Do Different Socio-Economic-Demographic Factors Matter in COVID-19 Related Stay-at-Home-Tendencies Across the US States?

Abstract: This study investigates the potential impacts of different socio-economic-demographic (henceforth, SED), factors in COVID-19 related stay-at-home-tendencies (henceforth, COVID-19-SAHTs) in the US. This requires a state-level investigation rather than a country-level since the US states exhibit large SED differences from one another. To this aim, the K-Means Cluster analysis and the panel ARDL (autoregressive distributed lag) models are applied. The main empirical finding indicates that different SED factors in different US states matter in COVID-19-SAHTs. Additionally, people in the states which have more equal income distribution, higher rate basic literacy, and less population density stay at their homes more during the COVID-19 pandemic. These findings may provide some vital pre-information to the state policymakers about how much the people from different SED statuses will tend to comply with future COVID-19 state restrictions such as stay-at-home orders and others. Until the scientists create a proven vaccine for the coronavirus states will most likely continue to issue some COVID-19 restrictions to reduce the spread of this pandemic.

Keywords: COVID-19 Related Stay-at-Home-Tendencies, Socio-economic-demographic factors, US States.
Jel Codes: D53, I15, I18.

1. Introduction

Apart from the technical and terminological differences of different COVID-19 restrictions both the federal and state governments issue these restrictions to curb the spread of this pandemic in the US. Some studies reveal that such restrictions serve this purpose very effectively in this country. For instance, Fowler et al., (2020) combined data on stay-at-home orders (henceforth, SAHOs), daily confirmed COVID-19 cases and fatalities in the US. They found that SAHOs reduced weekly cases, and fatalities at 40% and 59.8%, respectively after three weeks. Masterman (2020) used the differences-in-differences approach and found that SAHOs prevented 1.7 million new cases and 55,000 deaths in the US between mid-March and May 9, 2020. Similarly, Castillo

This is a peer-reviewed, accepted author manuscript of the following article: Ongan, S., & Gocer, I. (2023). Do Different Socio-economic-demographic Factors Matter in COVID-19 Related Stay-at-home-tendencies Across the US States? *Journal of Health Management*, *25*(2), 219-224. https://doi.org/10.1177/09720634231177341

et al., 2020 used the linear regression techniques and found that these orders considerably reduced the infection rates across the US states.

Besides its human tragedy, the full economic damage of the COVID-19 pandemic won't be measured for quite a while. However, many projections sign catastrophic damages of this pandemic on the US economy. For instance, the CBO (the US Congressional Budget Office) estimates that the US GDP will decline by about 12 percent during the second quarter of 2020 (CBO, 2020). Similarly, the IMF's projection is a 5.9 percent annual decline in the US GDP for 2020 (IMF, 2020). The Federal Reserve Bank of Cleveland's report by Sahin et al., (2020) estimates that the unemployment rate will reach 15.8% in May 2020.

Until the scientists create a proven vaccine for this virus, the US government will most likely continue to issue some restrictions to prevent the spread of this pandemic. Therefore, these future restrictions will be vitally important in this battle. However, getting positive returns from these restrictions will mostly depend on how much people from different socio-economic-demographic (SED) statuses will tend to stay at their homes. Hence, this study investigates the potential impacts of different SED factors in current COVID-19 related stay-at-home-tendencies (COVID-19-SAHTs) across the US states which have large SED differences from one another. This investigation may provide some vital pre-information to the state policymakers about the success degrees of their future potential restrictions. They will also get chance to customize their preventive measures in accordance with SED factors to get more successful returns from these COVID -19 restrictions.

The measure of COVID-19-SAHT, as the dependent variable of this study, is defined as the changes of daily community mobilities detected by the Google Map application in the residential places (homes). This application also shows the mobility changes in different categories of places such as transit stations, groceries-pharmacies, and retails-recreations during this pandemic. The independent variable of this study is the number of confirmed COVID-19 cases. Clustered SED factors are provided and explained in the next section.

2. Data Set

COVID-19-SAHT is technically constructed as the daily residential mobility percent change from the baseline and the data of this variable were obtained from the Google Map report[1]. Simply, this data shows whether the people in different states stay more at their residential places (homes) during the COVID-19 pandemic. The cumulative[2] number of confirmed COVID-19 cases (*CASE*) were obtained from the website of the USA FACTS. Clustered SED factors are defined as per capita GDP (*GDPPC*), the GINI coefficient (*GINI*), and person per square mile (*PPSM*) and they were obtained from the websites of the Data Planet by Sage publishing source and Federal Reserve Bank of St. Louis. Literacy is defined as the percentage of adults having *basic literacy*[3] *(BL)*. They were obtained from the NCES (US National Center for Education Statistics) (NCES, 2020a). All variables are in the US state-level. The sample period of the study is FEB/22/2020- MAY/21/2020.

3. Empirical Model

In order to investigate the impacts of SED factors in COVID-19-SAHTs across the US states, the following model in regression form is used:

$$\text{COVID\_19\_SAHT}_{it} = \beta_0 + \beta_1 CASE_{it} + \varepsilon_{it} \qquad (1)$$

where COVID-19-SAHT and *CASE* are COVID-19 related stay-at-home-tendencies and the number of confirmed COVID-19 cases, respectively. ε_t is the innovation term. In this model, we expect the sign of β_1 to be positive since the rising number of confirmed COVID-19 cases (*CASE*) will cause the people to stay at their homes more (denotes increases in COVID-19-SAHTs).

4. Empirical Methodology

The empirical methodology of this study is constructed in two-phased. First, the US states are clustered based on SED factors. To this aim, K-Means Cluster Analysis is applied. Second, in order to estimate the coefficients, the panel ARDL (autoregressive distributed lag) model is

[1] For detailed technical instruction refer https://www.google.com/covid19/mobility/

[2] The reason of using cumulative numbers of the COVID-19 cases is that we assume that people consider increasing cumulative daily numbers rather than current single day when they decide to stay at home.

[3] Basic literacy is defined as the ability of the adults who are able to read, understand, and write any text in English nothing more advanced.

applied for each cluster. Cluster analysis simply classifies the objects (variables) into the number of different groups in which the similar variables are placed in the same group. In this analysis, first, the distances of the variables are measured. The Euclidean distance (d), as a most common distance measure, between variables of x_i and x_j is constructed in the following form (Han et al., 2012):

$$d(x_i, x_j) = \left[\sum_{k=1}^{n} \left|x_{ik} - x_{jk}\right|^2\right]^{1/2} \quad (2)$$

where k is number of variables (objects) and n is the dimension of Euclidean space. Following measuring distances, we produce cluster three to find the optimal number of clusters. Finally, we apply the K-Means Cluster analysis (technique) which minimizes within-cluster variances and classifies each variable with the closest average. This is one of the most used and simplest techniques to classify the variables (Morissette and Chartier, 2013). Following this analysis, in the second phase of the empirical methodology of this study, we apply the panel ARDL model by Pesaran et al. (1999) for each cluster based on SED factors. On the contrary of traditional cointegration analysis within the system of equations, this model tests the long-run relation for individual briefed form of equation regardless of variables whether are I(0), I(1), or both I(0) and I(1) (Pesaran and Shin, 1999; Narayan, 2005).

$$\mathrm{COVID\_19\_SAHT}_{it} = \sum_{j=1}^{p} \lambda_{ij} COVID\_19\_SAHT_{i,t-j} + \sum_{j=0}^{q} \delta'_{ij} CASE_{i,t-j} + \mu_i + \varepsilon_{it} \quad (3)$$

where p and q are the number of optimal lags of the dependent and the independent variables. $i = 1, .., N$: the total number of states, $t = 1, \dots T$: time dimension in the series, μ_i: fixed effects, $CASE_{it}$: independent variables vector ($kx1$), $\mathrm{COVID\_19\_SAHT}_{i,t-j}$:dependent lagged value of the variable, δ_{ij}: ($kx1$) coefficients vector and λ_{ij}: the coefficient of lags of the dependent variable. If we rewrite Eqn. 3 in error correction form, we obtain the following model in Eqn. 4:

$$\Delta COVID\_19\_SAHT_{it} = (\varphi_i COVID\_19\_SAHT_{i,t-1} + \beta'_i CASE_{it}) + \sum_{j=1}^{p-1} \lambda^*_{ij} \Delta COVID\_19\_SAHT_{i,t-j} + \sum_{j=0}^{q-1} \delta'^*_{ij} \Delta CASE_{i,t-j} + \mu_i + \varepsilon_{it} \quad (4)$$

where $\varphi_i = -\left(1 - \sum_{j=1}^{p} \lambda_{ij}\right)$, $\beta_i = \sum_{j=0}^{q} \delta_{ij}$, $\lambda_{ij}^{*} = -\sum_{m=j+1}^{p} \lambda_{im}$, $j = 1, 2, \ldots, p-1$ and $\delta_{ij}^{*} = -\sum_{m=j+1}^{q} \delta_{im}$, $j = 1, 2, \ldots, q-1$. Δ represents the differences of series. φ_i: the coefficient of speed of adjustment to the long run status, β_i': the long-run coefficients, λ_{ij}^{*} and $\delta_{ij}'^{*}$: short-run coefficients of dependent and independent variables, respectively.

5. Empirical Findings

In this section of the study, we provide the empirical findings of the steps taken in empirical methodology. First, clusters, obtained by K-Means Cluster analysis based on SED factors, are reported.

Table 1: Clusters Obtained by K-Means Based on SED Factors

Criteria		**Cluster 1**	**Cluster 2**	**Cluster 3**
GDPPC	Final Cluster Center	40,962	34,743	28,619
	Number of Cluster Members	5	17	28
PPSM	Final Cluster Center	1201.04	495.08	1.32
	Number of Cluster Members	4	6	40
GINI	Final Cluster Center	0.51	0.47	0.42
	Number of Cluster Members	5	29	16
BL	Final Cluster Center	94	86	77
	Number of Cluster Members	24	22	4

The K-Means Cluster analysis created three clusters based on SED factors. For instance, for *GDPPC*, 5 states in Cluster 1, 17 states in Cluster 2 and 28 sates in Cluster 3. The states in Cluster 1 have higher per capita incomes (*GDPPCs*), population densities (*PPSMs*), rates of literacy (*BLs*) but worse income distributions (*GINIs*) than the states in Clusters 2 and 3. Following the findings of the K-Means Cluster analysis, before estimating the panel ARDL model, we, first, must make sure whether the series are stationary. To this aim, we apply the LLC by Levin, Lin and Chu (2002), the IPS by Im, Pesaran and Shin, (2003) and the Hadri by Hadri (2000) panel unit root tests. While null hypotheses of the LLC and IPS tests are "s*eries has a unit root*", this is "s*eries stationary"* for Hadri test. The results of these two tests are reported in Table 2.

Table 2: Panel Unit Root Test Results

Criteria	Cluster	Variables	LLC		IPS		Hadri	
			Level	First Diff.	Level	First Diff.	Level	First Diff.
GDPPC	Cluster 1	COVID-19-SAHT	0.23	0.00***	0.00***	-	0.00	0.75***
		CASE	0.08*	0.00***	0.15	0.00***	0.00	0.90***
	Cluster 2	COVID-19-SAHT	0.09*	0.00***	0.00***	-	0.00	0.56***
		CASE	0.00***	-	0.03**	-	0.00	0.99***
	Cluster 3	COVID-19-SAHT	0.00***	-	0.00***	-	0.00	0.99***
		CASE	0.00***	-	0.00***	-	0.00	0.99***
PPSM	Cluster 1	COVID-19-SAHT	0.06*	0.00***	0.00***	-	0.00	0.91***
		CASE	0.03**	-	0.00***	-	0.00	0.88***
	Cluster 2	COVID-19-SAHT	0.03**	-	0.00***	-	0.00	0.95***
		CASE	0.06*	0.00***	0.00***	-	0.00	0.93***
	Cluster 3	COVID-19-SAHT	0.00****	-	0.00***	-	0.00	0.97***
		CASE	0.00***	-	0.00***	-	0.00	0.77***
GINI	Cluster 1	COVID-19-SAHT	0.06*	0.00***	0.00***		0.00	0.93***
		CASE	0.10	0.00***	0.07*	0.00***	0.00	0.91***
	Cluster 2	COVID-19-SAHT	0.00***	-	0.00***	-	0.00	0.96***
		CASE	0.00***	-	0.00***	-	0.00	0.80***
	Cluster 3	COVID-19-SAHT	0.00***	-	0.00***	-	0.00	0.99***
		CASE	0.00***	-	0.00***	-	0.00	0.99***
BL	Cluster 1	COVID-19-SAHT	0.00***	-	0.00***	-	0.00	0.99***
		CASE	0.00***	-	0.00***	-	0.00	0.99***
	Cluster 2	COVID-19-SAHT	0.00***	-	0.00***	-	0.00	0.99***
		CASE	0.00***	-	0.00***	-	0.00	0.99***
	Cluster 3	COVID-19-SAHT	0.07*	0.00***	0.00***	-	0.00	0.91***
		CASE	0.04**	-	0.00***	-	0.00	0.88***

Note: Table provides P-values. ***, ** and * denote statistical significances at 1%, 5% and 10% levels, respectively.

Test results in Table 2 indicate that the series are stationary in different levels. Hence, we apply the ARDL model. The estimated coefficients of the panel ARDL model are reported in Table 3.

Table 3: Panel ARDL Model Results

			GDPPC	***PPSM***	***GINI***	***BL***
Cluster 1	Long Run	$CASE_t$	0.0001*** (0.00)	0.0001*** (0.00)	0.00005*** (0.00)	0.0003*** (0.00)
	Short Run	ECT_{t-1}	-1.12*** (0.00)	-1.10*** (0.00)	-1.16*** (0.00)	-1.15*** (0.00)
		$\Delta CASE_t$	-0.0009** (0.03)	-0.0007* (0.09)	-0.0001*** (0.00)	-0.002*** (0.00)
		$\Delta CASE_{t-1}$	0.0004 (0.32)	0.0005 (0.35)	-0.00001 (0.12)	-0.0002 (0.23)
		$\Delta CASE_{t-2}$	0.0002 (0.90)	0.0001 (0.53)	-0.00002** (0.01)	-0.0008*** (0.00)
		$\Delta CASE_{t-3}$	0.001 (0.13)	0.001 (0.23)	0.0001*** (0.00)	0.002*** (0.00)
		$Constant$	14.13*** (0.00)	13.93*** (0.00)	12.63*** (0.00)	11.17*** (0.00)
Cluster 2	Long Run	$CASE_t$	0.0008*** (0.00)	0.0005*** (0.00)	0.0001*** (0.00)	0.0001*** (0.00)
	Short Run	ECT_{t-1}	-1.09*** (0.00)	-1.13*** (0.00)	-1.16*** (0.00)	-1.15*** (0.00)
		$\Delta CASE_t$	-0.001** (0.01)	-0.0002*** (0.00)	-0.001*** (0.00)	-0.0007*** (0.00)
		$\Delta CASE_{t-1}$	0.0005 (0.34)	0.0002 (0.44)	-0.0001 (0.29)	0.0002 (0.47)
		$\Delta CASE_{t-2}$	-0.0001 (0.76)	-0.0003*** (0.00)	-0.0004** (0.04)	0.00009 (0.78)
		$\Delta CASE_{t-3}$	0.002** (0.01)	0.0002** (0.03)	0.0009*** (0.00)	0.001* (0.05)
		$Constant$	12.16*** (0.00)	12.94*** (0.00)	10.99*** (0.00)	10.90*** (0.00)
Cluster 3	Long Run	$CASE_t$	0.0002*** (0.00)	0.0002*** (0.00)	0.0002*** (0.00)	0.00005*** (0.00)
	Short Run	ECT_{t-1}	-1.19*** (0.00)	-1.16*** (0.00)	-1.13*** (0.00)	-1.17*** (0.00)
		$\Delta CASE_t$	-0.001*** (0.00)	-0.001*** (0.00)	-0.002*** (0.00)	-0.0001*** (0.00)
		$\Delta CASE_{t-1}$	-0.0003** (0.01)	0.0001 (0.95)	0.0003 (0.52)	-0.0001** (0.02)
		$\Delta CASE_{t-2}$	-0.0004** (0.00)	-0.0003 (0.16)	-0.0003 (0.58)	-0.00002** (0.02)
		$\Delta CASE_{t-3}$	0.0008** (0.02)	0.001*** (0.00)	0.002** (0.00)	0.00007*** (0.00)
		$Constant$	10.27*** (0.00)	10.85*** (0.00)	11.39*** (0.00)	12.45*** (0.00)

Note: ***, ** and * denote statistical significances at 1% and 5% and 10% levels, respectively. The error correction mechanisms work since their coefficients are significantly negative.

Test results in Table 3 indicate that rising confirmed COVID-19 cases cause the people to stay at their homes more (denotes increases in COVID-19-SAHTs) in all clusters for each SED (socio-economic-demographic) factor in the long-run since the coefficients of $CASE_t$ are significant and positive. However, this situation is completely opposite in the short-run since the coefficients of $\Delta CASE_t$ are significant but negative. This means that rising number of cases cause the people to stay at their homes less. This discrepancy can be interpreted that people in the states regardless their SED statuses consider ever-increasing daily cumulative number of cases in the long-run rather than current single day in the short-run. In regards to *GDPPC*, the people in the states which have the highest and the lowest per capita income (in Clusters 1 and 3, respectively) stay at their homes less than the people in Cluster 2. If we can define the people in Cluster 2 as middle-class just for our own research classification scale, we can conclude that middle-class people stay at their homes the most (0.0008). Interestingly, the people in the states which have the highest per capita income in Cluster 1 stay at their homes the least. The low responses in Cluster 3 may be explained that the people in this cluster might not have health insurance due to high cost and feel themselves unprotected outside. However, we cannot make the same interpretation for the people in Cluster 1 since they have high income and most likely high level health insurance, but they do not stay at their homes more than the people in Clusters 2 and 3. Other potential factors should lead to this result in Cluster 1.

Furthermore, maybe, income distribution (*GINI*) can explain above findings better than *GDPPC* or change them. Because income is distributed the worst in the states in Cluster 1 which have the highest *GDPPC*. Significantly positive coefficients of *GINI* clearly indicate that people in the states which have more equal income distributions stay at their homes more than the people in the states which have less equal distributions in the long-run (0.0002 and 0.0001 in Clusters 2 and 3 > 0.00005 in Cluster 1). This means that more equal income distributions, the more staying at homes. In the comparison of the findings of *GINI* and *GDPPC*, it seems that income distribution (*GINI*) plays a more determining and measurable role in staying at home than *GDPPC*.

In regards to *BL* (basic literacy), significantly positive coefficients of *BL* indicate that literacy plays a determining role in COVID-19-SAHTs in the long-run. The people in the states which have higher rates of basic literacy (Clusters 1 and 2) stay at their homes more than the people in the states which have lower rates of basic literacy (Cluster 3). The highest basic literate people in Cluster 1 stay at their homes the most (0.0003). This can be interpreted that the higher rates of basic literacy, the more staying at homes (the more comprehensions of the dangers of this pandemic). At this point, it should be also noted that about 43 million (NCES, 2020b) US adults who do not possess basic literacy levels pose a serious danger in preventing the spread of this virus. This finding may guide the state and federal government policymakers to increase the awareness of the risks of this pandemic on these people who are not able to read, understand, and write anything in English.

In regards to *PPSM*, significantly positive coefficients of *PPSM* indicate that population density matters in staying at homes in the long-run. The people in the states which have the highest population density (Cluster 1) stay at their homes less than the people in the states in Clusters 2 and 3. This may be interpreted that people in rural areas do not need to be outside more than the people who live in crowded cities.

6. Conclusion

Scientists have been working hard to create a proven coronavirus vaccine. However, it seems that this vaccine will take such a long time. Therefore, the US government will most likely continue to issue such COVID-19 restrictions to prevent the spread of this pandemic. At this point, the success of these potential future restrictions will most likely be determined by how much the people from different socio-economic-demographic (SED) statuses will tend to stay at their homes. Hence, this study investigates the potential impacts of different SED factors in current COVID-19 related stay-at-home-tendencies (COVID-19-SAHTs) across the US states which have large SED differences from one another. Empirical findings indicate that people in the states which have more equal income distribution, higher level basic literacy and less population density stay at their homes more. Hence, this investigation may provide some vital pre-information to the state policymakers about the success degrees of future potential restrictions. They will also get chance to customize their preventive measures in accordance with SED factors to get more successful feedbacks from these COVID -19 restrictions.

References

Castillo, R.C., Staguhn, E.D. and Weston-Farber, D. (2020). The effect of state-level stay-at-home orders on COVID-19 infection rates. *American Journal of Infection Control*, (Forthcoming).

CBO (2020). US Congressional Budget Office, CBO's Current Projections of Output, Employment, and Interest Rates and a Preliminary Look at Federal Deficits for 2020 and 2021. https://www.cbo.gov/publication/56335/, (accessed 06 June 2020).

Data Planet (2020). Sage Publishing Source.
https://www.data-planet.com/data-planet-statistical-datasets/, (accessed 03 June 2020).

Fowler, J.H., Hill, S.J., Levin, R. and Obradovich, N. (2020), The Effect of Stay-at-Home Orders on COVID-19 Cases and Fatalities in the United States. file:///C:/Users/serda/Downloads/The_Effect_of_Stay-at-Home_Orders_on_COVID-19_Infe.pdf/, (accessed 06 June 2020).

FRED (Federal Reserve Bank of St. Louis). (2020). FRED Economic Data. https://fred.stlouisfed.org/, (accessed 06 June 2020).

Google. (2020). COVID-19 Community Mobility Reports https://www.google.com/covid19/mobility/, (accessed 06 June 2020).

Hadri, K. (2000). Testing for stationarity in heterogeneous panels. *The Econometrics Journal,* 3(2), 148-161.

Han, J., Kamber, M. and Pei, J. (2012). *Data Mining* (Chapter 10; Cluster Analysis: Basic Concepts and Methods). (Third Edition), Morgan Kaufmann, San Francisco.

Im, K., Pesaran, H., and Y. Shin. (2003). Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, 115(1), 53-74.

IMF Report (2020). The Great Lockdown: Worst Economic Downturn Since the Great Depression. https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/, (accessed 25 May 2020).

Levin, A., Lin, F. and C. Chu. (2002), Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of Econometrics*, 108(1), 1-24.

Masterman, J.C. (2020). Stay-at-Home Orders and COVID-19 Fatalities. file:///C:/Users/serda/Downloads/SSRN-id3600905.pdf/. (accessed 06 June 2020).

Morissette, L. and Chartier, S. (2013). The K-Means Clustering Technique: General Considerations and Implementation in Mathematica, *Tutorials in Quantitative Methods for Psychology,* 9(1), 15-24.

Narayan, P. K. (2005). The Saving and Investment Nexus for China: Evidence from Cointegration Tests. *Applied Economics*, 37(17), 1979–1990.

NCES. (2020a). The US Center for Education Statistics, National Assessment of Adult Literacy. https://nces.ed.gov/naal/estimates/StateEstimates.aspx/, (accessed 07 June 2020).

NCES. (2020b). Adult Literacy in the United States https://nces.ed.gov/datapoints/2019179.asp, (accessed 07 June 2020).

Pesaran, M.H. and Shin, Y. (1995). Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis; DAE Working Paper Series No. 9514; Department of Applied Economics, University of Cambridge.

Pesaran, M. H., Shin, Y. and Smith, R. P. (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, 94(446), 621 – 634.

Pesaran, M. H.; Shin, Y. (1999). An Autoregressive Distributed-Lag Modelling Approach to Cointegration Analysis. In: Strom, S. (ed.). Econometrics and Economic Theory in the 20th Century, 371-413.

Sahin, A., Tasci, M. and Yan, J. (2020). The Unemployment Cost of COVID-19: How High and How Long? Economic Commentary, Number: 2020-09, The Federal Reserve Bank of Cleveland.

USA FACTS. (2020). Coronavirus Locations: COVID-19 Map County and State https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/