COVID-19 effects on NASDAQ

Hongye Li, 918964869

Mar 13, 2022

1 Abstract

In this paper, we focus on answering the question that if the changes in daily newly reported COVID-19 cases will influence the stock market index prices. We started from a simple model to see the direct relation between cases changes and prices. To eliminate some confounding effects, we added some variables to the model under some assumptions. Then, we considered the effect of time-series in the model solved by Autoregressive model. In the end, we got the conclusion that increasing in daily newly report COVID-19 cases may draw a higher stock market index prices. We conducted model diagnostics to the model which shows that the assumptions we proposed hold. We admit that insufficient covariates, outcome variable selection, and generalization ability of model are the caveats that will influence the results. The conclusion of this report may provide some insights to investors during and possibly a long time after the pandemic.

2 Introduction

In this paper, we are interested to study the relationship between new daily COVID-19 cases fluctuation and the NASDAQ index prices. We happened to find a potential correlation between these two variables, so we decided to find if there is an causal effect between them. Since we are experiencing tons of uncertainties for the recent years, these uncertainties make our life more unpredictable than ever which may lead to a severe problem to the economics and personal life. The stock market is connected to many people's personal financial status, and also the economics of the world. The BBC news states that "More than five million people became millionaires across the world in 2020 despite economic damage from the Covid-19 pandemic." (https://www.bbc.com/news/business-57575077), and the Reuters says "Forbes' annual world's billionaires list this year included a record-breaking 2,755 billionaires with a combined worth of 13.1 trillion dollars, up from 8 dollars trillion last year." (https://www.reuters.com/business/pandemic-boosts-super-rich-share-global-wealth-2021-12-07/), while the World Bank Blogs gives the title "COVID-19 leaves a legacy of rising poverty and widening inequality" (https://blogs.worldbank.org/developmenttalk/covid-19-leaves-legacy-rising-poverty-and-widening-inequality). These news reports tells us that the wealth of the world has gone through a huge drift during the pandemic. In the meanwhile, the stock market just went trough a bull market except the panic period at the beginning of the pandemic and recent policy risk. This phenomenon seems not a healthy and normal transformation of wealth drift, while it may be exacerbated by the pandemic, but we are not sure about this. To move forward to the answer of such problem, we decided to conduct research on the following question:

- Whether the fluctuation of daily newly reported cases has influences on NASDAQ index or not.

- What are the influences between them?

In this paper, we focus our attention on the US stock market NASDAQ (https://www.nasdaq.com/), with time ranging from the start of the pandemic to recent (from **Jan 06, 2020** to **Feb 02, 2022**). The result may provide some insightful ideas to the policy makers, investors and everyone suffering during the pandemic to get closer view of the impact of the pandemic to the world.

The COVID-19 data set used in this report is from WHO Data Table on their website (https://covid19.who.int/info?openIndex=2), and the stock market data set is from the institutional customer service of a Chinese financial information company Wind (https://www.wind.com.cn/en/default.html). We combined the `NASDAQ index prices` and several stock market `control variables` with the `COVID-19 cases` variables as the data set used in the report. We generated a new dummy variable `fluctuation of daily newly reported cases` and added it to the data set. Since our question needed to be answered in this report is the influences of `fluctuation of daily newly reported cases` to the NASDAQ index, this new dummy variable works as a treatment in ANOVA model, or estimator in linear regression model. The explanation of the variables is as follows.

Variables	Explanation	Notation
date	Time stamp, from the beginning of the pandemic Jan 1, 2020 to Feb 7, 2022	t
index.p	NASDAQ index prices during the period	P_{NASDAQ}
cases	Newly report COVID-19 cases during the period	C
cases.chg	Regarded as treatment to the outcome, represents `fluctuation of daily newly reported cases`, a dummy variable by setting 1 when $cases_t - cases_{t-1} > 0$, and setting 0 when $cases_t - cases_{t-1} <= 0$.	F
VIX	CBOE Volatility index	V
gold	COMEX consecutive gold future prices	G
brentoil	Consedutive brent crude oil future prices	B

3 Background

The pandemic has brought lots of hurt to our life and society, while at the same time, the stock market seems flourishing after a short fluctuation due to the panic of COVID-19. Although, there are many factors have some impacts to the stock market, the COVID-19 pandemic undoubtedly influences the world deeper than others. It changes people's life, working style, transportation, etc. For sure, it brings a big impact to the economy. "The toll the COVID-19 pandemic has exacted on the global economy has been significant, with the International Monetary Fund (IMF) estimating that median global GDP dropped by 3.9% from 2019 to 2020, making it the worst economic downturn since the Great Depression. While the global economy was estimated to have recovered in 2021, recovery has been uneven and disparities in vaccine access and coverage could threaten improvement in much of the world." (https://www.kff.org/global-health-policy/issue-brief/economic-impact-of-covid-19-on-pepfar-countries/), reported by Stephanie Oum, Jennifer Kates, and Adam Wexler on KFF. As many countries are struggled to bolster the economy, the saving and wealth surge have different trend. "Household saving increased sharply during the COVID-19 crisis in many countries." (https://blogs.imf.org/2021/11/09/the-unequal-covid-saving-and-wealth-surge/), reported by By Cian Allen and Cyril Rebillard on IMFBlog. These news presents the changes due to the pandemic, not only to the economy, but also to people's awareness of savings and personal financial situation. Together with the news mentioned in the **Introduction** section, we are interested in the question of the pandemic impacts on the stock market. By the answer to this question, someone may obtain some insights or ideas to investment, life style changing, or even policy make, etc. In this paper, we refer NiallO'Donnell, DarrenShannon, and BarrySheehan's paper "Immune or at-risk? Stock markets and the significance of the COVID-19 pandemic". Their study on COVID-19 influences to the stock market give me the idea to build the model to answer our questions of this report.

4 Descriptive analysis

We started from some visualizations to the data to get a preview of their direct and indirect relationship. We found that during different period, some other uncertainties may impact the stock market severely which leads to departure of the trends of the two variables. To deal with such problem, we propose several control variables to reduce this influence to the results.

We first examined the time-series plot of cumulative COVID-19 cases and NASDAQ index prices by the scale of these two variables and got $Figure 4.1$.
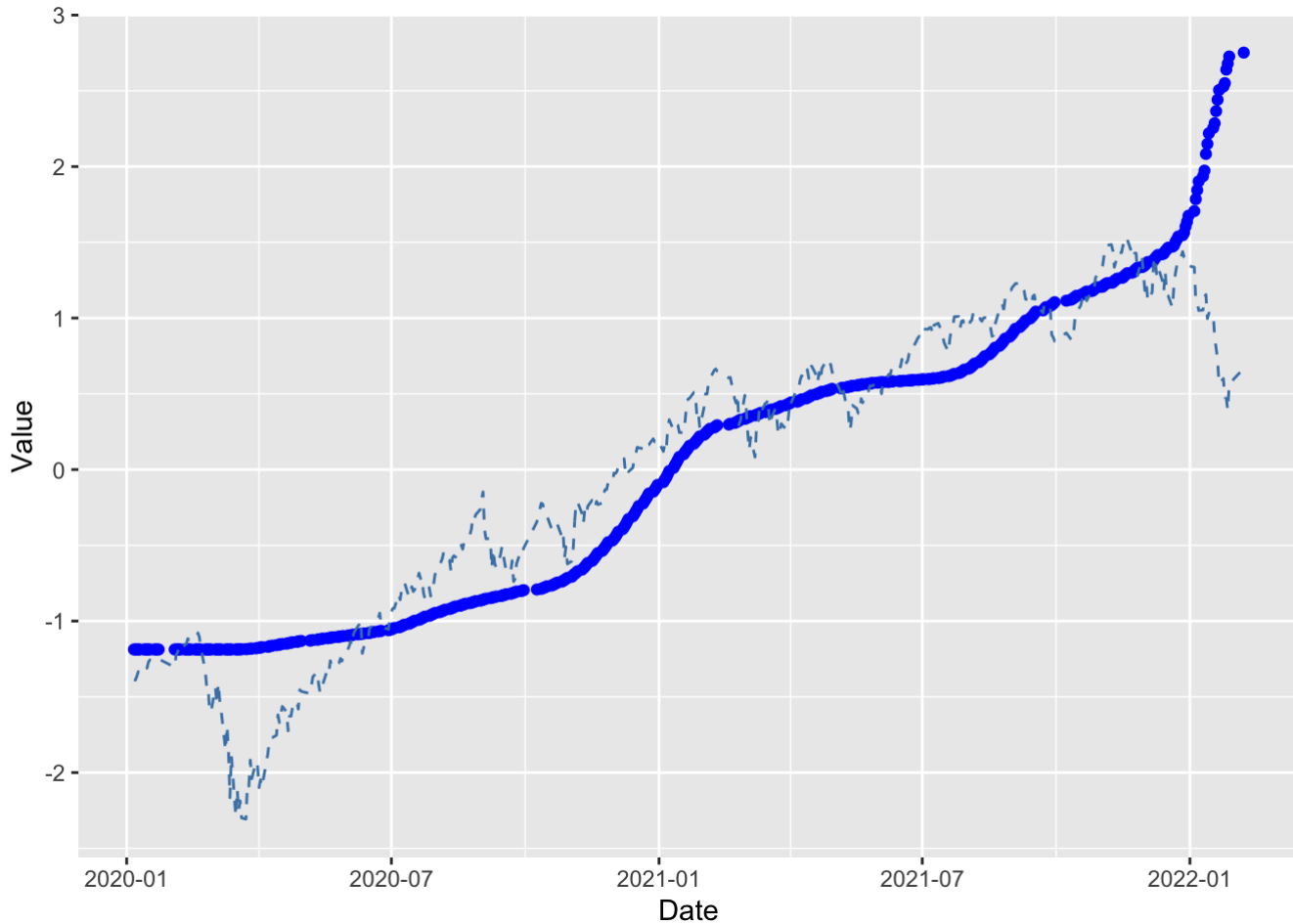

Figure 4.1

From the above plot, we notice that cumulative cases of COVID-19 and NASDAQ index prices have the same trend of increasing during the time series. Therefore, it attracted our attention to get a deeper insight of their relationship.

We first transform the `cases` to a dummy variable `cases.chg` by the following function:

$$F = \begin{cases} 1, & C_t - C_{t-1} > 0 \\ 0, & C_t - C_{t-1} \leq 0 \end{cases}$$

where F represents the fluctuation of daily newly reported COVID-19 cases `cases.chg`, C_t represents the number of newly reported COVID-19 cases on day t, C_{t-1} represents the number of newly reported COVID-19 cases on day t-1. By the dummy variable, we drew the time-series plot of NASDAQ index colored by `cases.chg` to have a deeper insight of two treatments' (i.e. $F = 1$ and $F = 0$) distribution.
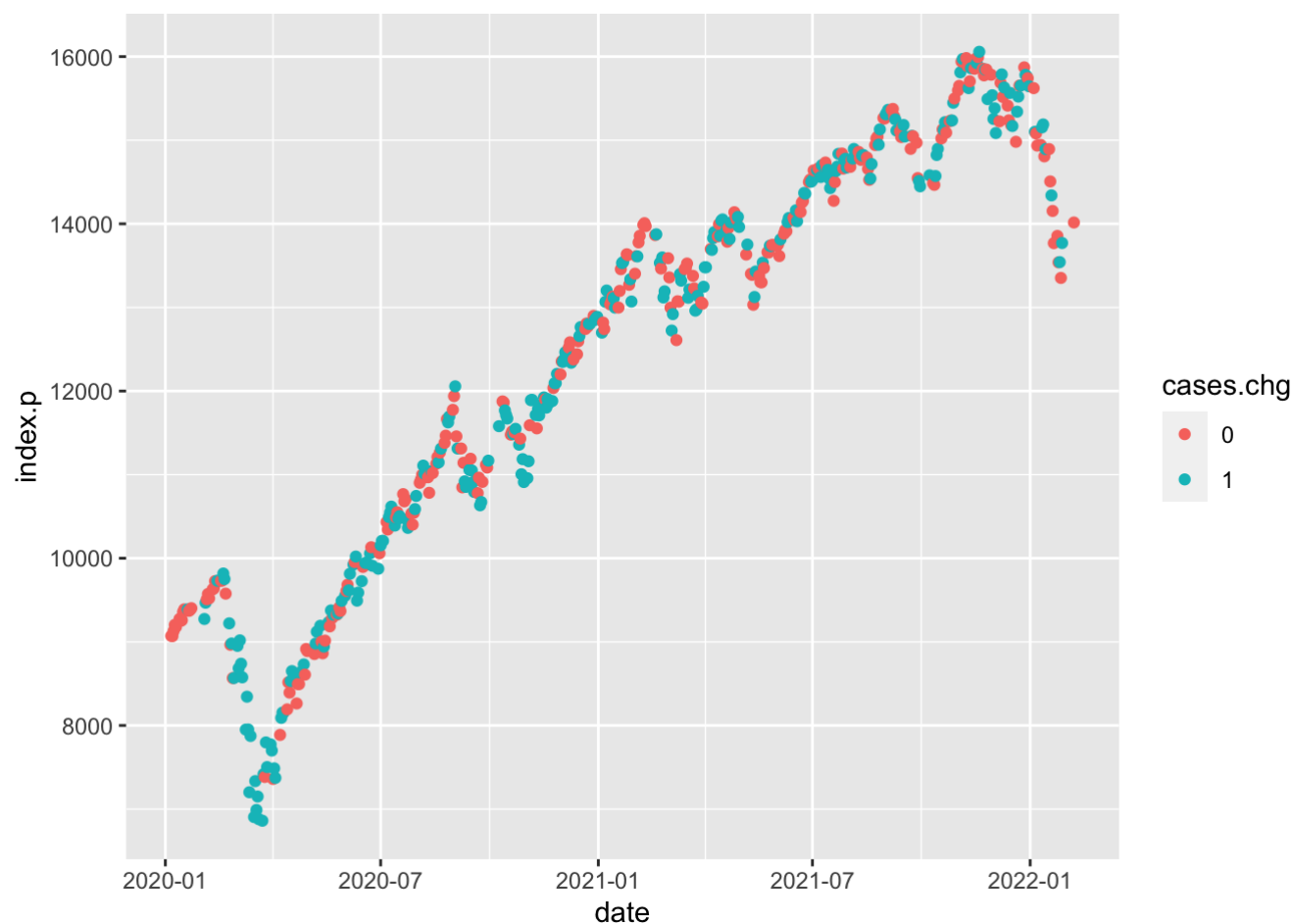
Figure 4.2

There is no obvious difference given different level of F in $Figure$ 4.2. To our intuitive, the stock prices should be small given $F = 1$. This plot indicates that there may be some influential factors to the index prices. We then obtained the main effect plot by using F as the categorical variable to see its effects on index prices.
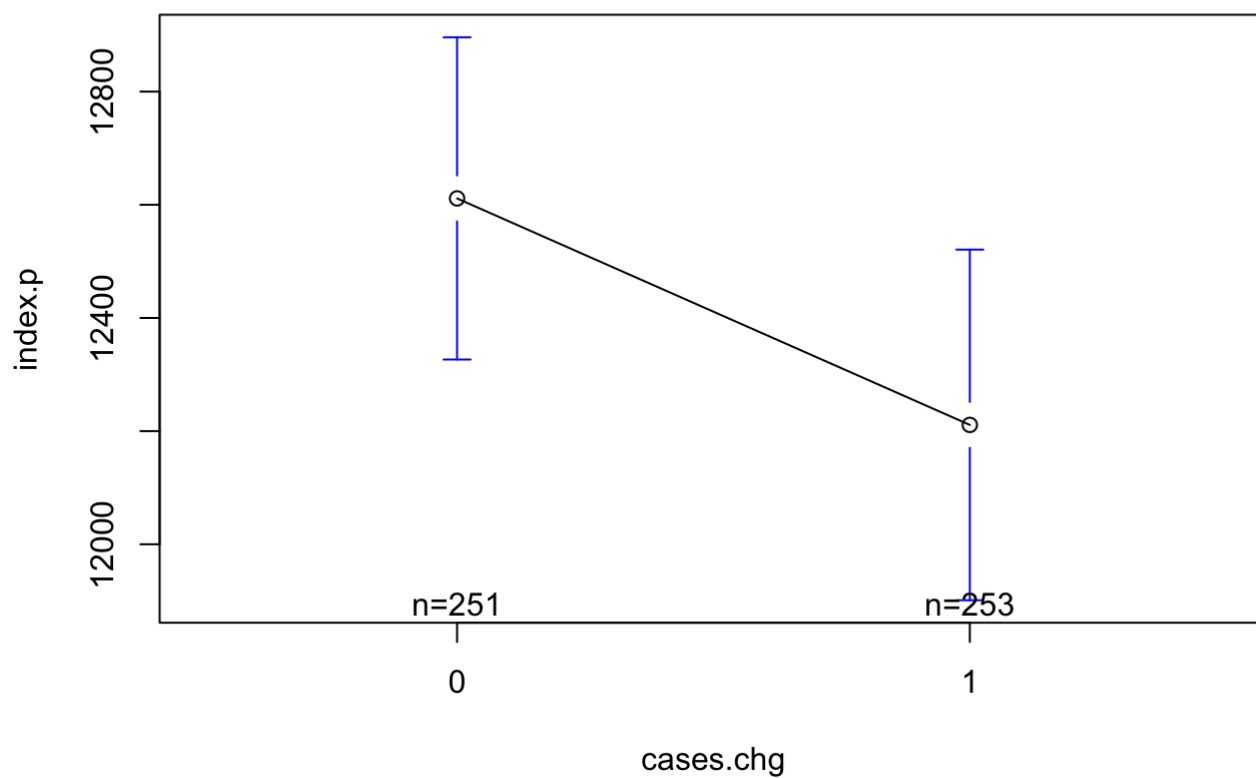
Figure 4.3

The main effect plot implies that an increasing of new daily cases has a lower index prices, which is the same with our intuitive. To obtain a further view of the data, we drew a side-by-side boxplot by using daily changes of index prices as outcome to compare the differences under different levels of F.
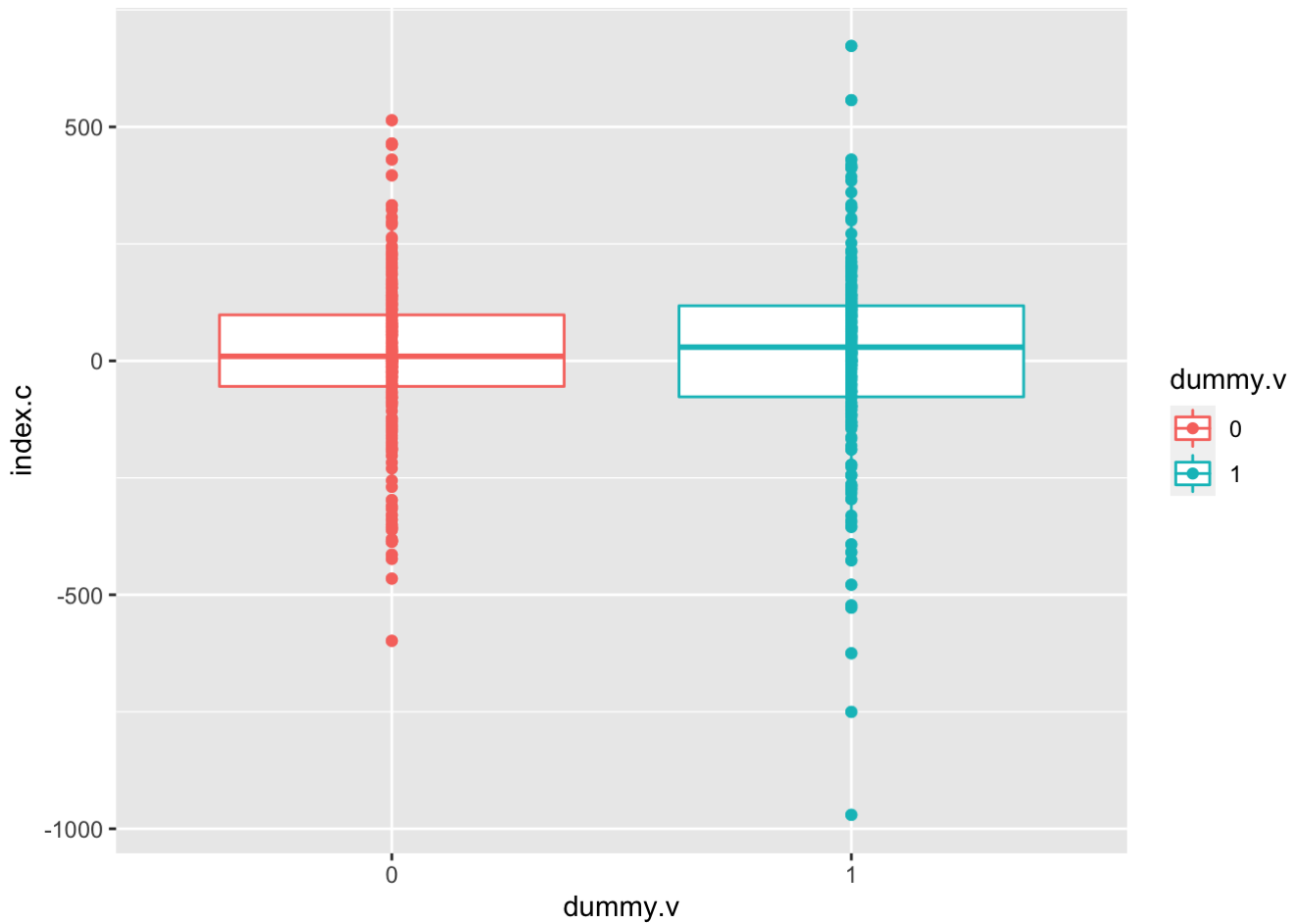
Figure 4.4

Although the main effect plot tells us that increasing in daily new cases may cause lower stock prices, the side-by-side boxplot presents that increasing in daily new cases may draw a higher median index prices daily changes and a larger range of distribution. This paradox induces us to a further study of the question.

Due to our intuitive and the result of the main effect plot, we could propose a preliminary guess of the answer to our questions:

- The increasing fluctuation of daily newly reported cases has negative effects on NASDAQ index.

To obtain the convincing answer to the questions we are interested in, we constructed several models and conducted tests to approach to the true answer.

5 Inferential analysis

Intuitively, we started with a simple model with only F as the input to the output `index.p` . By this model, we tested the effect of `fluctuation of daily newly reported cases` solely to the index prices. To construct such model, an ANOVA model fits best where F can be the treatment variable. The factor effect form of the model is as follows:

$$P_{NASDAQ} = \mu + \tau_i + \epsilon_i, \ i \in F = \{1, 0\}, \ \epsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{5.1}$$

where, P_{NASDAQ} represents the prices of NASDAQ index `index.p`, μ represents the mean of whole trials, τ_i represents the i-th effect given $i \in F = \{0,1\}$ where F is a dummy variable represents `fluctuation of daily newly reported cases`, ϵ_i represents the error term with 0 mean and equal variance.

The summary table of model (5.1) indicates that F is significant under 10% significant level, which indicates it has an association relation with the index and can explain the NASDAQ index.

```
##                Df    Sum Sq   Mean Sq F value Pr(>F)
## cases.chg       1 2.019e+07 20188739   3.507 0.0617 .
## Residuals     502 2.890e+09  5756273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 5.1

To get deeper in the differences between treatments (i.e. different `fluctuation of daily newly reported cases`), we obtained the Tukey-Karmer Confidence Intervals plot to test whether P_{NASDAQ} is different among the two treatments of `cases.chg`.
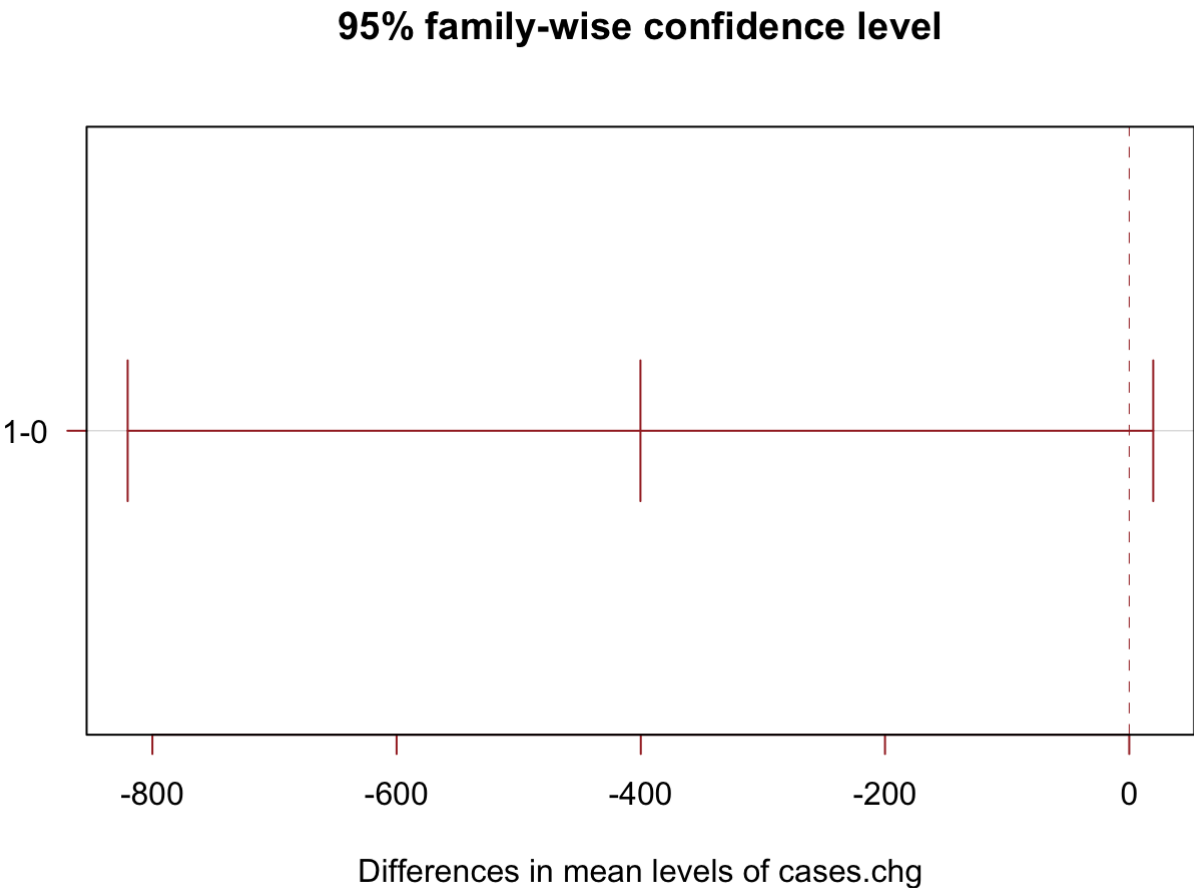
95% family-wise confidence level

Figure 5.1

The plot tells us that although there is a difference of index prices in these two treatments, but since the interval contains 0, they are not significantly different under 95% confidence level. The Tukey interval is as follows:

```
##               diff               lwr               upr            p adj
## -400.28815977  -820.22626868    19.64994914       0.06168146
```

However, it may not represent the causal relationship between them, because there are lots of potential confounding variables that are causally related to the outcome and correlated with the treatments. Since it's nearly impossible to design a randomization experiment to the stock market, we try to use some statistical methods to reduce the influence of confounding variables.

By the conclusion from the paper "Immune or at-risk? Stock markets and the significance of the COVID-19 pandemic", they use COVID-19 together with six additional variables as control variables to explain the index (Niall O, et al., 2021). They found only Volatility, Brent crude oil, Gold and COVID-19 variables are significant to explain the index in the US (Niall O, et al., 2021). These variables are corresponding with `VIX`, `brentoil`, `gold`, and `cases.chg`, respectively.

The selection of these variables are also not random. `VIX` represents the expectation of volatility in financial market, and it is suitable as a control variable since it is an ex-post measure of market activity, is considered a proxy for the current sentiment of investors, and has been shown to be significantly linked with market returns (Poshakwale et al., 2019).

`gold` can be described as a safe assets in the volatile market (Baur and Lucey, 2010), which implies it may be a significant factor to the market during the period of panic.

`brentoil` has a similar role as gold to the market, it is inverse (Sakaki, 2019). In light of the highly volatile global oil markets during March 2020 (Masson and Winter, 2020), the price of oil is a relevant and potentially significant factor in recent stock market prices (Niall O, et al., 2021). It's highly possible that it can explain the market turbulent fluctuation at the beginning of the pandemic and recent.

Since these variables do have causal relation to the index, and they are also correlated with the cases because of the wide-spread panic impact on the whole market, it's very convincing to use them as control variables in the model.

Because what happens in the stock market is too much complicated and it flows with time, we cannot conduct randomization to eliminate the confounding effects of obvious and potential confounders. Therefore, the problems comes to an observational study. By the notes of STA 207, we can use linear outcome models for the potential outcomes (Shizhe Chen, 2022), which is as follows:

$$Y_i = \alpha_0 + (\alpha_1 - \alpha_0)Z_i + X_i'\gamma_0 + Z_iX_i(\gamma_1 - \gamma_0) + \epsilon_i. \tag{5.2}$$

where Z_i represents the treatment, X_i represents the covariates, α_1 and α_0 represent the intercept in each causal effect model, γ_1 and γ_0 represent the parameter of covariates in each causal effect model.

For our problem in this report, we need to propose an important assumption beforehand:

- The covariates' influences to the potential outcome are independent with the treatment.

The reason for the above assumption is because we want to eliminate some confounding effects and keep the effects by treatment singly, it is natural to assume that other variables should be independent with the treatment. This is also the assumption for observational study. By the assumption, we can conclude that $\gamma_1 = \gamma_0$ in the above model. Therefore, we can remove the interactive term in the model.

By changing the variable in model (5.2) with our notation, the model should be written as:

$$P_{NASDAQ} = \beta_0 + \beta_1 F + \beta_2 V + \beta_3 G + \beta_4 B + \epsilon_i, \qquad (5.3)$$

$$i = 1, 0, \ \epsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2).$$

where F represents the dummy variable of `fluctuation of daily newly reported cases`, V represents the `CBOE Volatility index`, G represents the `COMEX consecutive gold future prices`, B represents the `Consedutive brent crude oil future prices`.

Model (5.3) is a linear model, where the parameter of F can be regarded as the effect of the treatment to the outcome. Therefore, the value of β_1 is what we should focus on. By using the function `anova` in **car** library, we constructed ANOVA table for this model, which is as follows.

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
cases.chg	1	20188739	20188739.0	25.37549	6.606361e-07
VIX	1	1233670342	1233670342.1	1550.61643	3.363192e-155
gold	1	162328635	162328635.2	204.03299	4.730016e-39
brentoil	1	1096645933	1096645932.9	1378.38866	1.106028e-145
Residuals	499	397004370	795599.9	NA	NA

5 rows

Table 5.2

The ANOVA table tells that `cases.chg` is significant to explain the NASDAQ index. In the meanwhile, the influences of confounding variables are reduced by plugging in the control variables. The added covariates are also significant in this model. The ANOVA table tells us that the variables included in model (5.3) are all siginificant and useful to explain the outcome. For a further study of the treatment effect, we need to construct the confidence interval for dummy variable $F's$ parameter to observe its distribution.

We need to check if the fluctuation of daily cases does have some effect to the market for further investigation. By the following confidence interval plot, we can obtain a clearer view of the relation.
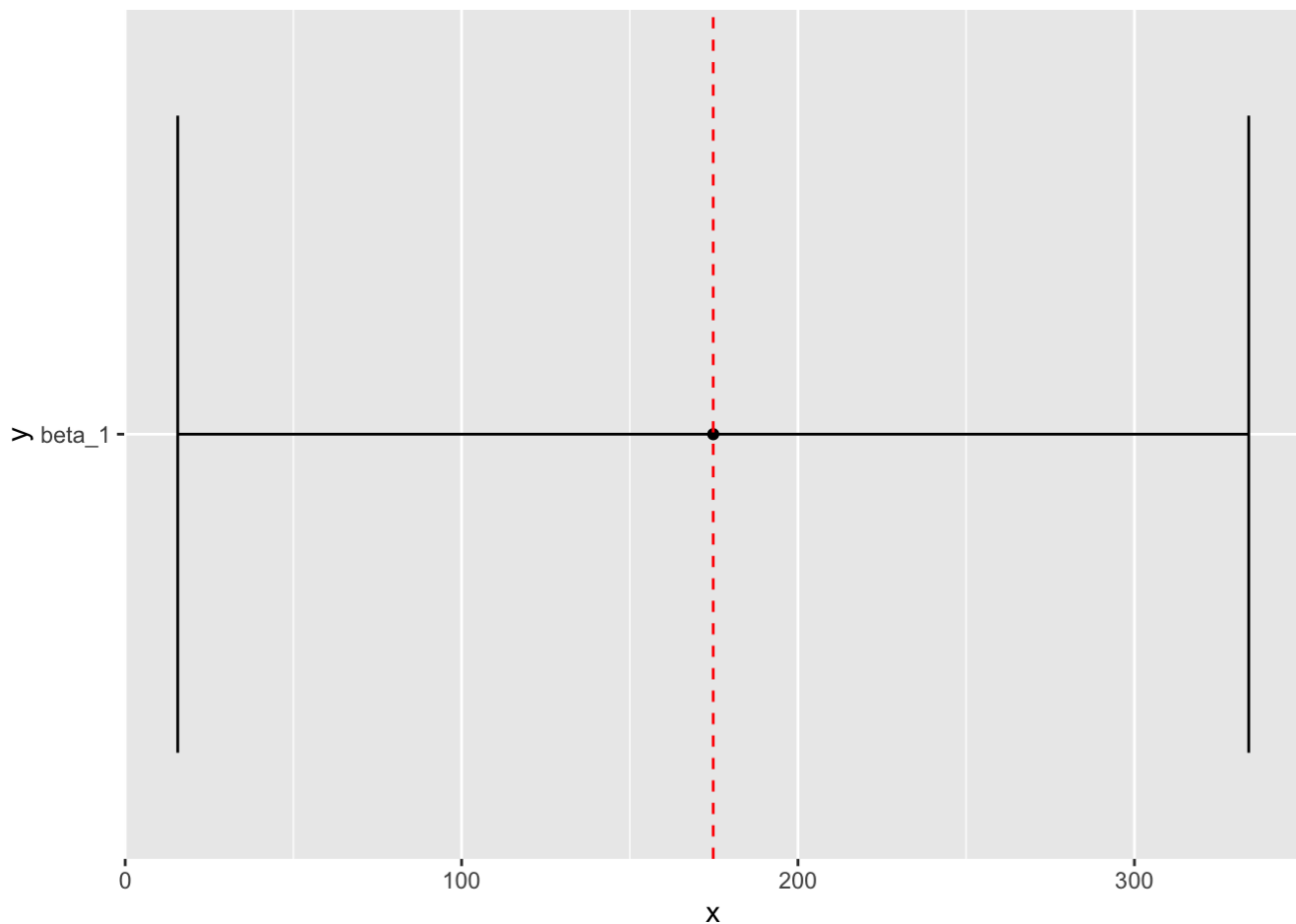
Figure 5.2

The 95% confidence intervals presents a confidence interval settled at the left of 0, which mean a positive effect of the dummy variable to the outcome exists. This result indicates that by removing some confounding influences, the increase of daily newly reported cases `cases.chg` (i.e. when $F = 1$) causes a higher index prices than the situation when daily newly reported cases decrease, which is contrary to our intuition and former results by the simple model.

The 95% confidence interval is as follows:

```
##                  2.5 %  97.5 %
## cases.chg1 15.6207 333.984
```

As most people regard stock market as a time series, we think that the lag feature of a time series may have some influence to our model. Therefore, we propose a transformation to the outcome to deal with the lag problem. Before the transformation, we need to conduct several test to check if autocorrelation exist in the index prices, and try to decide which autoregressive model should be used by ACF and PACF plot.

By the Durbin Watson test, we conclude that there is autocorrelation in the residuals of model (5.3). Test result is as follows.

```
##   lag Autocorrelation D-W Statistic p-value
##    1          0.9357    0.09352446       0
##  Alternative hypothesis: rho != 0
```

where p-value is smaller than 0.05, which means we are 95% confident that autocorrelation exists in the residuals.

Now, we need to decide which model is better to express the lagged pattern of index prices. By checking the ACF and PACF plot, we can make such decision.
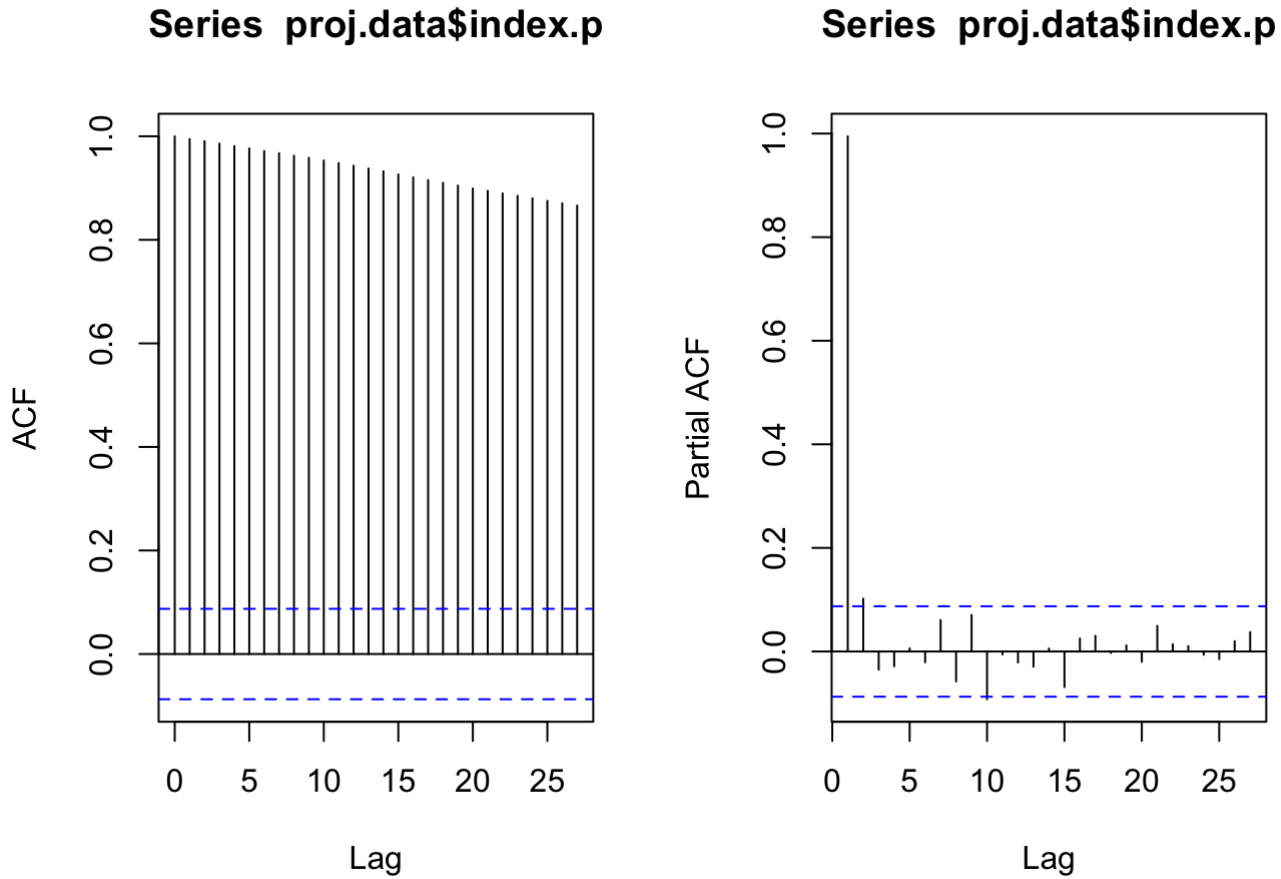

Figure 5.3

The ACF plot shows a slightly decrease pattern, while the PACF plot presents two spikes exceed the blue dashed confidence interval horizontal lines. It implies that AR(2) model is best fit to capture the time-series pattern of the data. The AR(2) model can be written as follows:

$$P^{time}_{NASDAQ_t} = c + \alpha_1 P^{time}_{NASDAQ_{t-1}} + \alpha_2 P^{time}_{NASDAQ_{t-2}} + \epsilon_t \tag{5.4}$$

where $P^{time}_{NASDAQ_t}$ represents the index prices at time t, $P^{time}_{NASDAQ_{t-1}}$ represents the index prices at time t-1, α_1 and α_2 are parameter of corresponding previous values, ϵ_t represents white noise.

In this way, we can capture the time-series influences by the model above. Then, we substituted the outcome in model (5.3) by the time-series model fitted values to eliminate the time influences to our model, and constructed model (5.5) as follows:

$$P^{time}_{NASDAQ} = \beta_0 + \beta_1 F + \beta_2 V + \beta_3 G + \beta_4 B + \epsilon_i, \tag{5.5}$$

$$i = 1, 0, \ \epsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2).$$

The only difference between model (5.3) and model (5.5) is the changting the outcome by time-series fitted outcome.

The plot of comparison between time series fitted values and true values is as follows, which has an obvious lag pattern.
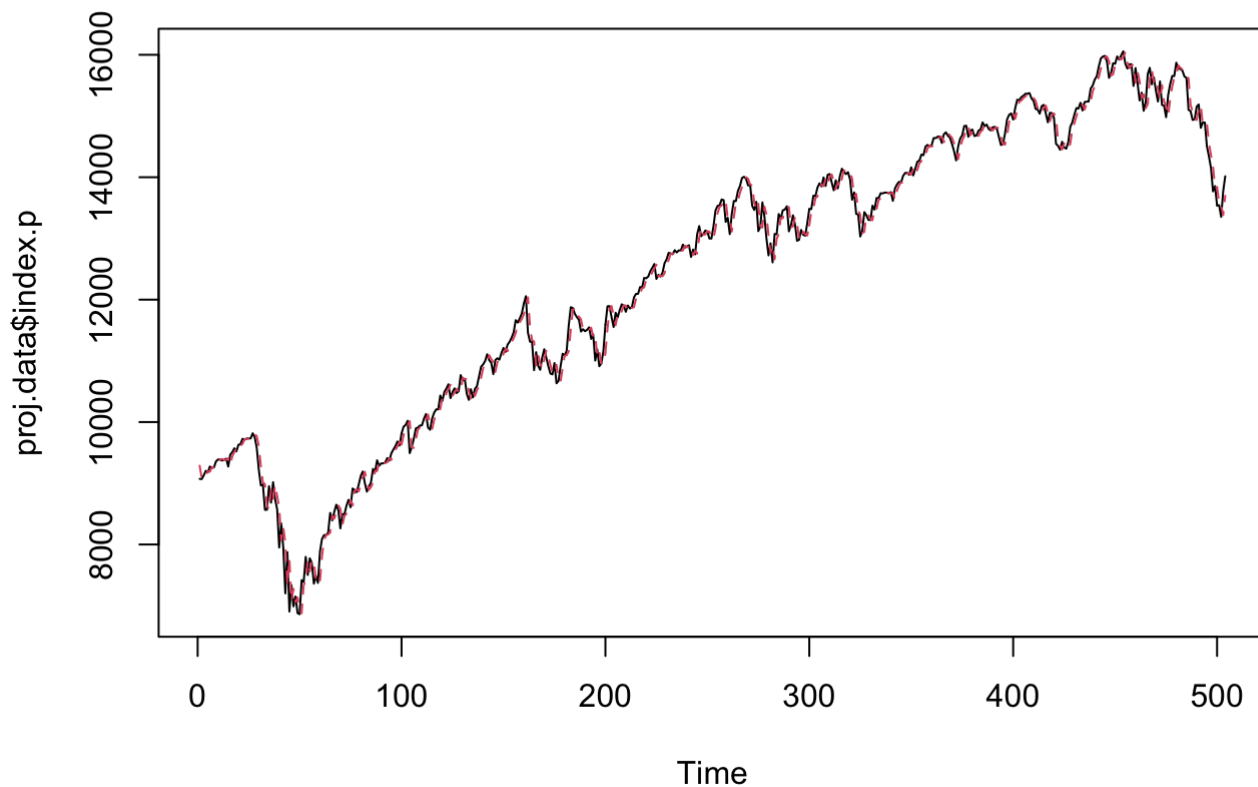

Figure 5.4

By fitting model (5.5), we constructed the ANOVA table of model (5.5), which implies that the fluctuation of daily cases is significant to explain the outcomes, which is the same as model (5.3). The added covariates are also significant. In the meanwhile, the influences of confounding variables are reduced by plugging in the control variables. The table is shown in *Table* 5.2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
cases.chg	1	19947562	19947562.3	25.20766	7.176670e-07
VIX	1	1179360355	1179360354.6	1490.35349	5.788230e-152
gold	1	159569876	159569876.4	201.64789	1.108889e-38
brentoil	1	1149240926	1149240926.3	1452.29168	7.197941e-150
Residuals	499	394873311	791329.3	NA	NA
5 rows					

Table 5.3

For checking if there is impact of `fluctuation of daily newly reported cases` on index prices, we need to construct the confidence interval of dummy variable's parameter and see if the interval covers 0. We present the confidence intervals of model (5.3) and model (5.5) for a clearer view of the differences between these two models.
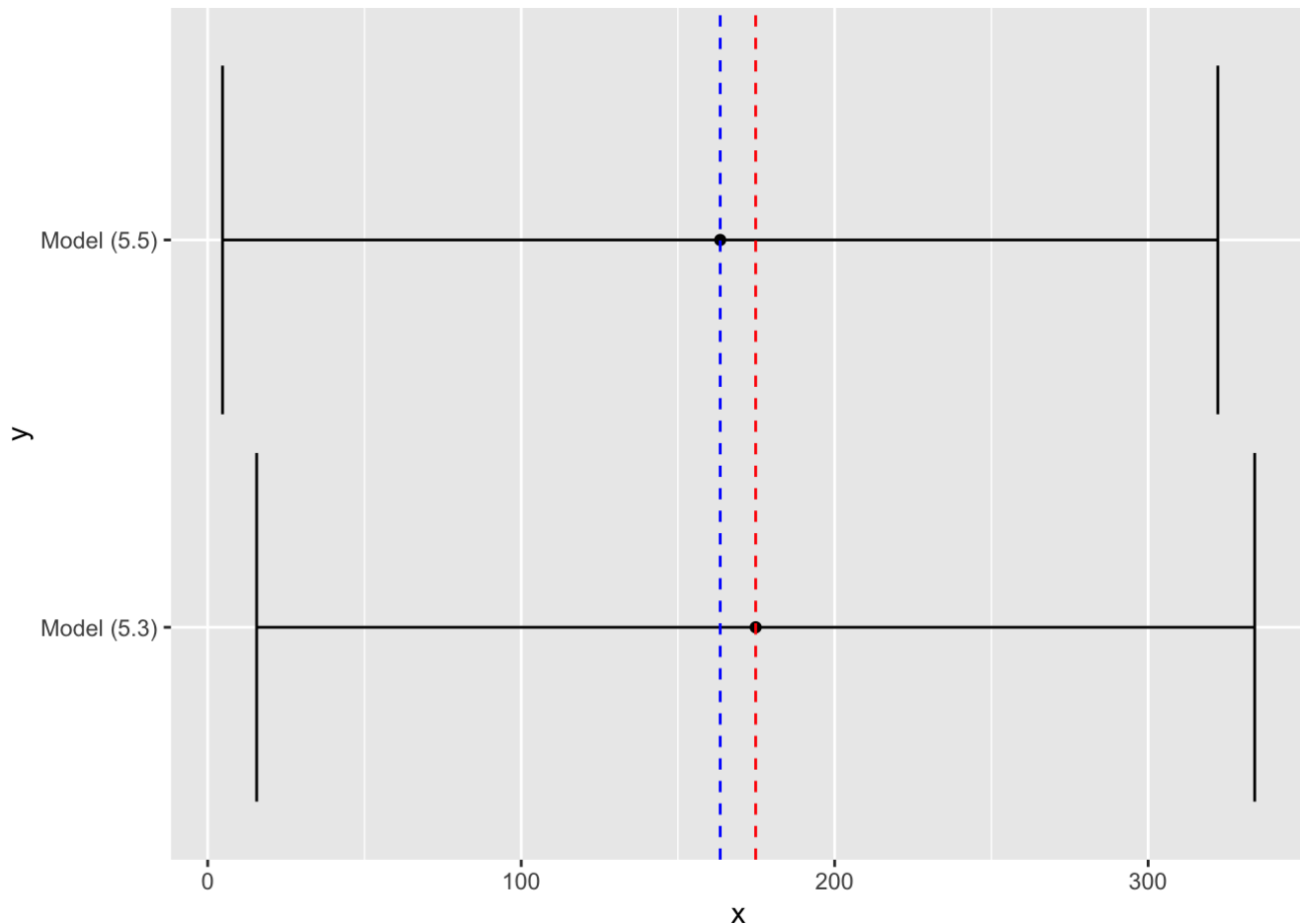

Figure 5.5

The blue dashed line is the estimated value of the parameter of dummy variable by model 5.5, and the red dashed line is the estimated value of the parameter of dummy variable by model 5.3. The confidence intervals plot tells us that model (5.5) has a smaller interval (i.e. closer to 0) than model (5.3), but its lower bound is still bigger than 0. Although the results stay the same, we found that our transformation to the outcome variable has some influences. Since the confidence intervals do not cover 0, we can conclude that the effect of `fluctuation of daily newly reported cases` on index prices is significant. When daily newly reported cases increases, the stock market index prices tends to be higher too.

The 95% confidence interval is as follows:

```
##                    2.5 %     97.5 %
## cases.chg1 4.731162 322.2388
```

By what we have discussed above, we believe that model (5.5) is the best model we have got for answering the questions we are interested in, and we will conduct model diagnostics to it in the following section.

6 Model diagnostics

We proposed several assumptions for the model (5.5). In this section, we will diagnose if these assumptions hold. The diagnostics is conducted in two parts: first, equal variance diagnosis; second, normality diagnosis.

6.1 Equal variance diagnosis

We use the function `leveneTest` in R to conduct Levene Test for testing if the assumption of equal variance hold. The null hypothesis of Levene Test is "The variance among groups is equal", while the alternative hypothesis is "The variance among groups is *not* equal". The result of the test is as follows.

	Df <int>	F value <dbl>	Pr(>F) <dbl>
group	1	2.687703	0.101752
	502	*NA*	*NA*
2 rows			

Table 6.1.1

We notice that the p-value of the test is around 0.1, which means under 0.05 significant level, we cannot reject the null hypothesis. Therefore, the equal variance assumption holds for the model (5.5).

6.2 Normality diagnostics

To test normality assumption, we decided to use visualization methods: Q-Q plot and histogram of the residuals, which are presented as follows.
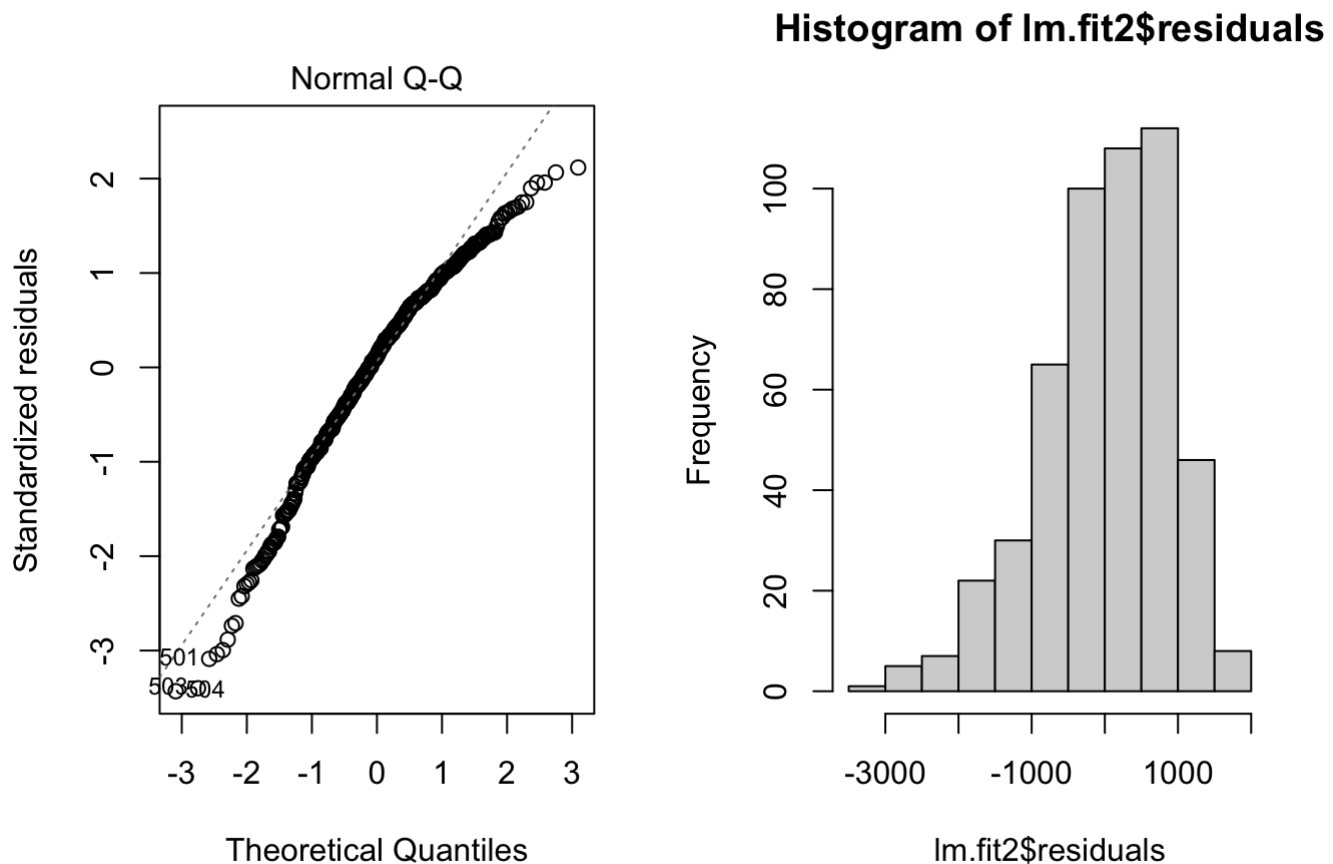
Histogram of lm.fit2$residuals

Normal Q-Q

Figure 6.2.1

Although the Q_Q plot has a slightly left skewed pattern, we could still say that the normality assumption holds since the skewed happens at the tail of the distribution, which may be caused by outliers or leverage points. The histograms also presents a slightly left skewed pattern. We could say that the normality assumption holds. However, we still need to check if we can adjust the Q-Q plot to be more "normal" by some transformation to the model.

6.3 Remedies diagnostics

We use `boxcox` function in `MASS` library in R to check if we need to transform the model (5.5) in some ways.
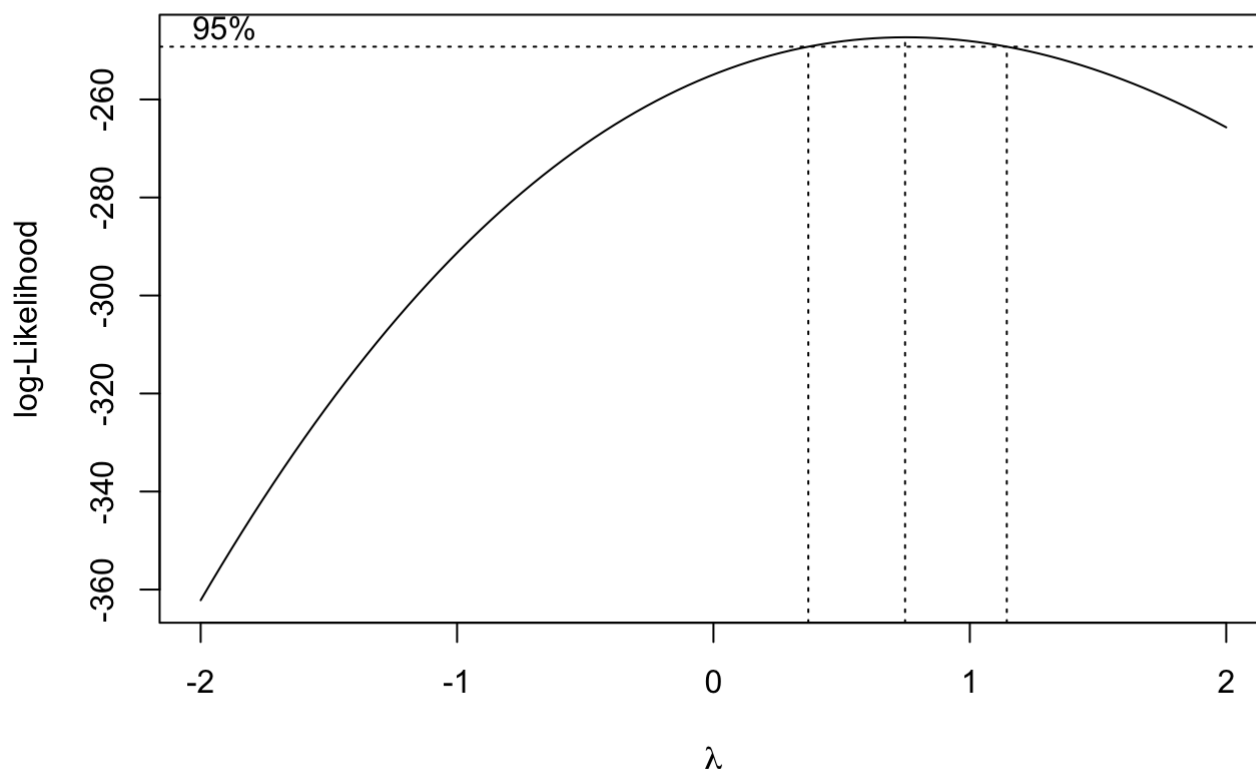
Figure 6.3.1

The plot shows that under 95% confidence level, the `boxcox` function suggests that no transformation is needed. Therefore, we conclude our model fulfill the assumptions we made and is valid.

7 Discussion

As what we have discussed in this paper, we make the conclusion that answers our interested questions. We conclude that the increase of daily newly reported cases has a positive causal effect to the NASDAQ index after controlling several potential confounding variables. This phenomenon contradicts with our commonsense and intuitive. Nevertheless, the model tells us that all variables are significant and all the assumptions of the model hold. The diagnostics also suggests no transformation to the model is needed. It implies that our model is valid. Therefore, we draw the conclusion from our study. It is not easy to explain why this phenomenon happens. We may say that the stock market is too complicated so that there may be other correlated factors together with COVID-19 data affect the market. We can also say that our assumptions for building the causal model (5.3) and (5.5) are too strong that violates the real world rules. In addition, there are some **caveats** of our model.

- The covariates added to our model may be not good enough as the control variable to present the causal relation between `fluctuation of daily newly reported cases` and index prices.

- The outcome we used in the model is index prices which are highly autocorrelated within the data. Probably, logarithm returns may be a better choice. Because, it can not only reflects the prices changes, but also get rid of the autocorrelated issues.

- Our model is only valid under our assumptions to the model. The real world problems are much more complicated. We started this study for some heuristic purpose. (e.g. The covariates used in our study probably can not explain the stock market well.)

For future study, we could focus on discover more useful covariates and confounders of such problem, such as periodic pattern of the COVID-19 cases, other factors affects the stock market, instrumental variables, etc.

The reason for the study in this report is to provide some insights of the influences that the pandemic takes to our world. The conclusion presents an unexpected result, that is "more new cases generates higher index prices". It will lead to many more problems to not only the financial market, but also everyone's daily life. The direct harm caused by pandemic may be easy to visualize, while the great changes under the surface are profound.

Acknowledgement

I acknowledge that some of the methods and codes are sourced from the notes of course STA 207 provided by Prof. Shizhe Chen.

Thanks for the help of Prof. Shizhe Chen for his interpretation of the methods used in this paper.

Thanks for the help of our TA Zitong Zhang. She provides very useful materials for the report and interpretation of some confusing topics.

Thanks for the help of my teammates. Some of the ideas of dealing with the problems in this paper come from my teammates Yulu Jin, Kaiming Fu and Xue Sheng.

Reference

Niall O'Donnell, Darren Shannon, Barry Sheehan (2021). Immune or at-risk? Stock markets and the significance of the COVID-19 pandemic. Journal of Behavioral and Experimental Finance, Volume 30, ISSN 2214-6350.

Poshakwale S.S., Chandorkar P., Agarwal V. (2019). Implied volatility and the cross section of stock returns in the UK. Res. Int. Bus. Finance, 48, pp. 271-286

Baur D.G., Lucey B.M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold Financ. Rev., 45 (2), pp. 217-229

Sakaki H. (2019). Oil price shocks and the equity market: Evidence for the s&p 500 sectoral indices Res. Int. Bus. Finance, 49, pp. 137-155

Shizhe Chen. (2022). 9.4.3 Elementary strategies. Chapter 9 Causal Inference, Notes, StatDataScience.

Session info

```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Big Sur 11.6
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/lib/libRblas.0.d
ylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/lib/libRlapack.d
ylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] MASS_7.3-54   car_3.0-12    carData_3.0-5 gplots_3.1.1  zoo_1.8-9
## [6] ggplot2_3.3.5
##
## loaded via a namespace (and not attached):
##  [1] highr_0.9          pillar_1.6.4       compiler_4.1.1     jquerylib_0.1.4
##  [5] bitops_1.0-7       tools_4.1.1        digest_0.6.29      jsonlite_1.7.2
##  [9] lattice_0.20-44    evaluate_0.14      lifecycle_1.0.1    tibble_3.1.6
## [13] gtable_0.3.0       pkgconfig_2.0.3    rlang_0.4.12       DBI_1.1.1
## [17] yaml_2.2.1         xfun_0.29          fastmap_1.1.0      withr_2.4.3
## [21] stringr_1.4.0      dplyr_1.0.7        knitr_1.37         caTools_1.18.2
## [25] gtools_3.9.2       generics_0.1.1     vctrs_0.3.8        grid_4.1.1
## [29] tidyselect_1.1.1   glue_1.6.0         R6_2.5.1           fansi_1.0.2
## [33] rmarkdown_2.11     farver_2.1.0       purrr_0.3.4        magrittr_2.0.1
## [37] scales_1.1.1       ellipsis_0.3.2     htmltools_0.5.2    abind_1.4-5
## [41] assertthat_0.2.1   colorspace_2.0-2   labeling_0.4.2     KernSmooth_2.23-20
## [45] utf8_1.2.2         stringi_1.7.6      munsell_0.5.0      crayon_1.4.2
```

Code Appendix

```r
knitr::opts_chunk$set(fig.pos = 'H')
library(ggplot2)
library(zoo)
library(gplots)
library(car)
library(MASS)

# Data preparation
# Read and combine different data sets
# COVID-19 data
whole.data <- read.csv('WHO-COVID-19-global-data.csv', header = T)
US.data <- whole.data[whole.data$Country=='United States of America',]

# Stock market data
nasdaq.data <- read.csv('Data-index.csv', header = F)[,1:2]
VIX.data <- read.csv('Data-volatility.csv', header = F)
gold.data <- read.csv('Data-gold.csv', header = F)
brentoil.data <- read.csv('Data-brent.csv', header = F)

# Data combination and transformation
time.ind.covid <- US.data$Date_reported %in% nasdaq.data$V1
US.new <- US.data[time.ind.covid,]
time.ind.market <- nasdaq.data$V1 %in% US.new$Date_reported

# Transform daily new cases changes into dummy variable (named as 'daily new cases fluct
uation')
obs.length <- nrow(US.new)
cases.change <- US.new$New_cases[2:obs.length]-US.new$New_cases[1:obs.length-1]
cases.change[cases.change>0] <- 1
cases.change[cases.change<=0] <- 0

# Aggregate the data into one single data set
A.data <- data.frame(date = as.Date(US.new$Date_reported)[-1],
                     index.p = as.numeric(nasdaq.data$V2[time.ind.market])[-1],
                     cases = US.new$New_cases[-1],
                     cases.chg = as.factor(cases.change),
                     VIX = as.numeric(VIX.data$V2[time.ind.market])[-1],
                     gold = as.numeric(gold.data$V2[time.ind.market])[-1],
                     brentoil = as.numeric(brentoil.data$V2[time.ind.market])[-1])

# Complement the NA's with most previous values
A.data <- na.locf(A.data, fromLast = T)

# Detect the potential relation between COVID-19 cases and NASDAQ index prices
ggplot(mapping = aes(x = date), data = A.data) +
  geom_point(aes(y = scale(cumsum(cases))), col = 'blue') +
  geom_line(aes(y = scale(index.p)), col = 'steelblue', linetype = 'dashed') +
  labs(x = 'Date',
       y = 'Value'
    ,color = 'Legend') +
  scale_color_manual(values = colors)
```

```r
# Remove the uncorrelated risk period, when the news of "Fed officials discussed raising
rates sooner and faster in 2022" was released.

proj.data <- A.data

# Data visualization
# Preliminary view of NASDAQ index prices distribution colored by daily new cases fluctu
ation
ggplot(aes(x = date), data = proj.data) +
  geom_point(aes(y = index.p, col = cases.chg))

# Glance of effects from simple model
plotmeans(index.p~cases.chg, data = proj.data)

index.chg <- proj.data$index.p[2:504]-proj.data$index.p[1:503]
F.val <- proj.data$cases.chg[2:504]
boxp.data <- data.frame(index.c = index.chg,
                        dummy.v = F.val)
ggplot(data = boxp.data) +
  geom_boxplot(aes(x = dummy.v, y = index.c, col = dummy.v)) +
  geom_point(aes(x = dummy.v, y = index.c, col = dummy.v))
# Inferential analysis
alpha <- 0.05

# Simple model
# Fit an ANOVA model with cases changes variable
anova.fit1 <- aov(index.p~cases.chg, data = proj.data)
summary(anova.fit1)

# Tukey plot to test the difference between populations
plot(TukeyHSD(anova.fit1, conf.level = 1-0.05), las=1 , col="brown")
# Tukey intervals
idx <- proj.data$cases.chg
means.comb <- tapply(proj.data$index.p, INDEX=idx,mean)
T.ci1 <- TukeyHSD(anova.fit1,conf.level = 1-alpha)
(contrast <- T.ci1$cases.chg['1-0',])
# Model with explanatory variables to stock market index prices
# Fit an lm model with cases changes variable and other explanatory variables of index p
rices
lm.fit1 <- lm(index.p~cases.chg+VIX+gold+brentoil, data = proj.data)

# Construct an ANOVA table
anova(lm.fit1)
# Generate a 95% confidence interval for the parameter of dummy variable F
# We can check if fluctuation fluctuation of daily newly reported cases has what kind of
influences to the index prices.
conf.int <- data.frame(x = lm.fit1$coefficients[2],
                       lower = confint(lm.fit1, 2, 1-alpha)[1],
                       upper = confint(lm.fit1, 2, 1-alpha)[2],
                       y = 'beta_1')
ggplot(data = conf.int, aes(x,y)) +
```

```r
  geom_point() +
  geom_errorbar(aes(xmin = lower, xmax = upper)) +
  geom_vline(xintercept = lm.fit1$coefficients[2], linetype = 'dashed', col = 'red')
# 95% confidence interval
confint(lm.fit1, 2, 1-alpha)
durbinWatsonTest(lm.fit1)
# Autocorrelation plots
par(mfrow=c(1,2))
acf(proj.data$index.p)
pacf(proj.data$index.p)
par(mfrow=c(1,1))
AR <- arima(proj.data$index.p, order = c(2,0,0))
ts.plot(proj.data$index.p)
AR_fit <- proj.data$index.p - residuals(AR)
points(AR_fit, type = "l", col = 2, lty = 2)
lm.fit2 <- lm(AR_fit~cases.chg+VIX+gold+brentoil, data = proj.data)
anova(lm.fit2)
# Generate two models' 95% confidence intervals for the parameter of dummy variable F
# We can check if fluctuation fluctuation of daily newly reported cases has what kind of
influences to the index prices.
conf.int.both <- data.frame(x = c(lm.fit1$coefficients[2], lm.fit2$coefficients[2]),
                     lower = c(confint(lm.fit1, 2, 1-alpha)[1], confint(lm.fit2, 2, 1-
alpha)[1]),
                     upper = c(confint(lm.fit1, 2, 1-alpha)[2], confint(lm.fit2, 2, 1-
alpha)[2]),
                     y = c('Model (5.3)','Model (5.5)'))
ggplot(data = conf.int.both, aes(x,y)) +
  geom_point() +
  geom_errorbar(aes(xmin = lower, xmax = upper)) +
  geom_vline(xintercept = lm.fit1$coefficients[2], linetype = 'dashed', col = 'red') +
  geom_vline(xintercept = lm.fit2$coefficients[2], linetype = 'dashed', col = 'blue')
# 95% confidence intervals
confint(lm.fit2, 2, 1-alpha)
# Model diagnosis
# Equal variance (not hold)
proj.data$res_abs <- abs(lm.fit2$residuals)
leveneTest(res_abs~cases.chg, data = proj.data)
# Normal distributed residuals
par(mfrow=c(1,2))
plot(lm.fit2, which = 2)
hist(lm.fit2$residuals)
par(mfrow=c(1,1))
boxcox(lm.fit2)
sessionInfo()
```