



Clusters in the Spread of the COVID-19 Pandemic: Evidence From the G20 Countries

Tian Meng*

Shanghai Sci-Tech Finance Institute, Shanghai University, Shanghai, China

This study tests the validity of the club convergence clustering hypothesis in the G20 countries using four measures of the spread of the COVID-19 pandemic: total number of confirmed cases per million people, new cases per million people, total deaths per million people, and new deaths per million people. The empirical analysis is based on the daily data from March 1, 2020, to October 10, 2020. The results indicate three clusters for the per capita income, two clusters for total cases per million people, and new cases per million people. Besides, there are only one and two clusters for total deaths per million people and new deaths per million people. Potential policy implications are also discussed in detail.

Keywords: the COVID-19 pandemic, the COVID-19 cases, the COVID-19 deaths, convergence clustering test procedure, the G20

OPEN ACCESS

Edited by:

Chi Lau,
University of Huddersfield,
United Kingdom

Reviewed by:

Haiping Li,
Beijing Institute of Petrochemical
Technology, China
Zezeng Li,
University of Huddersfield,
United Kingdom

*Correspondence:

Tian Meng
mengtian69@yahoo.com

Specialty section:

This article was submitted to
Health Economics,
a section of the journal
Frontiers in Public Health

Received: 12 November 2020

Accepted: 27 November 2020

Published: 18 January 2021

Citation:

Meng T (2021) Clusters in the Spread of the COVID-19 Pandemic: Evidence From the G20 Countries. *Front. Public Health* 8:628789. doi: 10.3389/fpubh.2020.628789

INTRODUCTION

In this paper, we examine the validity of the club convergence clustering hypothesis in the G20 countries using four indicators of the spread of the COVID-19 pandemic: total number of confirmed cases per million people, new confirmed cases per million people, total deaths per million people, and new deaths per million people. It is essential to examine the validity of the convergence clustering hypothesis in the G20 countries related to the indicators of the spread of the COVID-19 pandemic. Indeed, whether there are significant clusters in the spread of the COVID-19 pandemic can be particularly important for policy implications, such as lockdowns and limitations on business and social life. COVID-19 pandemic significantly affects every aspect of the global economy (1, 2). Therefore, forecasting the COVID-19 pattern in different countries is significant.

There are previous papers to analyze the spread of the pattern of the COVID-19 pandemic. For example, Katul et al. (3) show a significant global convergence in the generic spread mechanisms of the COVID-19. However, the authors focus on the data until early 2020. Kuniya (4) also examines the impact of an emergency state for the first wave of the COVID-19 in Japan for the period from April 7 to May 25, 2020. The author finds that the state of emergency has provided to 80% decline in the contact rate. Therefore, there is a significant convergence in the spread of the COVID-19 pandemic in Japan during concern. Chimmula and Zhang (5) forecast the infectious diseases related to the COVID-19 outbreak in Canada. The authors show that the spread of the COVID-19 pandemic in Canada follows a stationary forecasting process. Shabani and Shahnazi (6) considered the data of the COVID-19 cases from February 9, 2020, to July 27, 2020, to analyze COVID-19's spatial distribution dynamics. For this purpose, the authors applied the Markov Chain, while also used the Spatial Markov Chain. The findings indicate that the COVID-19 in 40 Asian countries have a unit root characteristics with the domestic policies. Besides, the neighboring countries have significant effects on the spread of COVID-19. Ismail et al. (7) confirm the evidence of convergence for the indicators on the spread of COVID-19 in 187 countries.

This study follows the current developments in the literature. It aims to examine the validity of the club convergence clustering hypothesis in the G20 countries using four indicators of the spread of the COVID-19 pandemic: total cases per million people, new cases per million people, total deaths per million people, and new deaths per million people. We use the daily data from March 1, 2020, to October 10, 2020.

A thorough search of the relevant literature yielded only one related article. This is the first study to use the club convergence clustering method to examine the spread of the COVID-19 pandemic in different countries. The results indicate two clusters for the per capita income, three clusters for total cases per million people, and new cases per million people. Besides, there are only one and two clusters for total deaths per million people and new deaths per million people. These findings suggest some substantial implications in the G20 countries. For example, the policymakers in these should implement measures for controlling the spread of the COVID-19 pandemic, and some countries have different dynamics in the spread of the COVID-19 pandemic. This main evidence should be some significant policy implications for these countries since the risks related to the COVID-19 significantly greater in some countries than others. Furthermore, emerging countries are seemed to be heavily affected by the COVID-19 pandemic.

The remaining parts of the study are structured as follows: Section Data and Club Convergence Methodology provides the details of the data and the club convergence methodology. The empirical results are stated in Empirical Findings. Section Conclusion concludes the study with possible implications of the findings.

DATA AND CLUB CONVERGENCE METHODOLOGY

Data

We examine possible cluster and club convergence dynamics for four indicators of the spread of the COVID-19 pandemic: total cases per million people, new cases per million people, total deaths per million people, and new deaths per million people. The empirical analysis is based on the daily data for the period from March 1, 2020, to October 10, 2020, in the G20 countries (19 countries excluding the European Union): Argentina, Australia, Brazil, Canada, China PR, France, Germany, India, Indonesia, Italy, Japan, Mexico, the Russian Federation, Saudi Arabia, South Africa, South Korea, Turkey, the United Kingdom, and the United States. The list of countries, including the country id of the countries in the empirical analyses, are provided in **Table 1**. The frequency of the panel data is daily. The data are downloaded from the dataset of Hasell et al. (8), so-called the *Data on COVID-19 (Coronavirus) by Our World in Data* project (<https://github.com/owid/covid-19-data/tree/master/public/data>).

Descriptive statistics of four indicators of the spread of the COVID-19 pandemic: total cases per million people, new cases per million people, total deaths per million people, and new deaths per million people are reported in **Table 2**.

TABLE 1 | Countries in the dataset.

ISO Code	Country	Country ID
ARG	Argentina	1
AUS	Australia	2
BRA	Brazil	3
CAN	Canada	4
CHN	China, PR	5
FRA	France	6
DEU	Germany	7
IND	India	8
IDN	Indonesia	9
ITA	Italy	10
JPN	Japan	11
MEX	Mexico	12
RUS	Russian Federation	13
SAU	Saudi Arabia	14
ZAF	South Africa	15
KOR	South Korea	16
TUR	Turkey	17
GBR	The United Kingdom	18
USA	The United States	19

TABLE 2 | Descriptive statistics.

Indicator	Total Cases per Million People	New Cases per Million People	Total Deaths per Million People	New Deaths per Million People
Mean	2,858	34.83	144.8	1.287
Max.	23,785	380.8	703.9	74.14
Min.	0.002	0.000	0.000	0.000
Std. Dev.	4,135	53.36	199.6	2.588
Obs.	4,256	4,256	4,256	4,256

Club Convergence Methodology

Phillips and Sul (9, 10) propose a novel approach for identifying the stochastic properties of convergence and defining different convergence clubs among the panel units over time. The methodology assumes the time-varying model with nonlinear nature, and it offers a mechanism of nonlinear transition. The best way of this approach is that it can also be applied in the panel data with unit root, or it does not assume homogeneous (common) factors in the data-generating algorithm. Besides, Phillips and Sul (9, 10) club convergence methodology captures each country's heterogeneity within the panel dataset. Hence, the club convergence procedure considers the dynamics of the COVID-19 spread among the G20 countries in a panel dataset. The COVID-19 spread rate in each county can be defined by the panel dataset, which may follow different convergence dynamics. Therefore, the club convergence procedure is a suitable test for the convergence dynamics of the COVID-19 spread among the G20 countries. This paper aims to examine the different convergence club features in the COVID-19 spread among the

G20 countries. We can define the club convergence procedure as such:

The series X_{it} captures an indicator of the COVID-19 spread for country i at time t , and $i = 1, 2, \dots, N$; $t = 1, 2, \dots, 19$. At this stage, Phillips and Sul (9, 10) decompose the variable into two components: First is the common component of cross-sectional dependence in a panel dataset, g_{it} , and transitory component, a_{it} , as such:

$$X_{it} = g_{it} + a_{it} \quad (1)$$

Phillips and Sul (9, 10) define the Equation (1) as the common and the idiosyncratic components. At this stage, the variable follows nonlinear stochastic properties, as such:

$$X_{it} = \left(\frac{g_{it} + a_{it}}{\mu_t} \right) \mu_t = \delta_{it} \mu_t \text{ for all } i \text{ and } t, \quad (2)$$

Where, μ_t captures the common component and δ_{it} indicates the time-varying idiosyncratic component. δ_{it} denotes the relative difference between common trend component μ_t and the value of X_{it} is an indicator of the spread of the COVID-19 in a country i at time t .

Let us take the deaths from the COVID-19 per million people as an example. μ_t denotes a common trend of the COVID-19 per million people in whole 19 countries. δ_{it} captures each country's relative share in terms of the COVID-19 per million people in the common trend in the G20 countries. The baseline approach of club convergence approach of Phillips and Sul (9, 10) is to define the time-varying load δ_{it} , and time-varying load will determine the dynamics of the club convergence in terms of the power of convergence. Furthermore, Phillips and Sul (9, 10) calculate a transition coefficient, which can be defined as h_{it} . Transition coefficient is based on the time-varying factor loadings (δ_{it}), as such:

$$h_{it} = \frac{X_{it}}{\frac{1}{N} \sum_{i=1}^N X_{it}} = \frac{\delta_{it} \mu_t}{\frac{1}{N} \sum_{i=1}^N \delta_{it} \mu_t} = \frac{\delta_{it}}{\frac{1}{N} \sum_{i=1}^N \delta_{it}} \quad (3)$$

In Equation (3), h_{it} indicates a transition term, which measure δ_{it} related to the average of the panel at time t . At this stage, the transition term defines a transition nature for source countries i relative to the average of the panel dataset of the G20 countries. All indicators used the filter provided by Hodrick and Prescott (11) to remove the cyclical component. Following Ravn and Uhlig (12), lambda is defined $1600 \times (365/4)^{4}$ for daily data. The filtered coefficient for transition parameter is represented by \hat{h}_{it} , and an extracted time-trend is defined as \hat{X}_{it} .

Furthermore, the club convergence test procedure also defines the cross-sectional variance ratio, $\frac{H_t}{H_t}$, which can be defined as follows:

$$H_t = \frac{1}{N} \sum_{i=1}^N (\hat{h}_{it} - 1)^2 \quad (4)$$

At this stage, Phillips and Sul (9, 10) show that the transition parameter H_t is defined within a limit form, which can be written as such:

$$H_t \sim \frac{A}{L(t)^2 t^{2\alpha}} \text{ as } t \rightarrow \infty \quad (5)$$

In Equation (5) A is a constant term, and $A > 0$, $L(t)$ is the function of time, and α indicates the speed of convergence. Phillips and Sul (9, 10) define $\log t$ regression to test the validity of the null hypothesis of convergence. The null hypothesis can be written as $H_0: \delta_i = \delta$ and $\alpha \geq 0$ and against $H_1: \delta_i \neq \delta$ for all i or $\alpha < 0$.

Furthermore, Phillips and Sul (9, 10) estimate the following Ordinary Least Squares (OLS) equation:

$$\text{Log} \left(\frac{H_1}{H_t} \right) - 2 \log L(t) = \hat{a} + \hat{b} \log t + \hat{c} (\log t)^2 + \hat{u}_t \quad (6)$$

In Equation (6), $L(t) = \log(t + 1)$, the fitted coefficient of $\log t$ is $\hat{b} = 2\hat{\alpha}$, and $\hat{\alpha}$ is the estimate of α in the null hypothesis. The authors include the *squares of log t* to enhance the test procedure's power by capturing nonlinearity in the series. The test procedure considers the initial condition by removing a fraction of the sample in the estimated regression. The initial condition requires a starting point $t = [rT]$ with $r > 0$. Phillips and Sul (9, 10) set $r = 0.3$. The authors estimate the coefficient of \hat{b} by providing the standard errors in the use of Heteroskedasticity and Autocorrelation Consistent (HAC) of the long-run variance in residuals to perform the one-sided t -test of null $\alpha \geq 0$. Hence the t -test statistic $t_{\hat{b}}$ is based on the normal distribution, and if $t_{\hat{b}} < -1.645$, the null hypothesis of club convergence will be rejected.

Finally, Phillips and Sul (9, 10) discuss that the rejection of the null of club convergence does not mean that there cannot be subgroup convergence in the panel dataset. It is important to note that the club convergence test procedure is defined for detecting cluster units. Using the club convergence test procedure, we examine the club convergence dynamics in the G20 countries over the period under concern. The club convergence is defined as $\log t$ regressions with the following main issues:

- 1) Ordering: Order the X_{it} series following the last observation in the panel dataset.
- 2) Group Formation: Calculate t -statistic $t_{\hat{b}}(k)$ for each country (k) and select country or countries for the core group.
- 3) Membership of the Club: Find the country for membership in the core group by including each remaining country separately, following the results of $\log t$ tests. A new country will be added to the club if the calculated t -statistic is higher than zero.
- 4) Recursion and Stop: Finally, $\log t$ -tests are applied for the group of unselected countries. If the cluster of countries converges in the first club, a second club will be formed. If there is no club convergence, sub-convergent club clusters will be investigated. If no subgroups are defined for the remaining countries, they will be defined as countries with a divergence pattern.

EMPIRICAL FINDINGS

Table 3 provides the club convergence results for four indicators of the spread of the COVID-19 pandemic: total cases per million people, new cases per million people, total deaths per million people, and new deaths per million people.

TABLE 3 | Results of the club convergence tests for indicators of the spread of the COVID-19.

Indicators of the spread of the COVID-19	Clubs	\hat{b}	t-statistic
Total cases per million people	(1)	-0.025	-0.550
	1 3 5 7 8 9 11 12 13 15		
	16 17 19		
New cases per million people	(2)	-0.090	-0.146
	6 14 18		
	(3)		
Total deaths per million people	(1)	0.417	74.4
	2 4 10		
	(1)		
1 2 3 4 5 6 7 8 9 10			
11 12 15 16 17 18 19			
New deaths per million people	(2)	-0.820	-0.919
	13 14		
	(1)		
1 2 3 4 5 6 7 8 9 10			
11 12 13 14 15 16 17 18 19			
New deaths per million people	(1)	0.031	0.191
	1 2 3 5 8 9 11 12 13 14		
	15 16 17 19		
New deaths per million people	(2)	-0.868	-3.689
	4 6 7 10 18		

In terms of the findings of the club convergence test for the total COVID-19 cases per million people, there are three clubs. The log t regression results for the first club consisting of 13 countries with the t -statistic of -0.55 , and the null hypothesis of convergence can be rejected. The second club consists of three countries (France, Saudi Arabia, and the United Kingdom) with the t -statistic -0.146 , and the null hypothesis of convergence can be rejected. Finally, the third club shows three countries (Australia, Canada, and Italy) with the t -statistic 74.4 , and the null hypothesis of convergence cannot be rejected.

There are two clubs in terms of the club convergence test results for the new COVID-19 cases per million people. The log t regression findings for the first club consisting of 17 countries with the t -statistic of -2.011 and the null hypothesis of convergence cannot be rejected. The second club includes two countries (the Russian Federation and Saudi Arabia) with a t -statistic -0.82 , and the null hypothesis of convergence can be rejected.

When we look at the club convergence test findings for the total COVID-19 deaths per million people, only one club consists of all countries in the dataset. The log t regression results for the only club consisting of all countries with the t -statistic of 179.6 and the null hypothesis of convergence cannot be rejected.

There are two clubs in terms of the club convergence test results for new deaths per million people. The log t regression findings for the first club consisting of 14 countries with the t -statistic of 0.191 and the null hypothesis of convergence can be rejected. Furthermore, the second club consists of five countries (Canada, France, Germany, Italy, and the United Kingdom) with a t -statistic -3.689 , and the null hypothesis of convergence cannot be rejected.

CONCLUSION

In this paper, we examined the validity of the club convergence clustering hypothesis in the G20 countries using four indicators of the spread of the COVID-19 pandemic: total cases per million people, new cases per million people, total deaths per million people, and new deaths per million people. We used the daily data from March 1, 2020, to October 10, 2020. We followed the club convergence clustering methodology of Phillips and Sul (9, 10) to model the time-varying nature of the spread of the COVID-19 pandemic and capture different fighting policy pandemic strategies. We observed that the cases and deaths related to the COVID-19 pandemic have a nonlinear nature and converge among the G20 countries.

We observed three clusters for the per capita income and two clusters for total cases per million people and new cases per million people. Besides, there are only one and two clusters for total deaths per million people and new deaths per million people. These results indicate that although policymakers in different countries have different solutions to the total pandemic deaths per million, they have similar stochastic properties in the G20 countries. This evidence can be related to the fact that the treatment of the COVID-19 virus has not been fully provided in the globe and the deaths due to the COVID-19 virus has somehow a random nature. Our results also indicate that if there will be no prevention, the countries with the low-level of COVID-19 spread will converge toward a pandemic's long-run level, which is the United States' case. Different characteristics of the countries have negligible effects on the spread of the COVID-19, particularly when we focus on the club convergence dynamics of the death ratios related to the COVID-19.

In terms of new deaths, Canada, France, Germany, Italy, and the United Kingdom are different countries. The death ratios per million people have decreased in these countries over time, creating a new club for these countries. In terms of other developed and developing G20 countries, there is another club convergence procedure. When we look at the new cases for the COVID-19, only the Russian Federation and Saudi Arabia have a different nature for convergence. Other countries have a similar pattern for the new cases for the COVID-19. The differences between the Russian Federation and Saudi Arabia are related to these countries' leading oil-exporters in the World economy. Note that the oil prices have significantly declined during the COVID-19.

Given that there are autocratic regimes in these countries, they may be underestimating the number of new cases to show the situation better. In terms of total cases for the COVID-19,

there are three different clubs, and they are hard to explain. This issue is the limitation of our study. Future papers can focus on more countries to analyze the club convergence clustering hypothesis's validity in larger panel datasets, which should have more countries and higher time dimensions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/owid/covid-19-data/tree/master/public/data>.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

1. McKibbin WJ, Fernando R. *The Global Macroeconomic Impacts of COVID-19: Seven Scenarios*. Centre for Applied Macroeconomic Analysis (CAMA) Working Paper, No. 19/2020. Canberra: CAMA (2020). doi: 10.2139/ssrn.3547729
2. Ozili P, Arun T. *Spillover of COVID-19: Impact on the Global Economy*. Munich Personal RePEc Archive (MPRA) Paper, No. 99317. Munich: University Library of Munich (2020). doi: 10.2139/ssrn.3562570
3. Katul GG, Mrad A, Bonetti S, Manoli G, Parolari AJ. Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics. *PLoS ONE*. (2020) 15:e0239800. doi: 10.1371/journal.pone.0239800
4. Kuniya T. Evaluation of the effect of the state of emergency for the first wave of COVID-19 in Japan. *Infect Dis Model*. (2020) 5:580–7. doi: 10.1016/j.idm.2020.08.004
5. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada Using LSTM Networks. *Chaos Solit Frac*. (2020) 135:109864. doi: 10.1016/j.chaos.2020.109864
6. Shabani ZD, Shahnazi R. Spatial distribution dynamics and prediction of COVID-19 in Asian Countries: Spatial Markov Chain approach. *Regional Science Policy and Practice* (2020) forthcoming. doi: 10.1111/rsp3.12372
7. Ismail L, Materwala H, Znati T, Turaev S, Khan MA. Tailoring time series models for forecasting coronavirus spread: case studies of 187 countries. *Comp Struct Biotechnol J*. (2020) 18:2972–3206. doi: 10.1016/j.csbj.2020.09.015
8. Hasell J, Mathieu E, Beltekian D, Macdonald B, Giattino C, Ortiz-Ospina E, et al. A cross-country database of COVID-19 testing. *Sci Data*. (2020) 7:345. doi: 10.1038/s41597-020-00688-8
9. Phillips PCB, Sul D. Transition modeling and econometric convergence tests. *Econometrica*. (2007) 75:1771–855. doi: 10.1111/j.1468-0262.2007.00811.x
10. Phillips PCB, Sul D. Economic transition and growth. *J Appl Econ*. (2009) 24:1153–1185. doi: 10.1002/jae.1080
11. Hodrick RJ, Prescott EC. Postwar U.S. business cycles: an empirical investigation. *J Money Credit Bank*. (1997) 29:1–16. doi: 10.2307/2953682
12. Ravn MO, Uhlig H. On adjusting the Hodrick-prescott filter for the frequency of observations. *Rev Econ Stat*. (2002) 84:371–6. doi: 10.1162/003465302317411604

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Meng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.