, it was stated that electricity obtained from biomass energy has a decreasing effect on GDP, while electricity obtained from wind energy and hydraulic energy has an increasing effect on GDP.

Sakarya and Yıldırım (2017) used Monte Carlo Simulation (MCS) model in the evaluation of wind power plant (WPP) investments in their study. By simulating various combinations of input variables affecting the RES investment, determining the net present value (NPV) of the planned project, they concluded that RES investments are economically profitable if appropriate conditions are met.

Atay (2016) analyzed the relationship between wind energy and economic growth in G-7 and G-20 countries with a data set for 2003-2012. Accordingly, unit root tests, cointegration and causality tests applied with the help of panel data set using wind energy consumption and economic growth rates were used in the study. According to the results, a 1% increase in wind energy consumption increases economic growth by 6%

3 DATA SET AND METHODOLOGY

In econometrics-based empirical studies, three types of data are used: cross-section data, time series data and mixed data, which is a combination of these two data sets. If the same section unit is examined in a certain time period, such mixed data are called panel data (Gujarati, 1999). In other words, panel data includes the examination of countries, firms or household units at a certain time scale (Baltagi, 2001: 1). The general equation used in panel data analysis is expressed as follows:

$$Y_{it} = \alpha + X'_{it} + u_{it} \qquad i=1,,,,,N \qquad t=1,,,,,,T \qquad (1)$$

In equation no 1, i = 1,,, represents the data of N number of firms, households or countries, while t = 1,,,, T represents time. The term u_{it} error is assumed to be independent of all units (Sandalcılar, 2012: 8).

Static models established with the help of panel data set mostly use fixed effect and random effect models (Çetin, 2013). In the fixed effects model, each unit that cannot be observed is assumed to be constant with respect to time, while in the random effect model, it is assumed that there are time varying effects around a certain probability distribution and that these effects are unrelated to the explanatory variables of the model (Baltagi, 2005).

Hausman (1978) test is used in the literature to choose between fixed and random effect models. In the Hausman (1978) test, the null hypothesis expressed as H_0 is set up to express the accuracy of the random effect model. As a result of this test, it is concluded that if the Chi-Square probability value is lower than 1%, the model to be used can be preferred as a fixed effect model (Baltagi, 2005). The variables belonging to the model created to examine the effect of wind energy on economic growth are given in Table 2.

Table 2: Variables Used in the Model

Variables	Defining Variables	Source	Expected Sign
GDP	Real GDP	World Bank-WDI	
WIND	Primary Use of Wind Energy	International Renewable Energy Agency- IRENA	+
GFC	Real Fixed Capital Investments	World Bank-WDI	+

LABOR	Labor	World Bank-WDI	+/-

In this study, besides the primary use of wind energy, the effects of real fixed capital investments and labor variables, which are among the important dynamics of growth, on economic growth were investigated. In this context, the study was examined for 23 developing countries using annual data for the period 2004-2016. Countries used in the study; Argentina, Brazil, Bulgaria, Chile, China, Colombia, Costa Rica, Croatia, Egypt, Hungary, India, Iran, Mexico, Morocco, Peru, Philippines, Poland, Romania, Russia, South Africa, Sri Lanka, Turkey and Ukraine . These countries are determined by whether they use wind energy or not.

3.1. Empirical Findings

In this study, the effect of wind energy use on economic growth will be analyzed with the help of the model given below;

The model used in the analysis is expressed below.

$LOGGDP_{it} = \alpha_i + \lambda_t + \beta_1 LOGWIND_{it} + \beta_2 LOGGFC_{it} + \beta_3 LOGLABOR_{it} + \varepsilon_{it}$ (2)

In model number 2 i; t represents the country while t represents time. β, estimation coefficients, αi, country fixed effect, λt is the time constant and εit stands for error term. Descriptive test statistics of the variables used in the study are expressed in Table 3.

Table 3: Descriptive test statistics for variables

Variables	Mean	Standard deviation	Minimum	Maximum
GDP	8.731003	0.677523	6.804615	9.620793
WIND	5.023556	2.498339	0	12.03605
GFC	24.92718	1.389702	22.42275	29.11586
LABOR	16.84755	1.445668	14.41239	20.48374

Before performing panel regression analysis, it is necessary to determine the appropriate model in the study. In this context, first of all, in order to determine which of the fixed and random effects models to be used in the study, it was decided according to the result of Hausman test, which is frequently used in the literature. According to the results of the Hausman Test given in Table 4, it was seen that the most suitable method in this study was the fixed effects model. Following the determination that the appropriate model was the Fixed Effects Model in the study, the modified Wald Test for heteroskedasticity problem that could cause errors and deviations in the estimation results, the Wooldridge autocorrelation test for the autocorrelation problem, and the Pesaran (2004) CD test to determine the presence of cross-sectional independence were used in the study. According to the obtained test results, it was determined that all three heteroskedasticity, autocorrelation and cross-sectional independence problems exist in the model. According to the test findings, Driscoll and Kraay (1998) Standard Errors and Fixed Effects Regression, which are used in the presence of all three of these problems and eliminate these problems, were applied as the final model and the estimated results obtained are given in Table 4.

Table 4: Fixed Effects Model Results with Driscoll and Kraay Standard Errors

Variables	Coefficients
Logwind	0.0254*** (0.00)
Loggfc	0.445*** (0.00)
Loglabor	-0.246* (0.056)
R^2	0.84
Sample Number	299
F-statistic (W)	576.88

Hausman Test statistic	16.86
Modified Wald Test statistic	9184(0.00)
Wooldridge Test statistic	67.31 (0.00)
Pesaran Cross-sectional Independence Test Statistic	4.785 (0.00)

Note: : ( ) refers to the probability values of the variables. and *** p <0.01, ** p <0.05, * p <0.1 represent significance levels

Table 4 shows Driscoll and Kraay estimator results.. According to Fixed Effects Model Results with Driscoll and Kraay Standard Errors , R^2 value is estimated as 0.84. In other words, the power of independent variables to explain the dependent variable is 0.84. The F statistic shows that the model is generally significant. The coefficients of all variables were found to be statistically significant. According to the estimation results, the increase in wind energy positively increases the GDP in accordance with the expectations. In other words, a 1% increase in the use of wind energy increases the GDP by 0.02. Likewise, fixed capital investments have a positive effect on GDP as expected. A 1% increase in fixed capital investments increases GDP by 0.44. On the other hand, the labor variable negatively affects the GDP and a 1% increase in the said coefficient decreases the GDP by 0.24.

4 CONCLUSION

One of the dynamics of a sustainable economy that is thought to change the economic structure of the future is wind energy. Unlike the fossil-based energy resources that are widely used in the world, countries have turned to alternative energy sources in recent years. Wind energy, which is one of these alternatives, increases its importance day by day.

In the study, the effect of wind energy on economic growth between 2004-2016 for 23 developing countries was estimated by means of Driscoll and Kraay estimator. In addition to the primary use of wind energy, labor and real fixed capital investments are also used in the model to measure the relationship between the use of wind energy and economic growth, as they are the dynamics of growth.

In the study, the coefficient for the primary use of wind energy was found to be statistically significant and positive. The demand for renewable energy has increased in recent years and wind energy, which is among the renewable energy sources, is technologically more advantageous than its alternatives, the installation phase is completed in a short time, it can be considered as a factor that increases economic growth due to social and economic contributions in the region where it is established. In addition, wind energy will reduce foreign dependency in energy and thus the decrease in energy import costs will cause an increase in GDP. Labor force, which is another variable considered, constitutes the dynamics of growth. The increase in the labor force in a growing economy is an issue in the economics literature. If productivity in the labor force is low and the wages correspond to a relatively small proportion of the production cost of the firm that will invest in the country, the abundant workforce in that country will not create an opportunity for the investing firm. In this context, 1% increase in labor force in the study reduces economic growth by 0.24%.The last variable used in the study is real fixed capital investments. Like other variables, this variable was found statistically positive and significant in the study. In summary, the study concludes that the use of wind energy and fixed capital investments increase economic growth, while the labor force reduces economic growth.

The rapid increase of the population in developing countries, and the supply of energy used as input in the increasing production process, at high rates, from fossil-based energy sources make these countries highly dependent on foreign countries. This foreign dependency constitutes one of the biggest obstacles to the development of countries. The current account deficit, which is largely given when importing energy, prevents countries from making new investments and causes a restriction on research and development expenditures to be allocated to this field. In order to minimize this negative situation and solve the energy deficit problem, developing countries should form their energy policies in favor of using renewable energy, just like developed countries. Wind energy, on the other hand, is a good alternative for developing countries due to the fact that its installation and maintenance costs are less than other energy sources and the time it can pay off is very short.

REFERENCE

Apergis, N. ve Danuletiu, D. C. (2014). Renewable energy and economic growth: evidence from the sign of panel long-run causality. *International Journal of Energy Economics and Policy*, 4(4), 578-587

Armeanu, D. Ş., Vintila, G., Gherghina, S. C. (2017). "Does Renewable Energy Drive Sustainable Economic Growth? Multivariate Panel Data Evidence for EU-28 Countries", *Energies*, 10 (3): 381.

Atay, G. (2016). "G-7 ve G-20 Ülkelerinde Rüzgâr Enerjisi ve Ekonomik Büyüme Arasındaki ilişkinin Panel Eşbütünleşme Yaklaşımı ile Analizi", *ESOGÜ Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi.*

Baltagi, B. H. (2005). Econometric analysis of data panel. *England: John Wiley & Sons Ltd*.

Baltagi, B. H. (Eds.). (2001). *A companion to theoretical econometrics* (Vol. 1). Oxford: Blackwell.

BP Statistical Review of World Energy Outlook, 2020

Çalışkan, Ş. (2009). Türkiye'nin enerjide dışa bağımlılık ve enerji arz güvenliği sorunu. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (25), 297-310.

Çetin, R. (2013). Ekonomik Özgürlüğün Gelir Düzeyi Üzerindeki Etkisinin Panel Veri Analizi Yöntemiyle İncelenmesi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *8*(2).

Driscoll, J. C., ve Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. Review of economics and statistics, 80 (4), 549-560.

Frankfurt School-UNEP Centre/BNEF. 2020. Global Trends in Renewable Energy Investment 2020, http://www.fs-unep-centre.org

Gujarati, D. N., & Porter, D. C. (1999). *Essentials of econometrics* (Vol. 2). Singapore: Irwin/McGraw-Hill.

Gültekin, E., & Uğur, A. (2019). OECD ülkelerinde yenilenebilir enerji tüketiminin makro ekonomik belirleyicileri: Rüzgâr enerjisi modeli. *Erciyes Üniversitesi Iktisadi ve Idari Bilimler Faküeltesi Dergisi*, (53), 325-342.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271.

Hayli, S. (2001). Rüzgâr enerjisinin önemi, dünya'da ve Türkiye'deki durumu. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, *11*(1), 1-26.

IRENA (2020), Renewable Power Generation Costs in 2019,

Karaca, C., ve Erdoğdu, M. (2012). "Türkiye'de Rüzgâr Çiftliklerinden Elektrik Üretilmesiyle Sağlanabilecek Çevresel ve Ekonomik Kazançlar", *Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 23: 156-188.

Koç, Ü., ve Apaydın, Ş. (2020). İktisadi Büyüme ve Rüzgar Enerjisi: Seçilmiş G-20 Ülkeleri İçin Bir Analiz. *Fiscaoeconomia 4*(3), 595-612

Koçaslan, G. (2010). "Sürdürülebilir Kalkınma Hedefi Çerçevesinde Türkiye'nin Rüzgâr Enerjisi Potansiyelinin Yeri ve Önemi", *Sosyal Bilimler Dergisi*, (1): 53-61.

Ohler, A., And Fetters, I. (2014). The Causal Relationship Between Renewable Electricity Generation and GDP Growth: A Study of Energy Sources", *Energy Economics*, 43: 125-139

Pegels, A., And Lütkenhorst, W. (2014). "Is Germany's Energy Transition a Case of Successful Green Industrial Policy? Contrasting Wind and Solar PV", *Energy Policy*, 74: 522-534.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels.University of Cambridge Faculty of Economics Cambridge Working Papers in Economics No.0435.

Sakarya, Ş. ve Yıldırım, H. H. (2017). "Rüzgâr Enerjisi Santral Yatırımlarının Değerlendirilmesinde Monte Carlo Simülasyonunun Kullanılması", *Maliye Finans yazıları (108).*

Sandalcılar, A. R. (2012). BRIC ülkelerinde ekonomik büyüme ve ihracat arasindaki ilişki: Panel eşbütünleşme ve panel nedensellik. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *17*(1), 161-179.

Turan, İ. (2019) China's Renewable Energy Policy.

Wind Energy İnternational, 2020 (https://library.wwindea.org/global-statistics/)

AWEA(2020). https://www.awea.org/wind-101/basics-of-wind-energy/wind-facts-at-a-glance

Journal of Statistics & Applied Science, Volume 1, Issue 2
ISSN 2718-0999

Inference in "One-Way" Random Designs - Discussing Sub-D and ANOVA based Estimators

Adilson Silva*[1,2]

[1] Faculty of Science and Technology, University of Cape Verde, Praia, Cape Verde

[2]CMA, Faculty of Science and Technology, New University of Lisbon, Caparica, Portugal

*Correspondence Author, e-mail: adilson.dasilva@docente.unicv.edu.cv

Abstract

Recently it was shown through simulations studies that Sub-D produces estimates with unbiased and lower variance-covariance estimates than the ANOVA-based estimator, except in case of random "one-way" balanced designs. In this designs the simulations studies suggested that they have the same variance-covariance estimates. This paper aims to compare the common ANOVA-based estimator to Sub-D in random "one-way" designs with two groups of treatment and in random "one-way" balanced designs. The comparison will be conducted through theoretical results and corroborated with simulation studies. It will be proved that the ANOVA-base estimator and Sub-D have exactly the same variance-covariance estimates in both above referred designs. The proof will be given firstly for random "one-way" designs with two groups of treatment and then for random "one-way" balanced designs.

Keywords: Sub-D, ANOVA, Variance Components, One-way Designs.

1. Introduction

Due to necessity of incorporate the amount of variations caused by certain uncontrollable sources of variations in statistical designs with fixed effects, for example the amount of variations within and/or between groups of treatments for that the experimenters are not able to control and those whose the levels must be randomly selected, in research field such as genetic, agriculture, animal breeding, and quality control and improvement, in early 1960 several designs with both fixed and random effects terms were introduced and widely investigated (see Khuri [4] and Silva [11]).

Among those designs we highlight the well known and widely discussed random "one-way" designs:

$$z_{ij} = \mu + \alpha_i + \epsilon_{ij},\ i = 1, \ldots, k;\ j = 1, \ldots, n_i, \tag{1.1}$$

$$\text{where} \begin{cases} k \ \text{ is the the number of groups of treatment;} \\ n_i \text{ is the number of observations within the } i\text{th group of treatment;} \\ \mu \ \text{ is the general mean (the fixed effect);} \\ \alpha_i \text{ is the random effect due to the } i\text{th group of treatment;} \\ \epsilon_{ij} \text{ is the random error due to the } j\text{th observation within the } i\text{th} \\ \quad \text{group of treatment.} \end{cases}$$

It is assumed that:

$$\begin{cases} \alpha_i \sim (0,\ \gamma_\alpha), \ \text{ that is } \alpha_i\text{'s are i.i.d. with mean zero and variance } \gamma_\alpha; \\ \epsilon_{ij} \ \sim (0,\ \gamma_\epsilon), \text{ that is } \epsilon_{ij}\text{'s are i.i.d. with mean zero and variance } \gamma_\epsilon; \\ cov(\alpha_i,\ \epsilon_{ij}) = 0,\ i = 1, \ldots, k \text{ and } j = 1, \ldots, n_i. \end{cases}$$

When all groups of treatment have the same number of observations, that is $n_i = n$, the model 1.1 is called random "one-way" balanced design. Otherwise it is called random "one-way" designs.

Random "one-way" designs are useful tools for modeling repeated measured data and, in particular, small sample and longitudinal data (see Wallace [15] and Khuri et al. [5]). For this designs several techniques and tools focussing on variance components estimation has been developed. Among than the most popular are those based on likelihood and ANOVA (see Demidenko [1] and Pinheiro and Bates [7], for instance). Recently, while doing research for his PhD Thesis, Silva (2017) developed a new estimator for variance components named Sub-D (see Silva [11], [12], [13] and Ferreira et al. [2]). On its approach Silva constructed and applied a finite sequence of orthogonal transformations (which he called sub-diagonalizations) to the covariance structure of the restricted design producing a set of sub-models which he used to create pooled estimators for the variance components.

Through simulations it was Shown that Sub-D produces very realistic estimative in random "one-way" balanced and unbalanced designs (see Silva [12]); in nested and crossed "two-way" unbalanced designs (see Silva [11]); and in nested "three-way" unbalanced designs (See Silva et al.[13]). In fact, the numerical simulation show that Sub-D produces reasonable and comparable estimates, sometimes slightly better than those obtained with REML and mostly better than tose obtained with Anova. However, due to the correlation between the sub-models on it's foundation, the variability of estimates produced with Sub-D is slightly greater then tose obtained with REML except in random "one-way" balanced designs. But, when compared with Anova, Sub-D produces estimates with unbiased and lower variance estimates than Anova-based estimator except in case of random "one-way" balanced designs. In this case, simulations studies suggested that Sub-D and Anova-based estimator has the same variance. Thus, this work aims to prove through theoretical results that for this designs Anova-based estimator and Sub-D have exactly the same variance. Moreover, this work also aims to propose a correction for a result in the deduction of one of the Sub-D's estimators for variance components estimators given in Silva [12].

First section is devoted to the introduction, and the second one to the background. Thirty section is reserved to prove that Anova-based estimator and Sub-D has exactly the same variance-covariance in random "one-way" balanced designs. Forth section is reserved to simulations studies, and the last one for the discussions.

From now on, the following *notations* will be used without any additional comments:

- $P_{R(X)}$ denotes the projection matrix onto the subspace spanned by the columns of a matrix X and $P_{R(X)^\top}$ the projection matrix onto the orthogonal complement of the subspace spanned by the columns of X;

- $\Sigma(x)$ denotes the variance-covariance matrix of a random vector x, i.e
$\Sigma(x) = Eb \quad \top$
$\mathbf{0}_{n,m}$ denotes an $n \times m$ matrix, while $\mathbf{0}_n$ denotes a null vector of dimension n; $\mathbf{1}_n$ denotes a vector of ones having both dimension n;
$\mathbf{J}_n$ denotes a $n \times n$ matrix of ones;
$z \sim (w, \Sigma)$ denotes a random vector z with mean w, and variance-covariance matrix Σ;
$z \sim N(w, \Sigma)$ denotes a random vector z with a normal distribution with mean w, and variance-covariance matrix Σ;
$r(A)$ denotes the rank of a matrix A;
$tr(A)$ denotes the trace of a matrix A
$\sum_{i \neq j}^{n}$ denotes $\sum_{i=1}^{n} \sum_{j=1}^{n}$ for $i \neq j$.

2. The Estimators: Anova and Sub-D

In this section we introduce and briefly discuss Sub-D and ANOVA-based estimators on design 1.1. Their MSE will be discussed. We will focus on case when the design has two groups of treatment, i.e $k = 2$, as well as the case when the design is balanced, that is $n_i = n, i = 1, \ldots, k$.

2.1. ANOVA-based Estimator

The analysis of variance (ANOVA) method of estimating the variance components γ_α and γ_ϵ in model 1.1 consists of equating observed values of the between group mean squares (MS_B) and within group mean square (MS_W) to their expected values, and solving the resulting equations for γ_α and γ_ϵ. This method produces unbiased estimators of γ_α and γ_ϵ. Such estimators are respectively given as

$$\begin{aligned}
\widehat{\gamma_\alpha}^A &= \frac{1}{n_o}(MS_B - MS_W) \\
&= \frac{1}{n_o}\left[\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{z_{i\bullet}} - \bar{z_{\bullet\bullet}})^2 - \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \bar{z_{i\bullet}})^2\right] \text{ and} \\
\widehat{\gamma_\epsilon}^A &= MS_W = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \bar{z_{i\bullet}})^2, \qquad (2.1)
\end{aligned}$$

where $n_o = \frac{N^2 - \sum_{i=1}^k n_i^2}{N(k-1)}$, $N = \sum_{i=1}^k n_i$, $\bar{z_{i\bullet}} = \sum_{j=1}^{n_i} \frac{z_{ij}}{n_i}$ and $\bar{z_{\bullet\bullet}} = \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{z_{ij}}{N}$.

Following Searle [8], [9] (see Sahai and Ojeda [3]) the variance of ANOVA estimators $\widehat{\gamma_\alpha}^A$ and $\widehat{\gamma_\epsilon}^A$, are respectively given as

$$\begin{aligned}
\Sigma\left(\widehat{\gamma_\alpha}^A\right) &= \frac{2\gamma_\alpha^2}{\left(N^2 - \sum_{i=1}^k n_i^2\right)^2}\left[N^2\sum_{i=1}^{k} n_i^2 + \left(\sum_{i=1}^{k} n_i^2\right)^2 - 2N\sum_{i=1}^{k} n_i^3\right] \\
&\quad + \frac{4N\gamma_\alpha\gamma_\epsilon}{\left(N^2 - \sum_{i=1}^k n_i^2\right)} + \frac{2\gamma_\epsilon^2 N^2(N-1)(k-1)}{\left(N^2 - \sum_{i=1}^k n_i^2\right)^2(N-k)} \text{ and} \\
\Sigma\left(\widehat{\gamma_\epsilon}^A\right) &= \frac{2\gamma_\epsilon^2}{N-k}. \qquad (2.2)
\end{aligned}$$

Numerical studies carried out by Singh [14] and Caro et al. [6] for different configurations of γ_α and γ_ϵ suggested that the unbalancedness of the data results in an increase of variance-covariance of $\Sigma\left(\widehat{\gamma_\alpha}^A\right)$ and $\Sigma\left(\widehat{\gamma_\epsilon}^A\right)$. Khuri et al. [5] proved that $\Sigma\left(\widehat{\gamma_\alpha}^A\right)$ attains its minimum for all γ_α and γ_ϵ when the data are balanced.

2.2. Sub-D

Lets take the matrix formulation of design (1.1):

$$z = \mu 1_N + Z\beta + \epsilon, \qquad (2.3)$$

where

$$Z = \begin{bmatrix} 1_{n_1} & 0_{n_1} & 0_{n_1} & \ldots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & 0_{n_2} & \ldots & 0_{n_2} \\ 0_{n_3} & 0_{n_2} & 1_{n_3} & \ldots & 0_{n_3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{n_k} & 0_{n_k} & 0_{n_k} & \ldots & 1_{n_k} \end{bmatrix}, \ \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} \text{ and } \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{kn_k} \end{bmatrix}, \qquad (2.4)$$

with $\beta \sim (0_k, \gamma_\alpha I_k)$, $\epsilon \sim (0_N, \gamma_\epsilon I_N)$ and β and ϵ mutually independent. Thus the model (2.3) may be rewritten as follow

$$z \sim \left(\mu 1_N,\ \gamma_\alpha ZZ^\top + \gamma_\epsilon I_N\right). \tag{2.5}$$

Let B be the $N \times (N-1)$ matrix whose columns are the $N-1$ orthonormal eigenvectors associated to the null eigenvalue of $\frac{1}{N}J_N$, where J_N denotes an $N \times N$ matrix of 1's. Using B it is possible to define (see Silva [12]) a new design (a restricted one) by projecting the design (2.5) onto the orthogonal complement of the vectorial subspace spanned by $\mu 1_N$, as follow

$$y = B^\top z \sim \left(\mu 0_{N-1},\ \gamma_\alpha M + \gamma_\epsilon I_{N-1}\right), \text{where } M = B^\top ZZ^\top B. \tag{2.6}$$

Now let A_i be the matrix whose rows are the set of $g_i = r(A_i)$ orthonormal eigenvectors associated to the eigenvalue θ_i, $i = 1, \ldots, h$, of M; Let also $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\epsilon}^S$ denote the Sub-D estimator of γ_α and γ_ϵ, respectively. Thus, following Silva[5], we that

$$\begin{aligned}\widehat{\gamma_\alpha}^S &= \frac{1}{h^*}\sum_{i=1}^{h}\theta_i\left(hy^\top P_i y - \sum_{j=1}^{h} y^\top P_j y\right)\\ &= y^\top \Lambda_\alpha y,\end{aligned} \tag{2.7}$$

where $\Lambda_\alpha = \frac{1}{h^*}\sum_{i=1}^{h}\theta_i\left(hP_i - \sum_{j=1}^{h}P_j\right)$, $h^* = h\sum_{i=1}^{h}\theta_i^2 - \left(\sum_{i=1}^{h}\theta_i\right)^2$, and $P_i = \frac{A_i^\top A_i}{g_i}$, and

$$\begin{aligned}\widehat{\gamma_\epsilon}^S &= \frac{1}{h^*}\sum_{i=1}^{h}\theta_i\left(\theta_i\sum_{j=1}^{h} y^\top P_j y - \sum_{j=1}^{h}\theta_j y^\top P_j y\right)\\ &= y^\top \Lambda_\epsilon y,\end{aligned} \tag{2.8}$$

where $\Lambda_\epsilon = \frac{1}{h^*}\sum_{i=1}^{h}\theta_i\sum_{j=1}^{h}\left(\theta_i - \theta_j\right)P_j$.

2.2.1. The Correct Version of Sub-D. Unfortunately, it seems that the algebraic manipulation at the time of Sub-D's deduction did not work as well as Silva [12] wished since we found that his deduction of $\widehat{\gamma_\epsilon}^S$ is wrong. The correct one is the one we presented here at (2.8). It worth to remark that:

(1) The above elucidated error in the deduction of $\widehat{\gamma_\epsilon}^S$ at Silva[5] (Section 3) lies on (the wrong) computation of $(\Theta^\top\Theta)^{-1}$. Indeed, with $\Theta = \begin{bmatrix}\theta_1 & 1\\ \vdots & \vdots\\ \theta_h & 1\end{bmatrix}$, we found that

$$\Theta^\top\Theta = \begin{bmatrix}\sum_{i=1}^{h}\theta_i^2 & \sum_{i=1}^{h}\theta_i\\ \sum_{i=1}^{h}\theta_i & h\end{bmatrix} \text{ so that } (\Theta^\top\Theta)^{-1} = \frac{1}{h^*}\begin{bmatrix}h & -\sum_{i=1}^{h}\theta_i\\ -\sum_{i=1}^{h}\theta_i & \sum_{i=1}^{h}\theta_i^2\end{bmatrix}, \tag{2.9}$$

but unfortunately a miscalculation led Silva[5] to find $\frac{1}{h^*}\begin{bmatrix}h & -\sum_{i=1}^{h}\theta_i\\ -\sum_{i=1}^{h}\theta_i & \sum_{i=1}^{h}\theta_i\end{bmatrix}$ for $(\Theta^\top\Theta)^{-1}$ instead of the equation at right side of (2.9), which on it's turn let to a wrong deduction of $\widehat{\gamma_\epsilon}^S$.

(2) The miscalculation in the deduction of $\widehat{\gamma_\epsilon}^S$ did not reflected in the section 'Numerical Example' of Silva[5], since the computation of $(\Theta^\top\Theta)^{-1}$ was done through a software (R).

From now on we refer to the correct version of $\widehat{\gamma_\epsilon}^S$ given in (2.8).

The next Theorem proposes the variance-covariance of both $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\epsilon}^S$.

Theorem 2.1. *Let* $\lambda_s = h^2 \sum_{i=1}^h \frac{\theta_i^s}{g_i} - 2h \sum_{i=1}^h \theta_i \sum_{i=1}^h \frac{\theta_i^{s-1}}{g_i} + \left(\sum_{i=1}^h \theta_i\right)^2 \sum_{i=1}^h \frac{\theta_i^{s-2}}{g_i}$, $s = 2, 3, 4$. *Then:*

(a) $\Sigma\left(\widehat{\gamma_\alpha}^S\right) = \frac{2\gamma_\alpha^2}{h^{*2}}\lambda_4 + \frac{4\gamma_\alpha\gamma_\epsilon}{h^{*2}}\lambda_3 + \frac{2\gamma_\epsilon^2}{h^{*2}}\lambda_2$;

(b) $\Sigma\left(\widehat{\gamma_\epsilon}^S\right) = \frac{2}{h^{*2}} \sum_{j=1}^h \left[\frac{\left(\sum_{i=1}^h \theta_i(\theta_i - \theta_j)\right)^2}{g_j}\right] \left(\gamma_\alpha^2\theta_j^2 + 2\gamma_\alpha\gamma_\epsilon\theta_j + \gamma_\epsilon^2\right)$.

Proof. (See Shayle et al. [10] for variance-covariance of a quadratic form) Part (a):

$$
\begin{aligned}
\Sigma\left(\widehat{\gamma_\alpha}^S\right) &= 2tr\left(y^\top \Lambda_\alpha y\right) = 2tr\left[(\Lambda_\alpha(\gamma_\alpha M + \gamma_\epsilon))^2\right] \\
&= 2\gamma_\alpha^2 tr\left[(\Lambda_\alpha M)^2\right] + 4\gamma_\alpha\gamma_\epsilon tr\left[\Lambda_\alpha M \Lambda_\alpha\right] + 2\gamma_\epsilon^2 tr\left[\Lambda_\alpha^2\right] \\
&= \frac{2\gamma_\alpha^2}{(h^*)^2}\left[h^2 \sum_{i=1}^h \frac{\theta_i^4}{g_i} - 2h\sum_{i=1}^h \theta_i \sum_{i=1}^h \frac{\theta_i^3}{g_i} + \left(\sum_{i=1}^h \theta_i\right)^2 \sum_{i=1}^h \frac{\theta_i^2}{g_i}\right] \\
&\quad + \frac{4\gamma_\alpha\gamma_\alpha}{(h^*)^2}\left[h^2 \sum_{i=1}^h \frac{\theta_i^3}{g_i} - 2h\sum_{i=1}^h \theta_i \sum_{i=1}^h \frac{\theta_i^2}{g_i} + \left(\sum_{i=1}^h \theta_i\right)^2 \sum_{i=1}^h \frac{\theta_i}{g_i}\right] \\
&\quad + \frac{2\gamma_\alpha^2}{(h^*)^2}\left[h^2 \sum_{i=1}^h \frac{\theta_i^2}{g_i} - 2h\sum_{i=1}^h \theta_i \sum_{i=1}^h \frac{\theta_i}{g_i} + \left(\sum_{i=1}^h \theta_i\right)^2 \sum_{i=1}^h \frac{1}{g_i}\right] \\
&= \frac{2}{(h^*)^2}\left(\lambda_4\gamma_\alpha^2 + 2\lambda_3\gamma_\alpha\gamma_\epsilon + \lambda_2\gamma_\epsilon^2\right). \qquad (2.10)
\end{aligned}
$$

Part (b):

$$
\begin{aligned}
\Sigma\left(\widehat{\gamma_\epsilon}^S\right) &= 2tr\left(y^\top \Lambda_\epsilon y\right) \\
&= 2\gamma_\alpha^2 tr\left[(\Lambda_\epsilon M)^2\right] + 4\gamma_\alpha\gamma_\epsilon tr\left[\Lambda_\epsilon M \Lambda_\epsilon\right] + 2\gamma_\epsilon^2 tr\left[\Lambda_\epsilon^2\right] \\
&= \frac{2\gamma_\alpha^2}{(h^*)^2}\sum_{j=1}^h \frac{\theta_j^2}{g_j}\left(\sum_{i=1}^h \theta_i(\theta_i - \theta_j)\right)^2 + \frac{4\gamma_\alpha\gamma_\epsilon}{(h^*)^2}\sum_{j=1}^h \frac{\theta_j}{g_j}\left(\sum_{i=1}^h \theta_i(\theta_i - \theta_j)\right)^2 \\
&\quad + \frac{\gamma_\epsilon^2}{(h^*)^2}\sum_{j=1}^h \frac{1}{g_j}\left(\sum_{i=1}^h \theta_i(\theta_i - \theta_j)\right)^2 \\
&= \frac{2}{(h^*)^2}\sum_{j=1}^h \left[\frac{\left(\sum_{i=1}^h \theta_i(\theta_i - \theta_j)\right)^2}{g_j}\right]\left(\gamma_\alpha^2\theta_j^2 + 2\gamma_\alpha\gamma_\epsilon\theta_j + \gamma_\epsilon^2\right). \qquad (2.11)
\end{aligned}
$$

□

3. Estimation in Designs with two groups of treatments

It is not so evident a strict comparison between the variance-covariance of Sub-D and Anova-based estimators, but when the design has a fixed $k = 2$ groups of treatment, no matter the number of observation for each group, it seems that they are somehow comparable.

When $k = 2$ it follows that $N = n_1 + n_2$ and $n_0 = \frac{N^2-(n_1^2+n_2^2)}{N^2}$, and so the ANOVA-based estimators reduce to

$$\begin{aligned}
\widehat{\gamma_\alpha}^A &= \frac{1}{n_0}\left[n_1(\bar{z_{1\bullet}} - \bar{z_{\bullet\bullet}})^2 + n_2(\bar{z_{2\bullet}} - \bar{z_{\bullet\bullet}})^2)\right] \\
&\quad -\frac{1}{n_o(N-2)}\left[\sum_{j=1}^{n_1}\left(z_{1j} - \bar{z_{1\bullet}}\right)^2 + \sum_{j=1}^{n_2}\left(z_{2j} - \bar{z_{2\bullet}}\right)^2\right] \text{ and} \\
\widehat{\gamma_\epsilon}^A &= \frac{1}{N-2}\left[\sum_{j=1}^{n_1}\left(z_{1j} - \bar{z_{1\bullet}}\right)^2 + \sum_{j=1}^{n_2}\left(z_{2j} - \bar{z_{2\bullet}}\right)^2\right].
\end{aligned}$$

As we may easily conclude, their respective variance-covariance will be given as

$$\begin{aligned}
\Sigma\left(\widehat{\gamma_\alpha}^A\right) &= 2\gamma_\alpha^2 + \left(\frac{2N}{n_1n_2}\right)\gamma_\alpha\gamma_\epsilon + \frac{N^2(N-1)}{2(n_1n_2)^2(N-2)}\gamma_\epsilon^2 \text{ and} \\
\Sigma\left(\widehat{\gamma_\epsilon}^A\right) &= \frac{2\gamma_\epsilon^2}{N-2}. \qquad (3.1)
\end{aligned}$$

When $k = 2$, it follows that $h = 2$, that is M will only have two eigenvalues, θ_1 and θ_2, and since $r(M) = k - 1 = 1$ it follows that $\theta_2 = 0$. Under these conditions we have that

$$\Lambda_\alpha = \frac{P_1 - P_2}{\theta_1} \text{ and } \Lambda_\epsilon = P_2, \qquad (3.2)$$

and therefore the estimators boils down to

$$\widehat{\gamma_\alpha}^S = y^\top\left(\frac{P_1 - P_2}{\theta_1}\right)y \text{ and } \widehat{\gamma_\epsilon}^S = y^\top P_2 y.$$

The results for their respective variance-covariance follow as a consequente of Theorem 2.1.

Corollary 3.1. *Consider the conditions of Theorem 2.1, and let Let $k = 2$. Then,*

(a) $\Sigma\left(\widehat{\gamma_\alpha}^S\right) = 2\gamma_\alpha^2 + \frac{4}{\theta_1}\gamma_\alpha\gamma_\epsilon + 2\left(\frac{g_2+1}{g_2\theta_1^2}\right)\gamma\epsilon^2$;

(b) $\Sigma\left(\widehat{\gamma_\epsilon}^S\right) = \frac{2\gamma_\epsilon^2}{g_2}$.

Proof. Nothing that $h = 2$, and so $g_1 = 1$ and $\theta_2 = 0$, and applying Theorem 2.1 the results follow. □

It worth to notice that since both Sub-D and Anova-based estimators are unbiased their respective mean square error (MSE) are equal to their respective variance-covariance. This remark allows us to infer about the quality of these estimators.

Remark 3.1. *With $MSE(\hat{q})$ denoting the MSE of an estimator $\hat{q}$ of a parameter q, we notice the following:*

- *Sub-D: $MSE\left(\widehat{\gamma_\alpha}^S\right) = \Sigma\left(\widehat{\gamma_\alpha}^S\right)$ and $MSE\left(\widehat{\gamma_\epsilon}^S\right) = \Sigma\left(\widehat{\gamma_\epsilon}^S\right)$;*
- *Anova: $MSE\left(\widehat{\gamma_\alpha}^A\right) = \Sigma\left(\widehat{\gamma_\alpha}^A\right)$ and $MSE\left(\widehat{\gamma_\epsilon}^A\right) = \Sigma\left(\widehat{\gamma_\epsilon}^A\right)$.*

The next result gives a comparative framework of the estimators in design with two groups of treatment.

Proposition 3.1. *Let $k = 2$. Then:*

(a) $MSE\left(\widehat{\gamma_\epsilon}^S\right) = MSE\left(\widehat{\gamma_\epsilon}^A\right)$;

(b) $MSE\left(\widehat{\gamma_\alpha}^S\right) = MSE\left(\widehat{\gamma_\alpha}^A\right)$, *if* $\theta_1 = \frac{2n_1n_2}{n_1+n_2}$.

Proof. These results are consequences of Corollary 3.1. Indeed, since $r(M) = 1$ we have that $g_1 = k - 1 = 1$ and $g_2 = N - K = N - 2$, so that

$$\frac{4}{\theta_1} = \frac{2(n_1 + n_2)}{n_1 n_2} = \frac{2N}{n_1 n_2} \text{ and } \frac{2(g_2 + 1)}{g_2 \theta_1^2} = \frac{N^2(N-1)}{2\left(n_1 n_2\right)^2 (N-2)}, \tag{3.3}$$

provide $\theta_1 = \frac{2n_1 n_2}{n_1 + n_2}$. □

The condition $\theta_1 = \frac{2n_1 n_2}{n_1 + n_2}$ for which $MSE\left(\widehat{\gamma_\alpha}^S\right) = MSE\left(\widehat{\gamma_\alpha}^A\right)$ imposed in Proposition 3.1 consists in a measure to compare the quality of estimators, in the sense that if $\theta_1 < \frac{2n_1 n_2}{n_1 + n_2}$ it holds that Sub-D is better than Anova-based estimator for γ_α and Anova based estimator is better if $\theta_1 > \frac{2n_1 n_2}{n_1 + n_2}$. In fact, as we may see through simulations studies (see tables 1 and 2),

$$\theta_1 = \frac{2n_1 n_2}{n_1 + n_2}$$

whatever the values of n_1 and n_2, and so $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\alpha}^A$ have exactly the same MSE.

For some combinations of parameters γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$ we simulated $s = 10000$ repeated designs, using $\beta \sim \mathcal{N}(0,\ \gamma_\alpha)$ and $e \sim \mathcal{N}(0,\ \gamma_\epsilon)$, and $n_1 = 101$ and $n_2 = 20$. For each simulated design both estimators was applied and the parameters γ_α and γ_ϵ was estimated. Next, the average of the estimated values for the parameters was computed as well as the standard deviations of the respective estimated values. See the results in Tables 1 and 2 and an R function for simulating both estimators in tables 3. As we may see, independently of the configuration for the parameters γ_α and γ_ϵ as well as the configuration for the number of elements in each groups of treatment, the estimates and the respective standard deviations found are the same for both estimators.

Table 1. Simulations for different values of γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, with $n_1 = 101$, $n_2 = 20$ and $s = 10000$. **Actual value** denotes the actual values of the parameters; **Estimate** denotes the estimated values of the parameters; **Stand. Dev.** denotes the standard deviations of the estimated values.

Sub-D	γ_α	γ_ϵ	**ANOVA**	γ_α	γ_ϵ
Actual value	**0.5**	**1**	**AV**	**0.5**	**1**
Estimate	0.50129	0.99912	**Estimate**	0.50129	0.99912
Stand. Dev.	0.75534	0.12768	**Stand. Dev.**	0.75534	0.12768
Actual value	**1**	**0.5**	**AV**	**1**	**0.5**
Estimate	0.99809	0.50001	**Estimate**	0.9980	0.50001
Stand. Dev.	1.42519	0.06520	**Stand. Dev.**	1.42519	0.06520
Actual value	**0.75**	**0.5**	**AV**	**0.75**	**0.5**
Estimate	0.756304	0.50095	**Estimate**	0.75630	0.50095
Stand. Dev.	1.08095	0.06576	**Stand. Dev.**	1.08095	0.06576
Actual value	**0.5**	**0.75**	**AV**	**0.5**	**0.75**
Estimate	0.50144	0.74989	**Estimate**	0.50144	0.74989
Stand. Dev.	0.75143	0.09628	**Stand. Dev.**	0.75143	0.09628
Actual value	**0.5**	**0.1**.	**AV**	**0.5**	**0.1**
Estimate	0.49695	0.10004	**Estimate**	0.49695	0.10004
Stand. Dev.	0.71919	0.01301	**Stand. Dev.**	0.71919	0.01301
Actual value	**0.1**	**0.5**.	**AV**	**0.1**	**0.5**
Estimate	0.10171	0.50099	**Estimate**	0.10171	0.50099
Stand. Dev.	0.16582	0.06458	**Stand. Dev.**	0.16582	0.06458

Table 2. Simulations for different values of γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, with $n_1 = 20$, $n_2 = 101$ and $s = 10000$. **Actual value** denotes the actual values of the parameters; **Estimate** denotes the estimated values of the parameters; **Stand. Dev.** denotes the standard deviations of the estimated values.

Sub-D	γ_α	γ_ϵ	**ANOVA**	γ_α	γ_ϵ
Actual value	**0.5**	**1**	**AV**	**0.5**	**1**
Estimate	0.50746	1.00002	**Estimate**	0.50746	1.00002
Stand. Dev.	0.75025	0.12910	**Stand. Dev.**	0.75025	0.12910
Actual value	**1**	**0.5**	**AV**	**1**	**0.5**
Estimate	1.00721	0.50095	**Estimate**	1.00721	0.50095
Stand. Dev.	1.44209	0.06572	**Stand. Dev.**	1.44209	0.06572
Actual value	**0.75**	**0.5**	**AV**	**0.75**	**0.5**
Estimate	0.74427	0.50020	**Estimate**	0.74427	0.50020
Stand. Dev.	1.07430	0.06574	**Stand. Dev.**	1.07430	0.06574
Actual value	**0.5**	**0.75**	**AV**	**0.5**	**0.75**
Estimate	0.50204	0.75085	**Estimate**	0.50204	0.75085
Stand. Dev.	0.74323	0.09775	**Stand. Dev.**	0.74323	0.09775
Actual value	**0.5**	**0.1**.	**AV**	**0.5**	**0.1**
Estimate	0.50690	0.09980	**Estimate**	0.50690	0.09980
Stand. Dev.	0.71661	0.01299	**Stand. Dev.**	0.71661	0.01299
Actual value	**0.1**	**0.5**.	**AV**	**0.1**	**0.5**
Estimate	0.10217	0.49945	**Estimate**	0.10217	0.49945
Stand. Dev.	0.16710	0.06501	**Stand. Dev.**	0.16710	0.0650

Table 3. The R Code applied to Simulate and test Sub-D and ANOVA-based estimators in an unbalanced "one-way" random with two groups of treatments. Tables 1 and 2 show some examples.

With regard to the optimality of design (1.1) Sub-D allows to set theoretical and consistent results. Optimality designs provide accurate statistical inference by choosing the number of groups of treatments and number of observations at each group in oder to minimize the variance of estimating interested parameters, such as $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\alpha}^A$, which is our case.

According with Corollary 3.1,

$$\begin{aligned} \Sigma\left(\widehat{\gamma_\alpha}^S\right) &= 2\gamma_\alpha^2 + \frac{4}{\theta_1}\gamma_\alpha\gamma_\epsilon + \frac{2(N-1)}{(N-2)\theta_1^2}\gamma\epsilon^2 \text{ and} \\ \Sigma\left(\widehat{\gamma_\epsilon}^S\right) &= \frac{2\gamma_\epsilon^2}{N-2}, \end{aligned} \tag{3.4}$$

recalling $g_2 = N - 2$ and $\theta_1 = \frac{2n_1n_2}{n_1+n_2}$

Noting that θ_1 depends on N through n_1 and n_1, and $\frac{N-1}{N-2} \approx 1$ providing N is a large natural number, results in (3.4) allow us to remark that the bigger is θ_1 the smaller are $\Sigma\left(\widehat{\gamma_\alpha}^S\right)$ and $\Sigma\left(\widehat{\gamma_\epsilon}^S\right)$. More over, it can be proved that θ_1 is not greater than the maximum of n_1 and n_2.

Proposition 3.2. *Whatever n_1 and n_2,*

$$\theta_1 \leq max\{n_1, n_2\}.$$

Proof. Firstly, lets suppose $n_1 = n_2$. Then $\theta_1 = \frac{2n_1^2}{n_1} = n_1$.

Now, without lost of generality, let $n_1 > n_2$. Then there exists a natural number b holding $0 < b \leq n_1$ such that $n_2 = n_1 - b$. Thus,

$$\theta_1 = \frac{2n_1^2 - 2bn_1}{2n_1 - b}. \tag{3.5}$$

By contradiction, suppose $\theta_1 > n_1$, i.e.

$$\frac{2n_1^2 - 2bn_1}{2n_1 - b} > n_1 \leftrightarrow -bn_1 > 0,$$

which is an absurd since by definition $n_1 > 0$ and $b > 0$. Therefore, θ cannot be greater than n_1. For the case when $n_2 > n_1$ we proceed identically. □

The proof of Proposition 3.2 provide a robust tool to discuss the optimality of design (1.1) with respect to Sub-D. In fact, supposing (with out lost of generality) that $n_1 \geq n_2$ and so $n_2 = n_1 - b$ and $\theta = \frac{2n_1n_2}{n_1+n_2}$, for some natural b, we easily prove that

$$\theta_1 \longrightarrow n_1 \text{ as } b \longrightarrow 0. \tag{3.6}$$

In practice, this means the "more balanced" the model is the smaller the variances of $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\epsilon}^S$ are. In order to do that lets consider the real function $t(b) = \theta_1 = \frac{2n_1^2-2bn_1}{2n_1-b}$, $0 \leq b \geq n_1$. Thus, we found the following: since $t'(b) = \frac{-2n_1^2}{(2n_1-b)^2} < 0$ (meaning that θ_1 is a decreasing function of b) and $t''(b) = \frac{-4n_1^2}{(2n_1-b)^3} < 0$ (meaning that θ_1 is a face-down concavity function of b), results (3.6) follows. $t'(b)$ and $t''(b)$ denote the first and second derivate of function $t(b)$ at b, respectively.

4. Estimation in Balanced "One-Way" Designs

For random "one-way" balanced designs, that is the case when $n_i = n$, $i =, 1 \ldots, k$, the ANOVA estimators for variance components γ_1 and γ_2, are given as (see Sahai and Ojeda [3]).

$$\begin{aligned}
\widehat{\gamma_\alpha}^{Ab} &= \frac{1}{n}\left[\left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n(\bar{z_{i\bullet}} - \bar{z_{\bullet\bullet}})^2 - \left(\frac{1}{k(n-1)}\right)\sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z_{i\bullet}})^2\right] \\
\widehat{\gamma_\epsilon}^{Ab} &= \left(\frac{1}{k(n-1)}\right)\sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z_{i\bullet}})^2,
\end{aligned} \tag{4.1}$$

with $n = n_i$, $\bar{z_{i\bullet}} = \frac{1}{n}\sum_{j=1}^{n} z_{ij}$ and $\bar{z_{\bullet\bullet}} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}$. The variance of the ANOVA estimators $\widehat{\gamma_\alpha}^{Ab}$ and $\widehat{\gamma_\epsilon}^{Ab}$ are respectively given as

$$\begin{aligned}
\Sigma\left(\widehat{\gamma_\alpha}^{Ab}\right) &= \frac{2\gamma_1^2}{k-1} + \frac{4\gamma_1\gamma_2}{n(k-1)} + \frac{2k(n-1)\gamma_2^2}{kn^2(n-1)(k-1)} \text{ and} \\
\Sigma\left(\widehat{\gamma_\epsilon}^{Ab}\right) &= \frac{2\gamma_\epsilon^2}{k(n-1)}.
\end{aligned} \tag{4.2}$$

When discussing Sub-D for such a design, we found that M has only two eigenvalues: $\theta_1 = n$ with multiplicity $g_1 = k - 1$, and $\theta_2 = 0$ with multiplicity $g_2 = N - k = k(n-1)$.

In this case the respective Sub-D estimators for variance components γ_α and γ_ϵ, become:

$$\widehat{\gamma_\alpha}^{Sb} = y^\top \left(\Lambda_{\alpha b}\right) y \text{ and } \widehat{\gamma_\epsilon}^{Sb} = y^\top \left(\Lambda_{\epsilon b}\right) y, \tag{4.3}$$

where

$$\Lambda_{\alpha b} = \frac{A_1^\top A_1}{n(k-1)} - \frac{A_2^\top A_2}{nk(n-1)} \text{ and } \Lambda_{\epsilon b} = \frac{A_2^\top A_2}{k(n-1)}. \tag{4.4}$$

As a consequence of Proposition 2.1 we find that:

$$\begin{aligned} \Sigma\left(\widehat{\gamma_\alpha}^{Sb}\right) &= \frac{2}{k-1}\gamma_\alpha^2 + \frac{4}{n(k-1)}\gamma_\alpha^2\gamma_\epsilon^2 + \frac{2(kn-1)}{kn^2(n-1)(n-1)}\gamma_\epsilon^2 = \Sigma\left(\widehat{\gamma_\alpha}^{Ab}\right), \text{ and} \\ \Sigma\left(\widehat{\gamma_\epsilon}^{Sb}\right) &= \frac{2\gamma_\epsilon^2}{k(n-1)} = \Sigma\left(\widehat{\gamma_\epsilon}^{Ab}\right), \end{aligned} \tag{4.5}$$

and so, consequently, we have the following corollary.

Corollary 4.1. *Let $n_i = n$, $i = 1, \ldots, k$. Then:*

(a) $MSE\left(\widehat{\gamma_\epsilon}^S\right) = MSE\left(\widehat{\gamma_\epsilon}^A\right)$;
(b) $MSE\left(\widehat{\gamma_\alpha}^S\right) = MSE\left(\widehat{\gamma_\alpha}^A\right)$.

For simulation purpose, we toke the same combinations of the parameters γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, and simulated $s = 10000$ repeated designs, using $\beta \sim \mathcal{N}(0,\ \gamma_\alpha)$ and $e \sim \mathcal{N}(0,\ \gamma_\epsilon)$ and $k = 10$ and $n = 23$. For each simulated design, both estimators are applied and the parameters γ_α and γ_ϵ was estimated. Then the average of the estimated values for the parameters was computed as well as the standard deviations of the respective estimated values. The number of groups of treatments and number of observations for each groups was respectively chosen as $k = 10$ and $n = 23$. These values was chosen with no reason other then the simulation purpose. As shown through theoretical results the estimates for both estimators will be equal no matter the number of groups and number of observations for each groups are taken. The results are in Table 4 and an R function to simulate both estimators in Table 5. As we may see, independently of the configuration for the parameters γ_α and γ_ϵ, the estimates and the respective standard deviations found are the same for both estimators.

Table 4. Simulations for different values of γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, with $k = 10$, $n = 23$ and $s = 10000$. **Actual value** denotes the actual values of the parameters; **Estimate** denotes the estimated values of the parameters; **Stand. Dev.** denotes the standard deviations of the estimated values.

Sub-D	γ_α	γ_ϵ	**ANOVA**	γ_α	γ_ϵ
Actual value	**0.5**	**1**	**AV**	**0.5**	**1**
Estimate	0.49668	0.99980	**Estimate**	0.49668	0.99980
Stand. Dev.	0.25322	0.09545	**Stand. Dev.**	0.25322	0.09545
Actual value	**1**	**0.5**	**AV**	**1**	**0.5**
Estimate	0.99885	0.50044	**Estimate**	0.99885	0.50044
Stand. Dev.	0.47794	0.04757	**Stand. Dev.**	0.47794	0.04757
Actual value	**0.75**	**0.5**	**AV**	**0.75**	**0.5**
Estimate	0.75206	0.49970	**Estimate**	0.75206	0.49970
Stand. Dev.	0.36972	0.04706	**Stand. Dev.**	0.36972	0.04706
Actual value	**0.5**	**0.75**	**AV**	**0.5**	**0.75**
Estimate	0.50078	0.74935	**Estimate**	0.50078	0.74935
Stand. Dev.	0.25221	0.07101	**Stand. Dev.**	0.25221	0.07101
Actual value	**0.5**	**0.1**.	**AV**	**0.5**	**0.1**
Estimate	0.50214	0.09996	**Estimate**	0.50214	0.09996
Stand. Dev.	0.23539	0.00954	**Stand. Dev.**	0.23539	0.00954
Actual value	**0.1**	**0.5**.	**AV**	**0.1**	**0.5**
Estimate	0.10025	0.49985	**Estimate**	0.10025	0.49985
Stand. Dev.	0.05769	0.04768	**Stand. Dev.**	0.05769	0.04768

Table 5. The R Code applied to Simulate and test Sub-D and ANOVA-based estimators in random "one-way" balanced designs. Table 4 shows an example.

5. Discussion

As we may see in Silva et al. [2], Silva [12], Silva [11] and Silva et al.[13], through simulations studies, Sub-D has proven its value. When compared to Anova-based estimator it was shown that Sub-D produces estimates with unbiased and lower standard deviations, except in case of random "one-way" balanced designs. In this sense we tough convenient to investigate the performance of both estimators in such a designs; and we found that not only they have the same performance in random "one-way" balanced designs but also in random "one-way" designs with two groups of treatments. In fact this was proven through theoretical results (see Proposition 3 and Corollary 4.1), corroborated with simulations studies (see tables 1, 2, and 3, regarding the random "one-way" designs with two groups of treatments, and tables 4 and 5, regarding the random "one-way" balanced designs.)

Acknowledgements

References

[1] Demidenko, E., Mixed Models: Theory and Applications (Wiley Series in Probability and Statistics), Wiley-Interscience, 2004.

[2] Ferreira, D., Ferreira, S., Nunes, C., Fonseca, M., Silva, A. and Mexia, J., *Estimation and incommutativity in mixed models*, Journal of Multivariate Analysis **161**, 58-67, 2017.

[3] Hardeo, S. and Ojeda, M., *Analysis of Variance for Random Models. Unbalanced Data, Theory, Methods, Applications, and Data Analysis*, 2003.

[4] Khuri, A., *Design for variance components estimation: past and present*, International Statistical Review, Vol. 68 (3), 311–322, 2000.

[5] Khuri, A., Mathew, T. and Sinha, B., 'Statistical tests for mixed linear models', *John Wiley & Sons, Inc., New York*, 1998.

[6] Mulder, M. B., Caro, T. M., Chrisholm, J. S., Dumont, J. P., Hall, R. L., Hinde, R. A., and Ohtsuka, R., *The use of quantitative observational techniques in anthropology [and comments and replies*, Current Anthropology, **26** (3), 323-335, 1985.

[7] Pinheiro, J. and Bates, D., *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing Book Series (SCO), 2000.

[8] Searle, S. R., *Topics in variance components estimation*, Biometrics, **27**, 1-76, 1971.

[9] Searle, S .R., *Linear Models*, Wiley, New York, 1971.

[10] Searle, S. R., Casella, G. and McCulloch, C. E., *Variance Components*, Wiley Series in Probabliity and Statistics, 1992.

[11] Silva, A., *Variance components estimation in mixed linear models*, Ph.D Thesis, New University of Lisbon, 2017.

[12] Silva, A. and Fonseca, M., *The method sub-d for variance components estimation in random one-way designs*, International Journal of Multivariate Data Analysis, **1** (4), 327-336, 2018.

[13] Silva, A., Moreira, A. and Fonseca, M., *Inference in Mixed Linear Models with four Variance Components - Sub-D and Sub-DI*, Int. J. of Data Analysis Techniques and Strategies, **12** (4), 2020.

[14] Singh, B., *On effect of unbalancedness and heteroscedasticity on the ANOVA estimator of group variance component in one-way random model*, Biometrical J., **34**, 91-96, 1992.

[15] Wallace, D. D., *An approximate f statistic for testing population effects in longitudinal studies via mixed models*, Ph.D Thesis, University of North Carolina at Chapel Hill, 1993.