Michelle S. Nemiada
Econometrics Terminal Report
May 20, 2021

Investigating Relationship between Philippines' Covid-19 Cases and Covid-19-related Policies

It has already been more than a year since the Covid-19 pandemic struck the world and impacted people's lives and economies of almost all countries around the globe. As a result, governments mandated and enforced various measures, policies, and safety protocols to minimize the effect of the pandemic.

In the Philippines, as of May 17, 2021, DOH reported a total of 1,149,925 cases recorded across the country (*COVID-19 Case Tracker | Department of Health Website*, 2021). The government has implemented national and local policies since the start of the pandemic last 2020 to reduce the burdens on healthcare facilities while vaccines are not yet widely available.

This analysis investigates the relationship between the Covid-19-related policies and the daily growth rate of Covid-19 cases in the Philippines. This report has the following sections:

I. Economic theory and economic model
II. Model specification, assumptions, and justifications
III. Hypothesis testing
IV. Data and statistical software
V. Results and diagnostic tests
VI. Conclusion and recommendations

Economic theory and economic model

The economic theory and model for this analysis is similar to the one used by Hsiang et al. (2020), which analyzed the effects of Covid-19-related policies on Covid-19 growth rates among six countries, on national, and local levels. Their analysis measured local infection growth rates associated with changes in anti-contagion policies. A necessary condition for the study is the assumption that the timing of policy deployment is independent of infection growth rates, following the epidemiological theories also cited in Hsiang et al. (Chowell et al., 2016; Ma, 2020; Tillett, 1992). These theories suggest that cases would grow exponentially early in the epidemic without anti-contagion policies. Hence, pre-policy infection growth rates should be constant over time and uncorrelated with the timing of policies. Empirical evidence (Aylward et al., 2014; Nishiura et al., 2010) also supports these theories.

For this analysis, the researcher used a model similar to the Hsiang et al. (2020) study. A time-varying outcome growth rate g is modeled as:

$$g_t = \beta_0 + \beta_1 policy_t + \epsilon_t$$

where:

β_0 : *average growth rate without a policy*

$policy_t$: *count of how many* $policy_t$ *were implemented in time t*

β_1 : *average effect of the policy on growth rate g over all periods subsequent to the introduction of the policy thereby encompassing any lagged effects of policies*

ε_t : *a mean − zero disturbance term that captures interperiod changes not described by* $policy_t$

A multiple regression model will be estimated using OLS, adding into the equation different policy categories.

Model specification, assumptions, and justifications

As mentioned in the previous section, this analysis utilized a multiple linear regression model. From the data, there are a total of 9 policy categories curated. These are the policies included in the model; thus, the model is specified as follows:

$$g_t = \beta_0 + \beta_1 TRAVEL.BAN + \beta_2 TRAVEL.BAN.LIFT + \beta_3 EASE + \beta_4 POSTPONE + \beta_5 AID + \beta_6 PROTOCOL + \beta_7 TESTING + \beta_8 REOPEN + \beta_9 REPORTING.CHANGE + \varepsilon_t$$

where:

g_t : *infection growth rate at time t computed as the (number of cases on t subtracted by the number of cases in* $t-1$ *)/number of cases in* $t-1$

$TRAVEL.BAN$: *implementation of travel ban (local and national)*

$TRAVEL.BAN.LIFT$: *implementation lifting of travel ban (local and national)*

$EASE$: *easing of protocols such as travel restrictions, border restrictions, health and safety protocols*

$POSTPONE$: *postponement of activities (includes MECQ, ECQ, work from home enforcement, cancellation of classes and community activities)*

AID : *distribution of economic aid to members of the community greatly impacted by the pandemic* $(i.e.,$ *cash amelioration, aid for displaced workers and frontliners, etc*$)$

$PROTOCOL$: *stricter safety protocol pronouncements and implementation*

$TESTING$: *testing initiatives either by government, NGOs or private firms or individuals*

$REOPEN$: *reopening of recreation areas, restaurants, salons, tourist spots, etc.*

$REPORTING.CHANGE$: *changes on how DOH reports Covid* $-$ 19 *cases*

β_0 : *fixed* $-$ *effect; average growth rate without a policy*

β_1 *to* β_9 : *average effect of the policy on growth rate g over all periods subsequent to the introduction of the policy thereby encompassing any lagged effects of policies*

ε_t : *a mean* $-$ *zero disturbance term that captures interperiod changes not described by the policies*

The model used in Hsiang et al. (2020), which was the basis of the model used for this analysis, took advantage of the susceptible–infected–recovered (SIR) disease model, which considers the growth of infection rates during the early period of the pandemic. That is, the infection rates grow exponentially in the absence of policies. For this study, however, the early period cases were not available. Furthermore, the first case in the data already has an ongoing policy implemented, so the same assumption of exponential growth is not applicable. Thus, the researcher used the typical computation for growth rates (number of cases on *t* subtracted by the number of cases in *t-1*)/number of cases in *t-1*.

The assumptions in this analysis are the assumptions for linear regression modeling which includes:

1. A linear relationship between the response and predictor variables
2. Normality of residuals
3. Homoscedasticity or constant variance of residuals
4. Independence of residual errors
5. Multicollinearity between predictors does not exist

Hypothesis testing

The null hypothesis for the analysis is that policies have no significant effect on infection growth rates; mathematically, it is:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$

$H_1: at\ least\ 1\ \beta\ is\ nonzero$

Theoretically, the effect of $TRAVEL.BAN$, $POSTPONE$, $PROTOCOL$, $REPORTING.CHANGE$ should be negative since the first three policies restrict gatherings, travels, and activities and calls for stricter implementation of health and safety protocols. On the other hand, the reporting change based on the past reports (DOH Explains Highest Daily Spike in Infections, 2020; Neil, 2020) lowers the number of reported cases. In contrast, the other variables, $TRAVEL.BAN.LIFT$, $EASE$, $TESTING$, $REOPEN$, should have a positive relationship with the infection growth rates. These factors, except for $TESTING$, relax protocol implementations and permit more mass gatherings, activities, and travel. In comparison, more $TESTING$ may be associated with more documented cases. AID, however, may have varying effects based on the manner of its distribution. For example, aids given on a door-to-door approach require fewer people on the streets, thus, reducing the chances of infection. On the other hand, aids distributed in designated distribution centers force people to queue. Therefore, in this scenario, AID may increase infection growth rates.

Data and statistical software

The researcher used the 2020 reported Covid-19 cases across the country retrieved from a Case Information Data Drop from the DOH website[1]. The data has 473,653 case records reported from all regions of the Philippines. However, the analysis only included the 2020 data since new variants have emerged during the first quarter of 2021, which may influence the result of the study.

The DOH Covid-19 data has recorded the first case on January 23, 2020, and the last case on December 31, 2020. Table 1 shows the distribution of 2020 cases per region.

Table 1: Distribution of Covid-19 Cases in 2020 Across Different Regions

Region	Total Records
Unknown	3124
BARMM	3820
CAR	8230
CARAGA	4752
NCR	209987
ROF	16972
Region I: Ilocos Region	5173
Region II: Cagayan Valley	5498

[1] https://doh.gov.ph/covid-19/case-tracker

Region III: Central Luzon	31136
Region IV-A: CALABARZON	86343
Region IV-B: MIMAROPA	2690
Region IX: Zamboanga Peninsula	5842
Region V: Bicol Region	4577
Region VI: Western Visayas	22593
Region VII: Central Visayas	26482
Region VIII: Eastern Visayas	11289
Region X: Northern Mindanao	8401
Region XI: Davao Region	12751
Region XII: SOCCSKSARGEN	3993

The policies were web-scraped from a timeline of Covid-19 policies on the CNN website[2]. The policies are then categorized into the following classifications (this is to avoid sparsity of data in modeling):

Table 2: Covi-19-related Policies Categories

Policy Category	Description
TRAVEL BAN	travel ban (local and national)
TRAVEL BAN LIFT	lifting of a travel ban (local and national)
EASE	easing of protocols such as travel restrictions, border restrictions, health and safety protocols
POSTPONE	postponement of activities (includes MECQ, ECQ, work from home enforcement, cancellation of classes, and community activities)
AID	distribution of economic assistance to members of the community greatly impacted by the pandemic (i.e., cash amelioration, aid for displaced and frontline workers)
PROTOCOL	stricter safety protocol pronouncements and implementation
TESTING	testing initiatives either by government, NGOs or private firms, or individuals

[2] https://cnnphilippines.com/news/2020/4/21/interactive-timeline-PH-handling-COVID-19.html

REOPEN	reopening of recreation areas, restaurants, salons, tourist spots, etc.
REPORTING CHANGE	changes on how DOH reports Covid-19 cases

The is a total of 344 days in the data, which is from January 23, 2020, to December 31, 2020. It also just so happened that the first recorded case in the DOH Covid-19 case report occurred on the same day as the implementation of the first policy (travel ban for flights from Wuhan, China).

Counts of the total number of cases each day were recorded and merged with the policy data. Since there was no clear reporting of how long a policy would take effect, it was assumed that it was in effect until the end of the study period, which is December 12, 2020. However, since conflicting policies may be implemented simultaneously, it is assumed that their effects would cancel out each other. For instance, the government implements a travel at day *t* then a travel ban lift was then implemented at day *t+10*, the effect of the travel ban lift should cancel the effect of the travel ban which was first imposed.

For each day in 2020, the total new cases were counted, the total number of policies in each policy category was recorded. The summary statistics of the variables used for modeling are in the succeeding table.

Table 3: Summary Statistics of Variables

variable	count	mean	std	min	25%	50%	75%	max
Daily Cases	344	1,376.90	1,236.36	0	295.25	1,105.00	2,154.50	5,689.00
growth_rate	344	13.61	103.73	-100	-17.33	0.00	14.05	1,700.00
travel ban	344	8.85	2.01	1	9.00	10.00	10.00	12.00
travel ban lift	344	0.94	0.25	0	1.00	1.00	1.00	1.00
ease	344	3.15	3.88	0	1.00	1.00	3.00	14.00
postpone	344	9.88	5.19	0	6.00	12.00	14.00	18.00
aid	344	2.43	1.13	0	3.00	3.00	3.00	3.00
protocol	344	4.14	4.28	0	1.00	2.00	9.00	12.00
testing	344	4.84	2.79	0	3.00	7.00	7.00	7.00
reopen	344	5.38	4.47	0	0.00	5.00	10.00	11.00
reporting	344	1.08	0.90	0	0.00	1.00	2.00	2.00

change								

The researcher used Python for all data processing and R for the modeling.

Results and diagnostic tests

A multiple regression model estimated using OLS has the following results:

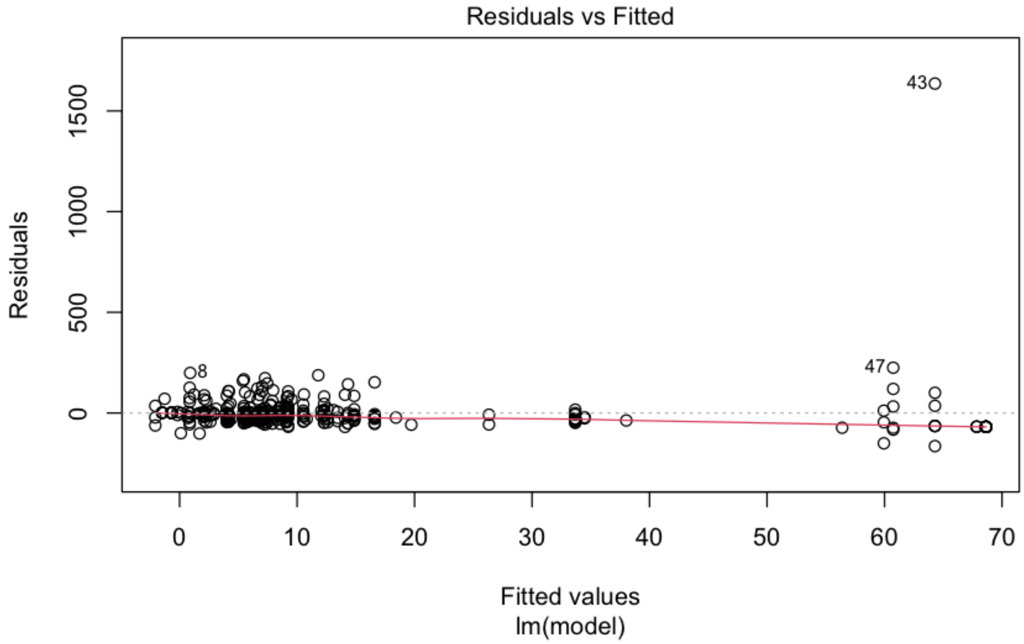
```
model<- growth_rate~travel.ban+travel.ban.lift+ease+postpone+aid+protocol+testing+reopen+reporting.change
fit<- lm(model, data)
summary(fit)
```

```
Call:
lm(formula = model, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-164.29  -32.33  -11.53    5.73 1635.71 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)  
(Intercept)        2.4927    40.6227   0.061   0.9511  
travel.ban        -0.7948    10.8721  -0.073   0.9418  
travel.ban.lift   68.8866    40.4098   1.705   0.0892 .
ease               1.2323     4.1110   0.300   0.7646  
postpone          -3.5511     9.1618  -0.388   0.6986  
aid              -18.3725    20.0811  -0.915   0.3609  
protocol          -1.3977     5.9149  -0.236   0.8133  
testing            6.6103    11.4204   0.579   0.5631  
reopen             2.2677    10.7075   0.212   0.8324  
reporting.change  -8.2847    30.2636  -0.274   0.7844  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 103.6 on 334 degrees of freedom
Multiple R-squared:  0.0295,	Adjusted R-squared:  0.003353 
F-statistic: 1.128 on 9 and 334 DF,  p-value: 0.3419
```

On a 5% level of significance, lifting a travel ban has a statistically significant positive effect on the growth rate of covid cases. However, the adjusted R^2 suggests that the fitted model only explains 0.3% of the variation of the case growth rates.

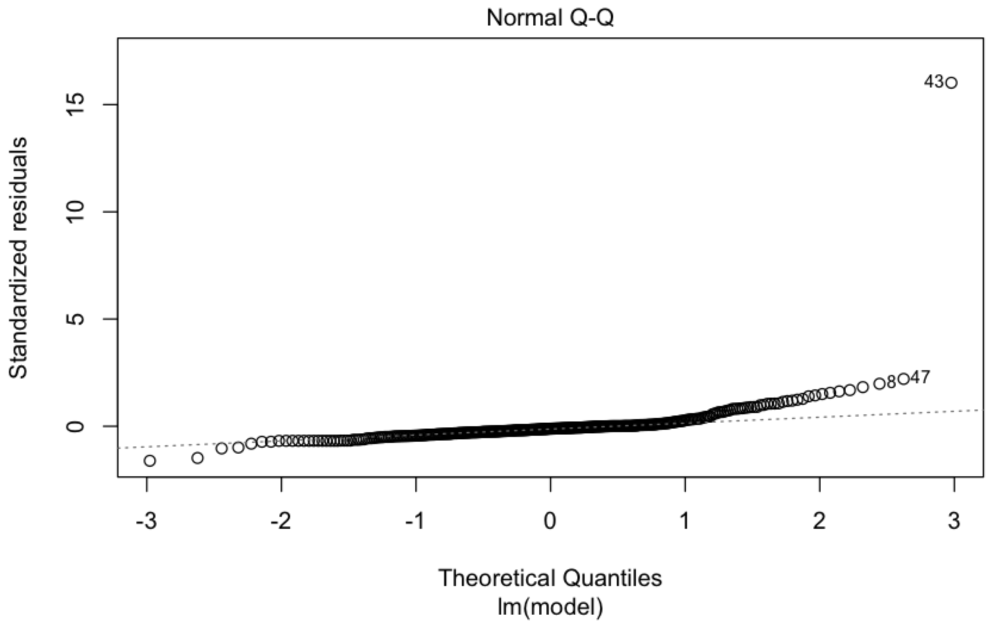
Checking on the OLS assumptions to know if the result is valid, the researcher conducted the subsequent tests.

A. *Linear relationship between the response and predictor variables*



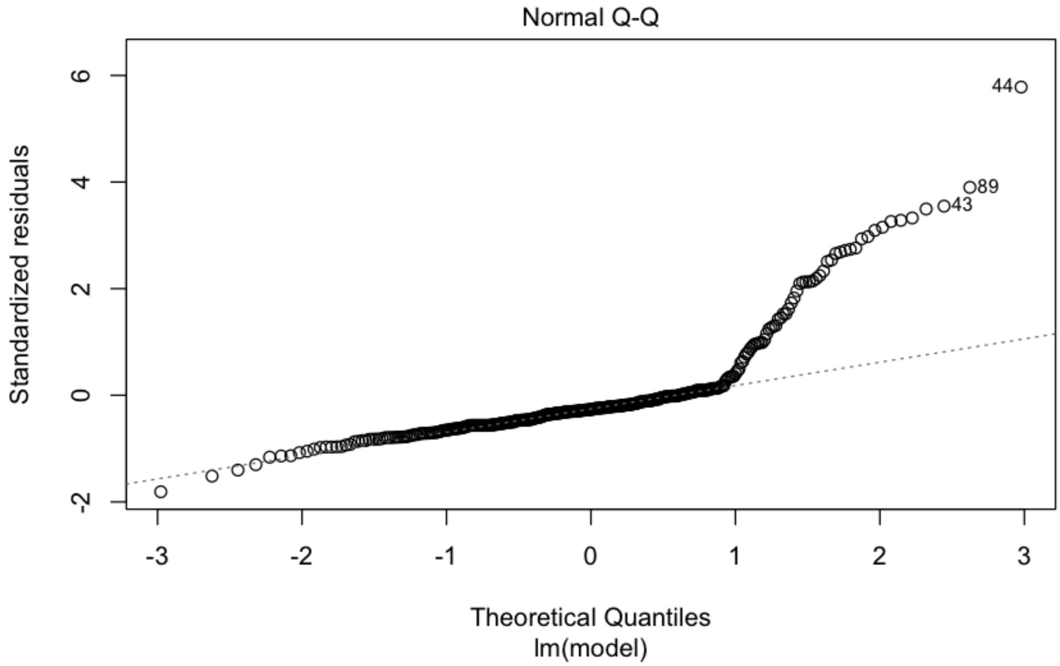
From the above scatter plot that visualizes the residuals, the points in the horizontal line dispersed with no distinct pattern, which is also exhibited by the horizontal residual line (red line) that lies very close to the dotted null hypothesis line (gray line). With this, the assumption that a linear relationship between the response and predictor variables is reasonable. Moreover, the horizontal residual line that lies very close to the null hypothesis line suggests that the variances of the error terms are equal.

B. Normality of residuals



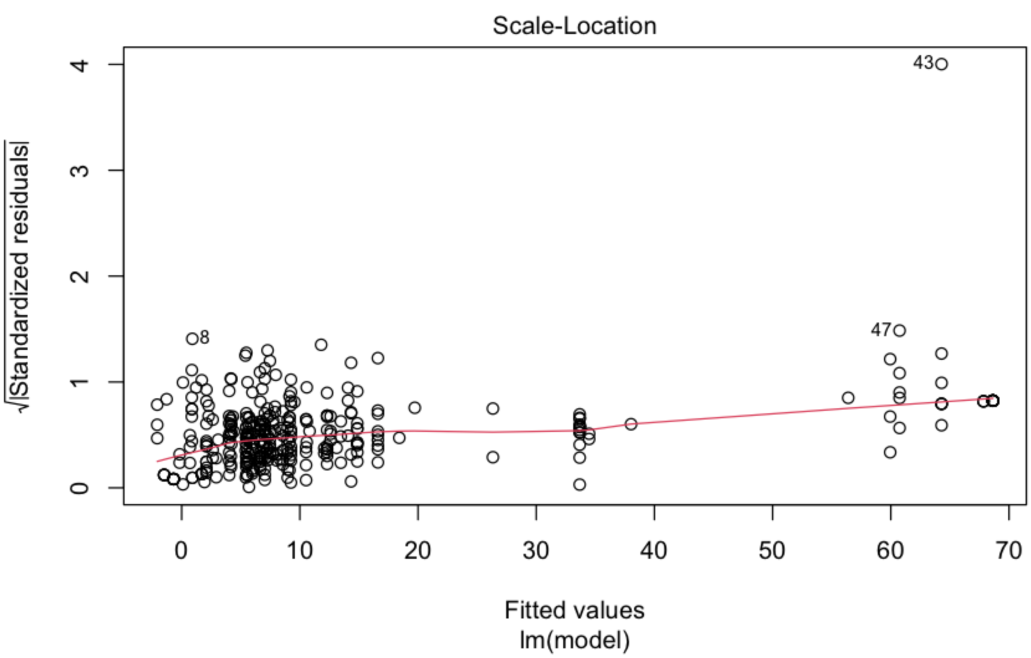
A Normal Q-Q plot detects non-normality in the residuals. The assumption is that the residuals of the fitted model come from a normal distribution. So, if the residuals are also normally distributed, the data points in the scatterplot should follow the dotted gray line. For this model, however, the plot shows that the residuals did not align perfectly with the theoretical values, indicating non-normality.

The researcher investigated the reason for non-normality and fitted a similar model using a dataset with fewer samples. The Normal Q-Q plot is as follows:



Comparing the two plots, the first one with more samples follows the dotted gray line better than the second model. The difference between the two plots shows that the normality of residuals may be due to a lack of samples, and adding more samples would resolve the issue.

C. Homoscedasticity or constant variance of residuals

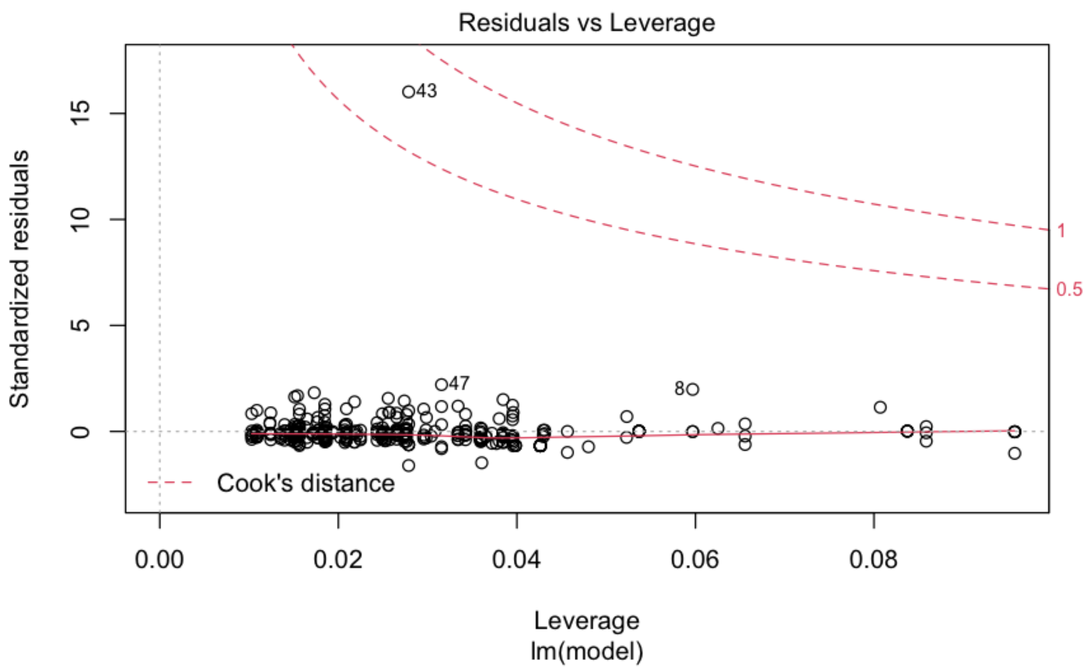


A Scale-Location plot draws residuals to check for homoscedasticity. From the plot, if the residuals have constant variance, the red line should be a horizontal line, which means that the residuals are spread equally along the ranges of predictors. However, the plot above shows that the red line may not be perfectly horizontal but still roughly is. Therefore, to make sure that the model does not violate the assumption of homoscedasticity, the researcher conducted a Breusch-Pagan test which gave the following results:

```
        studentized Breusch-Pagan test

data:  fit
BP = 9.5637, df = 9, p-value = 0.3869
```

The null hypothesis for the Breusch-Pagan test is that the residuals are homoscedastic. From the above result, the p-value is 0.3869, which means there is insufficient evidence to reject the null hypothesis. Thus, the assumption of homoscedasticity holds.

D. *Independence of residual errors*



The Residuals versus Leverage plot assesses if there exist residuals of extreme values that might influence the regression results when included or excluded from the analysis. The above plot shows that the 43rd observation may impact the regression result.

E. Multicollinearity between predictors does not exist

The correlation values between predictors tell us of the presence of multicollinearity. All values in the table below were less than 0.7, suggesting that multicollinearity is not present.

```
$estimate
                  travel.ban travel.ban.lift        ease    postpone         aid    protocol
travel.ban        1.00000000     0.683027107 -0.13008580  0.39744475  0.30392498  0.02555012
travel.ban.lift   0.68302711     1.000000000  0.05720912 -0.07083772 -0.14037756 -0.06502058
ease             -0.13008580     0.057209117  1.00000000  0.50497241 -0.25932339  0.65524263
postpone          0.39744475    -0.070837719  0.50497241  1.00000000  0.39434800 -0.17745707
aid               0.30392498    -0.140377559 -0.25932339  0.39434800  1.00000000  0.33380046
protocol          0.02555012    -0.065020578  0.65524263 -0.17745707  0.33380046  1.00000000
testing          -0.18787111    -0.047636473 -0.25669557  0.46432244  0.41704077 -0.24435687
reopen           -0.06344163     0.005252573 -0.18012246  0.35287354 -0.48284906  0.63946324
reporting.change  0.01826173    -0.006836034 -0.17327194  0.03213234 -0.01652262 -0.20401987
                     testing       reopen reporting.change
travel.ban       -0.18787111 -0.063441630      0.018261726
travel.ban.lift  -0.04763647  0.005252573     -0.006836034
ease             -0.25669557 -0.180122463     -0.173271935
postpone          0.46432244  0.352873539      0.032132338
aid               0.41704077 -0.482849065     -0.016522622
protocol         -0.24435687  0.639463244     -0.204019865
testing           1.00000000  0.305130497     -0.057166383
reopen            0.30513050  1.000000000      0.696506734
reporting.change -0.05716638  0.696506734      1.000000000
```

Conclusion and recommendations

The regression model estimated Covid-19-related policies’ effects on the growth rate of Covid-19 cases in the Philippines. The result suggests that the lifting of travel bans has a positive statistically significant impact on the growth rates of Covid-19 cases. The hypothesized effects of the other factors also surfaced on the study’s findings. However, there was insufficient evidence to conclude their effect. Estimating the model on more samples may resolve this issue, which would also solve the non-normality of residuals.

Nevertheless, the simple model used in this analysis showed how Covid-19-related policies affect Covid-19 reported cases growth rates. Future research may modify the model specification, dig deeper at the more localized policy implementation on a regional or LGU level, and incorporate recoveries and deaths in the modeling. One caveat for this analysis is that the policies in the data were mainly on the national level, but the cases were from the local governments. Since local policies may vary in every jurisdiction, it would be a step further to investigate the relationship between Covid-19-related policies and Covid-19 growth rates in areas where the cases continue to increase despite efforts to mitigate the impact of the pandemic.

References:

Aylward, B., Barboza, P., Bawo, L., Bertherat, E., Blake, I., Brennan, R., Briand, S., Chakauya, J., Chitala, K., Conteh, R., Cori, A., Croisier, A., Jean-Marie, D., Diallo, B., Donnelly, C., Dye, C., Eckmanns, T., Ferguson, N., & Yoti, Z. (2014). Ebola Virus Disease in West Africa—The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine*, *371*, 1481–1495. https://doi.org/10.1056/NEJMoa1411100

Chowell, G., Sattenspiel, L., Bansal, S., & Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, *18*, 66–97. https://doi.org/10.1016/j.plrev.2016.07.005

COVID-19 Case Tracker | Department of Health website. (n.d.). Retrieved May 19, 2021, from https://doh.gov.ph/covid-19/case-tracker

DOH explains highest daily spike in infections: Only 20% were "fresh cases." (2020, May 29). Cnn. https://cnnphilippines.com/news/2020/5/29/DOH-explains-spike-covid-19-cases-.html

Ericksen, E. P. (1974). A Regression Method for Estimating Population Changes of Local Areas. *Journal of the American Statistical Association*, *69*(348), 867–875. https://doi.org/10.2307/2286155

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J., & Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*(7820), 262–267. https://doi.org/10.1038/s41586-020-2404-8

Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, *5*, 129–141. https://doi.org/10.1016/j.idm.2019.12.009

Multicollinearity in R. (n.d.). *DataScience+*. Retrieved May 19, 2021, from https://datascienceplus.com/multicollinearity-in-r/

Neil. (2020, May 29). *DoH to change how it reports COVID-19 cases*. BusinessWorld. https://www.bworldonline.com/doh-to-change-how-it-reports-covid-19-cases/

Nishiura, H., Chowell, G., Safan, M., & Castillo-Chavez, C. (2010). Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theoretical Biology & Medical Modelling*, *7*, 1. https://doi.org/10.1186/1742-4682-7-1

Rekha M (2020, March 23). *Correlation and Collinearity—How they can make or break a model*. Medium. https://blog.clairvoyantsoft.com/correlation-and-collinearity-how-they-can-make-or-break-a-model-9135fbe6936a

Tillett, H. E. (1992). Infectious Diseases of Humans; Dynamics and Control. *Epidemiology and Infection*, *108*(1), 211.

TIMELINE: How the Philippines is handling COVID-19. (n.d.). Cnn. Retrieved May 17, 2021, from https://cnnphilippines.com/news/2020/4/21/interactive-timeline-PH-handling-COVID-19.html

Understanding Diagnostic Plots for Linear Regression Analysis | University of Virginia Library Research Data Services + Sciences. (n.d.). Retrieved May 19, 2021, from https://data.library.virginia.edu/diagnostic-plots/