CLUSTERING MACROECONOMIC IMPACT OF COVID-19 IN OECD COUNTRIES AND CHINA

COVID-19'un OECD Ülkeleri ve Çin'de Makroekonomik Etkisinin Kümeleme Analizi

Bige KÜÇÜKEFE*

Abstract

Keywords: Clustering, COVID-19, K-means Clustering, Regression Analysis

JEL Codes: C38, F32, I10, O47 The coronavirus pandemic (COVID-19) has caused the biggest economic contraction in global economy since the Second World War. COVID-19 pandemic has forced governments to take unprecedented measures to prevent the spread and to protect their economies that presented a dilemma because of their conflicting outcomes. This paper investigates the presumption of health-economy trade-off due to COVID-19 by comparing the GDP declines and deaths in per million population in OECD countries and China. The empiric data shows the countries with the highest death rates have seen the largest economic downturns. The clustering analysis by using k-means algorithm finds that there are three partitions of countries for current account balances, GDP growth rate, and deaths in per million population. The countries with current account surpluses above 2.5% of GDP managed to limit their GDP decline below -15% and are in the same cluster. On the other hand, the countries with higher death rates and current account deficits group another cluster and saw GDP declines as above 15% except for USA and Brasil.

Özet

Anahtar Kelimeler: Kümeleme, COVID-19, K-ortalama Kümelemesi, Regresyon Analizi

JEL Kodları: C38, F32, I10, O47 Koronavirüs salgını (COVID-19), İkinci Dünya Savaşı'ndan bu yana küresel ekonomideki en büyük ekonomik daralmaya neden oldu. COVID-19 pandemisi, hükümetleri hem bu hastalığın yayılmasını önlemek hem de ekonomilerini korumaya çalışmak gibi birbiri ile çelişki içinde görünen amaçlar için benzeri görülmemiş önlemler almaya zorladı. Bu makale, OECD ülkeleri ve Cin'de GSYİH düsüslerini ve milyon kişi başına düsen ölümleri karsılastırarak COVID-19 nedenivle bir sağlık-ekonomi değis tokusu olup olmadığı varsayımını araştırmaktadır. Ampirik veriler, en yüksek ölüm oranlarına sahip ülkelerin en büyük ekonomik gerilemeleri yaşadığını göstermektedir. K-ortalamalar algoritması kullanılarak yapılan kümeleme analizi, cari hesap dengesi, GSYİH büyümesi ve bir milyon kişi başına düşen ölüm sayısı açısından ülkelerin üç bölüme ayrıldığını bulmuştur. Cari hesap fazlası GSYH'nin %2,5'inin üzerinde olan ülkeler, GSYİH düşüşlerini % -15'in altında sınırlamayı başardılar ve aynı kümede yer almaktadırlar. Öte yandan, ölüm oranları ve cari açıkları yüksek olan ülkeler başka bir kümede ver alırlar ve bu ülkelerin GSYİH, ABD ve Brezilya dışında, %15'in üzerinde düşmüştür.

Makale Geliş Tarihi (Received Date): 15.10.2020

Makale Kabul Tarihi (Accepted Date): 22.12.2020

^{*} Lecturer, Tekirdag Namik Kemal University, M. Ereglisi Vocational School, Turkey, bkucukefe@nku.edu.tr, ORCID: 0000-0003-1945-3037

1. Introduction

Coronavirus pandemic (COVID-19) has infected millions of people and killed over a million worldwide since January 2020 (John Hopkins Coronavirus Resource Center [JHCRC], 2020) as of early October 2020. The spread of COVID-19 started in Wuhan city in China and has gradually reached other countries across the world. The COVID-19 epicenters shifted throughout the year starting with China and followed by Europe, USA, Brazil, and India. At the time of writing, the USA has the most infections globally with one quarter of confirmed cases (Statista, 2020).

The disruption caused by COVID-19 is not limited to individual health due to its ability to fast spread and kill in huge numbers without a vaccine. Lockdowns and other measures implemented by countries at various degrees have dramatically changed how people live, work, and interact with each other. Social, political, psychological, and economic impacts of COVID-19 have been studied by researchers. Brodeur, Gray, Islam, and Bhuiyan (2020) provided a comprehensive review of this growing literature. Fernandes (2020) estimated the potential global economic costs of COVID-19 and forecasted a global recession whose severity will depend on the success of spread prevention measures, government policies to alleviate liquidity problems, supporting families under financial distress, and securing jobs. Tisdell (2020) provided a selective history of pandemics and discussed moral and ethical questions about how human life should be valued. In a similar work, Acemoglu, Chernozhukov, Werning, and Whinson (2020) developed a multi-group version of the epidemiological SIR population-based model to identify benefits from targeted lockdowns. Serafini et al. (2020) reviewed the studies about the psychological impact of lockdown restrictions due to COVID-19 and identified risk factors. Anderson, Heesterbeek, Klinkenberg and Hollingsworth (2020) argues that, governments can not be able to minimize both deaths from coronavirus disease and economic impact of the pandemic. Keeping death rate as low as possible is the highest priority; so that governments would improve the inevitable economic recession.

COVID-19 pandemic crashed all countries, but there is a huge gap between economic performance of the countries. The Organisation for Economic Co-Operation and Development (OECD, 2020) predicts at the end of 2020, America's economy will be the same size as it was in 2019 but China's will be 10% larger. One of the most important factors about differences among countries spread of the disease (Chaudhry, Dranitsaris, Mubashir, Bartoszko and Riazi, 2020). The first wave of COVID-19 outbreak in China lasted around 20 days under strict lockdown policies and recovery also fast (Sun, Zhang, Yang, Wan and Wang, 2020). The UK extended restrictions on the movement of people on the contrary, Sweden has supported herd immunity', which has particularly fewer restrictions on the population, but provides specific guidance to protect the most vulnerable (Danielli, Patria, Donnelly, Ashrafian and Darzi, 2020). Atkeson (2020), concludes in his study, economic tradeoffs between public health and economy researches are urgently needed. Another difference is the structure of the economies before the pandemic. Manufacturing is easier to operate under social distancing. But service sector hits more severely because this sector rely on face-to-face contact (Seetharaman, 2020). The third important factor is the policy response. For example, The US Federal government passed the stimulus package called CARES Act on 27 March 2020 which covers over \$2 trillion in allocated funds (Coibion, Gorodnichenko and Weber, 2020). Policy also includes fiscal, monetary and exchange rate responses. Transfers to household and business, extension of social safety benefits, and healthcare system funds are typical fiscal policies (Elgin, Basbug and Yalaman, 2020). Monetary policies are liquidity support to banks (International Monetary Fund [IMF], 2020). Imtyaz, Haleem and Javaid (2020) aimed to investigate different governments' responses to the pandemic to find best method to fight Coronavirus. For this purpose, this study used some exploratory data analysis and k-Means clustering. According to research government response like lockdowns and social distancing norms can slow the spread of the COVID-19. Aydin and Yurdakul (2020) analyzed the countries efficiency performance against COVID-19 pandemic, used k-means, hierarchic clustering methods and, the weighted stochastic imprecise data envelopment analysis to assess the performances of 142 countries against COVID-19 outbreak. In this study optimum number of clusters for 142 countries is three.

COVID-19 poses a difficult dilemma that forces the governments to choose between lockdowns to prevent spread of COVID-19 that will potentially harm economy and allowing social and economic activities that will cause fast spread of COVID-19. This research investigated how OECD countries and China performed economically under a health-economy trade-off dilemma by comparing GDP growth rate, current accounts, and deaths per million population from COVID-19. Moreover, clustering analysis using these data with k-means algorithm is performed to group OECD countries and China based on economic performance during COVID-19 pandemic. To analyze k-means clustering efficiency, average silhouette width is calculated.

This paper is organized as follows. Section 2 presents data and methodology. Results are discussed in Section 3 and Section 4 concludes.

2. Data and Methodology

The dataset includes the GDP growth rate in the first half of 2020 and current accounts to GDP ratio in 2019 for 41 OECD countries and China. I use macroeconomic data available at OECD.stat. First half GDP growth rate data was computed by using the quarterly data which is based on previous period. Deaths per million population from COVID-19 data was obtained from Statista for the period from 01.01.2020 to 27.09.2020. Ethics of research and publication were followed in this study, which does not require permission from the ethics committee and / or legal / special permission.

Fig.1 shows confirmed deaths per million people as of 27.09.2020 and GDP decline in percent for OECD countries and China. I build a simple regression model to identify the correlation between deaths from COVID-19 and GDP growth rate. The model outcome is GDP growth rate and independent variable is mortality data per million population.

Based on work by Lloyd (1982), the k-means clustering method uses a local search approach to group the data points into k clusters. The k-means method has been applied to many areas from machine learning (unsupervised learning) to computer graphics (Arthur and Vassilvitskii, 2006). Let $X = \{x_1, x_2, ..., x_n\}$ be a set of data points in \mathbb{R}^d . After being seeded with a set of k centers $c_1, c_2, ..., c_k$ in \mathbb{R}^d , the k-means algorithm partitions these data points into k clusters. Optimal number of clusters was determined by the elbow method which plots the ratio of variance outcome as a function of number of clusters (Naeem and Wumaier, 2018). Three-dimensional dataset has different scales that prevent equal contribution. Therefore, a standardization procedure was applied before the k-means clustering.



Regression analysis and k-means clustering were made using the R software for statistical computing (R Core Team, 2018).

Figure 1. Confirmed Deaths Per Million People From COVID-19 Vs GDP Growth Rate in OECD Countries and China

3. Results

3.1. Health-Economy Trade-off

It is evident from Fig.1 that the countries with higher death rates from COVID-19 experienced a higher economic downturn in contrast with the conception that countries faced a trade-off between health and economy. Countries like Great Britain and Spain saw GDP contractions more than 20% with death rates as high as 700 per day. With noticeable exception for GDP growth rate, China did not record an economic decline due to correction in the second quarter of 2020. The highest drop in GDP happened in India with -25 % even though the country had relatively lower death rates. The highest mortality rate occurred in Belgium not only among OECD countries and China but in the world (World Health Organization [WHO], 2020) whereas the country saw a 15% decline in GDP. Turkey and Germany recorded similar death rates and economic contraction. Another pair of countries with similar records are Italy and Mexico.

The regression model in this study uses the confirmed deaths per million population from COVID-19 as an independent variable and GDP growth rate as an outcome.



Figure 2. Regression Model for Confirmed Deaths Per Million People From COVID-19 vs GDP Growth Rate

The slope coefficient in the linear model has a value of -0.002991 with p-value = 0.00257 which is sufficient to reject the null hypothesis that there is no relation between confirmed deaths in per million population from COVID-19 and GDP growth rate. The negative slope coefficient indicates a negative correlation between two variables. The adjusted coefficient of determination (Adjusted R-squared) is 0.1856. Relatively small value of adjusted R-squared indicates high variance in data. Fig.3 shows that the residuals are randomly distributed around the regression line and the linear model is valid. This finding voids the presumption that there is a trade-off between health and economy. The countries experiencing larger GDP decline also had higher death rates in general.





3.2. K-means Clustering

In this part, it is investigated the impact of current account balances on the GDP growth rate by clustering the countries according to deaths per million population from COVID-19, GDP growth rate and current account to GDP ratio. K-means clustering is one of the most popular algorithms to partition data into clusters and is commonly used in unsupervised machine learning.

Determining the optimal number of clusters (k) in a k-means clustering problem is important. One useful method for obtaining the optimal value of k is the elbow method which plots various values of cost with changing k. The scree plot from elbow method is provided in Fig.4. The elbow at 3 suggests the optimal value for k is 3.



Figure 4. Scree Plot for the K-means Cluster Analysis (k=3)

After determining the optimal number of clusters as 3 for three-dimensional dataset, the k-means clustering algorithm generated three partitions as shown in Fig. 5.

The silhouette coefficient (silhouette width) is used to evaluate validation and goodness of clustering. The silhouette coefficient is calculated in three steps: First average dissimilarity is calculated as follows. For each observation i, the average dissimilarity between i and all other points within the same cluster with i is calculated and called "D(i)". Second dissimilarity between i and the closest cluster to i right after i's own cluster is calculated and called "C(i)".

At last, the silhouette (S(i)) width is the difference between C(i) and D(i) (C(i) - D(i)) divided by the greatest of those two values (max(D(i), C(i))).

$$S(i) = (C(i) - D(i)) / \max\{D(i), C(i)\}$$
(1)

The observation is well clustered If S(i) > 0. On the other hand, a negative C(i) indicates poor clustering. S(i) = 0 means the observation is between two clusters.

In this research the silhouette coefficients of three clusters are as follows:

Cluster Type	Size	Average Silhouette Width		
Cluster B	15	0.28		
Cluster A	14	0.32		
Cluster C	13	0.33		
Cluster C	15			

Table 1. Cluster Type, Size and Average Silhouette Width

Source: Own elaboration

In this research each cluster's average silhouette width is positive and near 0.30. This average silhouette width has statistical significance.

Clusters silhouette plot with average silhouette width are provided in figure 5. In this figure cluster 1 with biggest size represents cluster B, cluster 2 represents cluster A and cluster 3 represents cluster C.



Figure 5. Clusters Silhouette Plot

The graphical representation of clusters and countries in the clusters are provided in Figure 6. Notably, countries with current account surplus (black font color) over 2.5 % differ

from the countries with current account deficit and saw lower GDP contractions between 5 - 15 %. The second cluster (green font color) includes countries that had current account balances lower than 2.5 % and experienced economic downturns like the first cluster. China is the only exception with positive GDP growth rate. Turkey is in the second cluster and borders with the first one with Estonia. The countries in the third cluster (red font color) were severely affected both in terms of GDP contraction and deaths per million population from COVID-19 pandemic. Except Italy and Spain, these countries had current account deficits up to 5 %. USA and Brazil are located within the second cluster countries on Fig.5 due to their lower GDP decline during the period. It is evident from Fig. 5 that the countries that managed the COVID-19 better (i.e. lower death rates) were able to limit GDP contraction even though they had current account deficit.



Figure 6. K-means Clustering of OECD Countries and China (k=3) for GDP Growth Rate in 2020 First Half, Current Account Balance to GDP Ratio (2019), and Deaths Per One Million Population as of 27.09.2020

K-means Clustering of OECD countries and China is shown in Table 1 with current account balance, GDP growth rate and COVID-19 death rates. Countries' clusters can be seen clearly in this table. As in Table 2, Cluster A consists of 14, cluster B consists of 15, cluster C consists of 13 countries.

Ekonomi, Politika & Finans Araştırmaları Dergisi, 2020, 5(Özel Sayı): 280-291 Journal of Research in Economics, Politics & Finance, 2020, 5(Special Issue): 280-291

	Current	GDP	COVID-19	Clustor	Cluster	Cluster C
COUNTRIES	Balance/GDP (%) (2019)	Growth (%) 2020-I	Death rates (per million)	A	B	
Argentina	-0.76986	-19.7051	343.904			ARG
Australia	0.51136	-7.24096	34.118		AUS	
Austria	2.624642	-12.8731	87.382	AUT		
Belgium	-1.22964	-15.1579	860.598			BEL
Bulgaria	4.036337	-9.74943	113.551	BGR		
Brazil	-2.68795	-11.9099	665.254			BRA
Canada	-1.96919	-13.3768	245.402		CAN	
Switzerland	12.25565	-10.4784	205.439	CHE		
China	0.985404	0.35	3.293		CHN	
Colombia	-4.26185	-16.7033	497.142			COL
Czechia	-0.35153	-11.7477	55.187		CZE	
Germany	7.146996	-11.514	112.874	DEU		
Denmark	7.923734	-8.7966	111.875	DNK		
Spain	1.98864	-22.1075	667.996			ESP
Estonia	2.19711	-7.69619	48.246		EST	
Finland	-0.76665	-6.28496	61.905		FIN	
France	-0.68306	-18.8724	485.649			FRA
United Kingdom	-3.78074	-22.1163	618.257			GBR
Greece	-1.39509	-14.6396	36.074		GRC	
Hungary	-0.77483	-14.8619	76.188		HUN	
Indonesia	-2.71507	-7.58885	37.686		IDN	
India	-0.93541	-24.7555	68.48			IND
Ireland	-9.35794	-7.99276	364.94		IRL	
Iceland	5.80568	-14.2764	29.304	ISL		
Israel	3.556327	-9.81122	166.483	ISR		
Italy	2.944803	-17.5529	592.407			ITA
Japan	3.631405	-8.45543	12.216	JPN		
Lithuania	4.239169	-5.77446	32.693	LTU		
Luxemburg	4.353185	-8.55889	198.091	LUX		
Latvia	-0.51617	-9.25697	19.086		LVA	
Mexico	-0.1942	-18.0185	591.339			MEX
Netherlands	10.20875	-9.86179	370.998	NLD		
Norway	3.983146	-6.72749	49.804	NOR		
New Zealand	-2.88302	-13.427	5.184		NZL	
Poland	0.468789	-9.20326	64.048		POL	
Portugal	-0.09929	-17.101	190.65			PRT
Romania	-4.55216	-12.0004	243.636		ROU	
Slovenia	6.55939	-13.9571	65.418	SVN		
Sweden	3.915164	-8.12515	582.22	SWE		
Turkev	1.152023	-11.1325	94.013		TUR	
USA	-2.33153	-10.2409	617.811			USA
South Africa	-3.03525	-16.7367	276.115			ZAF

 Table 2. OECD Countries' and China's Current Account Balance, GDP Growth, COVID-19 Death

 Rates and Clusters

Source: OECD, Statista and own elaboration

As reported in Table 3, big differences occur between the statistics parameters of the clusters. Considering the total number of cases, the first cluster has 152 thousand on average, while the second cluster has 92 thousand and the third cluster has 498 thousand. Similar situations are valid for other parameters.

	Cluster A (14 countries)		Cluster B (15 countries)		Cluster C (13 countries)				
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Current Account Balance/GDP (%) (2019)	2.62464	12.2556	5.73139	-9.35793	2.1971	-1.3311	-4.26185	2.94480	-1.15964
GDP Growth (%) 2020-I	-14.2764	-5.77446	-9.92567	-14.8618	0.35	-9.7399	-24.7555	-10.2408	-17.7675
COVID-19 Death rates (per million)	12.216	582.22	152.7391	3.293	364.94	92.6004	68.48	860.598	498.1232

Table 3. Statistic Analyses of the Clusters

Source: Own elaboration

4. Conclusion

The coronavirus pandemic (COVID-19) has caused the worst economic recession in global economy since the Second World War. This study investigates the presumption that there is a trade-off between health and economy during COVID-19 pandemic. The linear regression model for deaths per million population from COVID-19 and GDP decline in OECD countries and China reveals the negative correlation between two. As a result, this study argues that there is no empirical evidence indicating the health-economy trade-off in OECD countries and China. This study finds that most countries experiencing economic downturns also were severely affected by the pandemic in terms of deaths per million population. There is no doubt that many factors affect the COVID-19 death rate and economic performance of a country.

The k-means clustering algorithm identified three groups of OECD countries and China for three-dimensional data of deaths per million population from COVID-19, GDP contraction, and current account balance. The efficiency test is performed by silhouette width with. Result of average silhouette width is 0.31 indicates well k-means clustering. The main finding of this study is that the countries with current account surplus managed to limit economic downturn and death rates due to COVID-19.

References

- Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M.D. (2020). Optimal targeted lockdowns in a multi-group SIR model (NBER Working Paper No. w27102). Retrieved from: https://www.nber.org/w27102
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D. and Hollingsworth, T. (2020). How will countrybased mitigation measures influence the course of the covid-19 epidemic? *The Lancet*, 395(10228), 931-934. https://doi.org/10.1016/S0140-6736(20)30567-5
- Arthur, D. and Vassilvitskii, S. (2006, June). How slow is the K-means method?. In N Amenta and O Cheong (Eds.), *Proceedings of the 22. Annual Symposium on Computational Geometry* (pp.144-153). Papers presented at 22nd ACM Symposium on Computational Geometry (SoCG), Arizona. New York: Association for Computing Machinery.
- Atkeson, A. (2020). What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios (NBER Working Paper No. w26867). Retrieved from https://www.nber.org/w26867
- Aydin, N. and Yurdakul, G. (2020). Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms. *Applied Soft Computing*, 97, 106792. https://doi.org/10.1016/j.asoc.2020.106792
- Brodeur, A., Gray, D., Islam, A. and Bhuiyan, S. (2020), *A literature review of the economics of Covid-19* (IZA Discussion Paper, No. 13411). Retrieved from https://www.econstor.eu/bitstream/10419/222316/1/GLO-DP-0601.pdf
- Chaudhry, R., Dranitsaris, G., Mubashir, T., Bartoszko J. and Riazi, S. (2020). A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortalityand related health outcomes. *EClinical Medicine*, 25, 100464. doi:10.1016/j.eclinm.2020.100464
- Coibion, O., Gorodnichenko, Y. and Weber, M. (2020). How US consumers use their stimulus payments. *VoxEU & CEPR Coverage of the Covid-19 Global Pandemic*. Retrieved from https://european.economicblogs.org/voxeu/2020/gorodnichenko-weber-us-consumers-stimuluspayments
- Danielli, S., Patria, R., Donnelly, P., Ashrafian, H. and Darzi, A. (2020). Economic interventions to ameliorate the impact of COVID-19 on the economy and health: an international comparison. *Journal of Public Health*, 104. https://doi.org/10.1093/pubmed/fdaa10
- Elgin, C., Basbug, G. and Yalaman, A. (2020). Economic policy responses to a pandemic: developing the Covid-19 economic stimulus index. *Covid Economics*, 1(3), 40-53. Retrieved from http://www.ceyhunelgin.com/
- Fernandes, N. (2020). Economic effects of Coronavirus outbreak (COVID-19) on the world economy (IESE Business School Working Paper No. WP-1240-E). Retrieved from https://ssrn.com/abstract=3557504
- Imtyaz, A., Haleem, A. and Javaid, M. (2020). Analysing governmental response to the COVID-19 pandemic. *Journal of Oral Biology and Craniofacial Research*, 10(4), 504–513. https://doi.org/10.1016/j.jobcr.2020.08.005
- International Monetary Fund. (2020). *Policy-responses-to-COVID-19*. Retrieved from https://www.imf.org/en/Topic s/imf-and-covid19/Policy-Responses-to-COVID-19#U
- John Hopkins Coronavirus Resource Center. (2020). John Hopkins University & Medicine [Dataset]. Retrieved from https://coronavirus.jhu.edu.
- Lloyd, S. P. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2), 129-136. doi:10.1109/TIT.1982.1056489
- Naeem S. and Wumaier, A. (2018). Study and implementing K-mean clustering algorithm of English text and techniques to fine the optimal value of K. *Industrial Journal of Computer Applications*, *182*, 31, 7-14. Retrieved from https://www.ijcaonline.org/

- Organisation for Economic Co-operation and Development. (2020). *OECD economic outlook*. Retrieved from https://www.oecd-ilibrary.org/
- R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org
- Seetharaman, P. (2020). Business models shifts: Impact of Covid-19. International Journal of Information Management, 54, 102173. https://doi.org/10.1016/j.ijinfomgt.2020.102173
- Serafini, G., Parmigiani, B., Amerio A., Aguglia, A., Sher L. and Amore M. (2020). The psychological impact of COVID-19 on the mental health in the general population. *QJM: An International Journal of Medicine*, 113(8), 531–37. https://doi.org/10.1093/qjmed/hcaa201.
- Statista. (2020). Number of cumulative cases of coronavirus (COVID-19) worldwide from January 8 to October 7, 2020, by day [Dataset]. Retrieved from https://www.statista.com/statistics/1103040/cumulative-coronavirus-covid19-cases-numberworldwide-by-day.
- Sun, Z., Zhang, H., Yang, Y., Wan, H. and Wang, Y. (2020). Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China. *Science of The Total Environment*. 746, 141347. https://doi.org/10.1016/j.scitotenv.2020.141347,
- Tisdell, C. A. (2020). Economic, social and political issues raised by the COVID-19 pandemic. *Economic* Analysis and Policy 68, 17–28. https://doi.org/10.1016/j.eap.2020.08.002
- World Health Organization. (2020). World health organization data [Dataset]. Retrieved from https://who.int/data