# The Reaction of Stock Market Indexes to COVID-19 Infections and Its Variations Across Industries

by

Angelina (Ziyi) Shen

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Business and Economics Honors Program

NYU Shanghai

May 2021

Professor Marti G. Subrahmanyam
Professor Christina Wang
Professor Wendy Jin

Faculty Advisers

Professor Xin Zhou

Thesis Adviser

## Abstract

Unexpected changes in COVID-19 infection cases are one of the key predictors of U.S. and China's highly volatile industry index returns since the start of 2020. Its predictive ability is especially valid under times when number of confirmed cases was growing exponentially, and when COVID-related discussions dominated the market. The paper aims to better understand the rationale of investors using epidemic predictive models to update their beliefs about the market, and to provide a more concrete forecast of the financial consequences of the outbreak in real time. One key thing to note is that our discussion is most accurate during the initial phase of the outbreak. As investors get more accustomed to COVID news, the validity of our model gradually decreases. This timeframe varies between countries, for example, our model has more predicting power during Jan. 22$^{nd}$ to May. 21$^{st}$ for the Chinese market, counting trading days only. The paper is heavily driven by data, concluding on the effect of massive geopolitical events starting from the COVID-19 pandemic, paying specific attention to the asymmetry of impact across industries. Upon finalizing on the predictive model used, and finish performing the robustness check of our model, we intuitively justify such differences from a government policy and public sentiment perspective to shed more light on market behavior and investor decisions.

## Keywords:

COVID-19, epidemic outbreak predictive models, multivariate regression models, industry index returns, volatility

**Acknowledgement**

First of all, I would like to give special thanks to my thesis advisor Dr. Xin Zhou for all his help to guide me through the entire research process. Without him, it would be impossible for me to conclude on such valuable insights. Additionally, the paper is an extension of Professor Robert Engle's most recent paper on "Multi-regime Forecasting Model for the Impact of COVID-19 Pandemic on Volatility in Global Equity Markets". Special thanks to Professor Zhou for introducing Professor Engle's insightful paper to me and guiding my research.

Next, I would like to thank all the faculty advisers, Professor Marti G. Subrahmanyam, Professor Christina Wang, and Professor Wendy Jin. Thank you all for coordinating the entire program so well and closely following up on each student's work, providing useful feedbacks.

I would also like to thank our teaching assistant Xinyi Yang for her kind support and guidance.

Finally, I would like to thank my fellow classmates in the Business Honors Program. I learnt so many new areas and fields from you and you all broadened my horizon.

# 1. Data Collection

In this section we discuss how the data was obtained, outline how infectious disease outbreaks can be modeled in real time, and how financial data is analyzed.

## 1.1 COVID-19 Data

Our COVID-19 data for the United States is crawled from the Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The Repository provides data on number of confirmed cases and number of confirmed deaths, updated once per day on country-level data as well as state-level/ county-level data. Both the confirmed cases and confirmed deaths are excluding probable COVID-19 cases and deaths, as these data are not reported officially. In order to select the time period to study for the US market, we visualize how the number of cases increase throughout. Based on the graph below, we finalize on periods March 18th, 2020 to Feb 23rd, 2021, with a total of 302 days excluding weekends and public holidays.

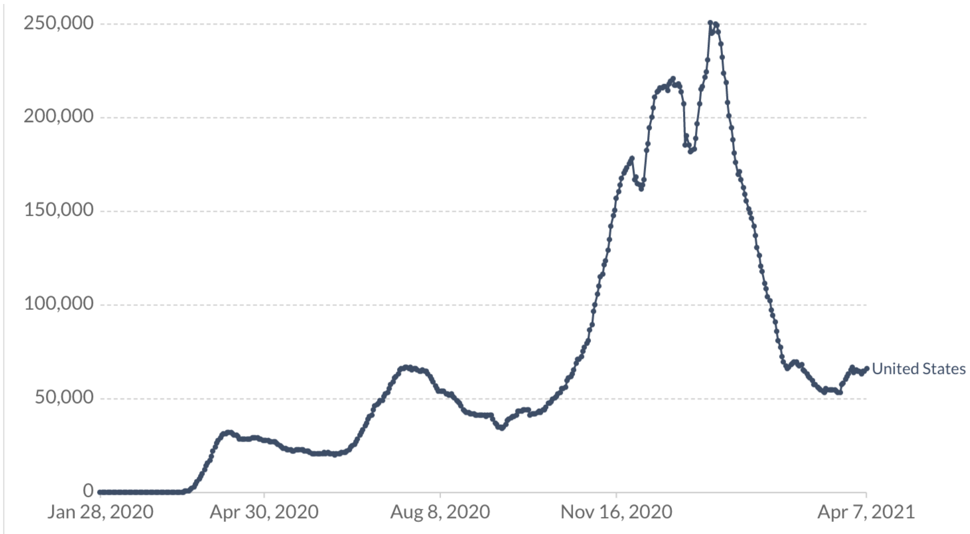


Figure 1: Number of Confirmed Cases- United States

The COVID-19 data for China is cited on province level and obtained from China CDC (CCDC) at http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm. The data is highly accurate as the data provider is the National Health Commission in coordination with local governments. Data is updated on 8am in the morning the next day, summarizing data including new cases, deaths and suspected cases. Though sources of each case are reported, for example 10 cases imported from the United States on December 13th, 2020, we will not be considering the imported cases separately from domestic ones. Similar to the US market, the time period paper selects for the Chinese market is from Jan 1st, 2020 to May 21st, 2020 with a total of 75 trading days, which includes the initial bursting period as well as the most volatilely affected period.

For the time being, we are aggregating to the country level and only using the confirmed cases column. Further research can be conducted on how state-level confirmed cases can impact the local stock exchanges and market behaviors. For example, how cases confirmed in Shanghai impact the volatility of the CSI 300 Index.

**1.2 Financial Data**

Financial data in China is obtained from Tonghuashun iFinD terminal, where we select 1-year treasury bond yield, CSI 300 Index and SSE Composite Index as benchmark for the Chinese market. We select industry as defined by the China Industry Classification system, querying for industry index data based on 28 segments of ShenWan Level One industries. For the above financial instruments, we obtain day-to-day return time series data. Financial data in the US is obtained from Yahoo finance and the Kaggle database, where we select returns similar to that in the Chinese market except for the fact that the number of days is significantly longer.

Additionally, we obtain control variables of the Chinese market from FactorWar BetaPlus team (https://www.factorwar.com/data/factor-models/). More specifically, we selected factors from the Fama-French Five Factor Model, incorporating risk-free rate, market return, small minus big SMB, high minus low HML, robust minus weak RMW and conservative minus aggressive CMA:

$$HML = 1/2\ (S/H + B/H) – 1/2\ (S/L + B/L)$$

$$RMW = 1/2\ (S/R + B/R) – 1/2\ (S/W + B/W)$$

$$CMA = 1/2\ (S/C + B/C) – 1/2\ (S/A + B/A)$$

$$SMB = 1/3\ (SMBBM + SMBROE + SMBINV)$$

Where

$$SMBBM = 1/3\ (S/H + S/M + S/L) – 1/3\ (B/H + B/M + B/L)$$

$$SMBROE = 1/3\ (S/R + S/N + S/W) – 1/3\ (B/R + B/N + B/W)$$

$$SMBINV = 1/3\ (S/C + S/N + S/A) – 1/3\ (B/C + B/N + B/A)$$

## 2. Empirical Design: Modelling Predictive COVID-19 Confirmed Cases

### 2.1 Statistical Design Approach

Four methods are frequently used to model epidemic outbreaks: the exponential growth model, the logistic growth model, the multi hump logit model, and a Markov Chain Monte Carlo Simulation.

Exponential growth model:

$$C_t = ae^{(rt)}$$

Logistic growth model:

$$C_t = k / (1 + ce^{(-rt)})$$

Upon initial reviews of literature and interviews with industry professionals, the authors finds that multi hump logit model, and Markov Chain Monte Carlo Simulation are rarely used in practice by finance professionals. The latter models are primarily used by professionals in pathology studies in academia. Therefore, we mainly focus on the exponential and the logistic.

The paper conducts comparisons on the exponential and the logistic model, finding that predictions for the two models are both reasonably well during the initial phase of the pandemic. However, their 95 percent confidence intervals stop overlapping at the ending 40% of the prediction period. After this point, the logistic model performs significantly better and converges with the actual number of cases, while the exponential model continues rising.

Therefore, the paper favors the logistic growth model. The parameters are estimated by a linear ordinary least squares (OLS) method to daily new COVID-19 confirmed cases. The model fits the time-series data very well for the whole of China, with a 95% confidence interval for parameters K, r, and P0. The r value for virus transmissibility that incorporates the level of physical quarantine and public sentiment was evaluated through stochastic Markov chain Monte Carlo (MCMC) methods and exponential adjustment (IDEA) model. In our paper, we assume r is fixed within each country, and from literature review, we find that the commonly adopted province average r for China is evaluated at r = 0.25. $P_0$ is the starting value of the number of confirmed cases for each period, and K is the carrying limit.

We then define our logistic_increase_function(t, K, $P_0$, r) as:

$$(K * \text{exponential}\ (r * (t - t_0)) * P_0) / (K + (\text{exponential}\ (r * (t - t_0)) - 1) * P_0)$$

We make two key assumptions during our estimation of logistic functions in implementation.

(1) COVID data is available each calendar day, whereas financial data is available only on trading days. With this in mind, we design special treatments for holidays during the period by omitting the data of the first returning day from the holiday. We apply this methodology because of the wealth of data during the holiday period, the accuracy of the regressed coefficient with our case estimate to financial metrics is likely to be an outlier and compromise its predicting ability.

(2) Estimates for each day t are sensitive to the choice of starting values P0 for that day, particularly in the initial days of the pandemic. This feature of the estimation is not surprising as when the number of cases is relatively small, a wide range of logistic curves may be consistent with the data, and the objective function across them may be relatively flat. Therefore, starting day (P0) is set to be t-15. And for days 2-15, starting day is day 1.

## 2.2 Deep Learning Approach

In this approach, the paper first uses the conventional transfer learning methods, using three pre-trained learning models. The Xception model shows a relatively ideal effect, and the diagnostic accuracy is above 90%. The deep learning model is inspired by and built upon several Github resources, with the most prominent one from MIT's repository: https://github.com/tarunk04/COVID-19-CaseStudy-and-Predictions.

Model overview, and key parameters from the Deep Learning Algorithm are attached below (Figure 2, Figure 3), since Machine Learning is not the primary focus in this paper, the paper will not go into details on the mining process.

```
In [49]: model = models.load_model("/Users/angelina/Desktop/model_confirmed.h5")
         model.summary()

Model: "model_16"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_19 (InputLayer)        [(None, 1)]               0
_________________________________________________________________
Dense_l1 (Dense)             (None, 80)                160
_________________________________________________________________
LRelu_l1 (LeakyReLU)         (None, 80)                0
_________________________________________________________________
Dense_l2 (Dense)             (None, 80)                6480
_________________________________________________________________
LRelu_l2 (LeakyReLU)         (None, 80)                0
_________________________________________________________________
Dense_l3 (Dense)             (None, 1)                 81
_________________________________________________________________
Output (LeakyReLU)           (None, 1)                 0
=================================================================
Total params: 6,721
Trainable params: 6,721
Non-trainable params: 0
```

Figure 2: Model Summary for China Confirmed Cases

```
In [43]: model_usa_c = models.load_model("/Users/angelina/Desktop/model_usa_c.h5")
         model_usa_c.summary()

Model: "model_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_6 (InputLayer)         [(None, 1)]               0
_________________________________________________________________
Dense_11 (Dense)             (None, 80)                160
_________________________________________________________________
LRelu_11 (LeakyReLU)         (None, 80)                0
_________________________________________________________________
Dense_12 (Dense)             (None, 80)                6480
_________________________________________________________________
LRelu_12 (LeakyReLU)         (None, 80)                0
_________________________________________________________________
Dense_13 (Dense)             (None, 1)                 81
_________________________________________________________________
Output (LeakyReLU)           (None, 1)                 0
=================================================================
Total params: 6,721
Trainable params: 6,721
Non-trainable params: 0
_________________________________________________________________
```

Figure 3: Model Summary for US Confirmed Cases

## 3. Empirical Design: Modelling the Shock Function

We employ two methods to model the unexpected shock of the severity of outbreak between two consecutive days.

The more intuitive approach is to design:

$$shock_{(t)} = \ln(C_{(t)}) - est.\ \ln(C_{(t-1)})$$

The second approach is:

$$shock_{(t)} = est.\ \ln(C_{(t-1)}) - est.\ \ln(C_{(t-2)})$$

Though method two is less common, it captures investor estimate from the previous two days using the number of infections on day t – 1 comparing to the number of cases for day t using t-2 estimates. The method is more credible especially during weekends when financial data is lacking, and noisy news to investors is highly unpredictable. For the Chinese market, the complicating factor that involves information reception is insignificant because previous day infection data is distributed before market opening each day.

Upon visualizing the exponential parameter estimates as defined in Method 1 and 2 respectively, where left axis reports the log change for each day, the paper decides on employing Method 2 as the magnitude for the changes is more consistent with our previous expectations.

Similarly, shock estimates are plotted for the deep learning approach. The magnitude of shocks is even smaller, within very ideal range.

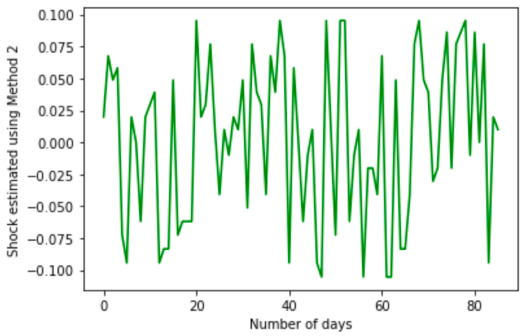


Figure 4: Shock Estimates for COVID-19 in China with Method 2 – Deep Learning Approach

Therefore, the paper selects the Deep Learning estimates moving forward, which is also more consistent with industry practices at the time of COVID-19 crisis.

## 4. Empirical Design: Regression Analysis- Multivariate Linear Regression

As introduced in Section 1.1, we examine the link between changes in model predictions and stock market behavior through benchmarks 1-year treasury bond yield, CSI 300 Index SSE Composite Index, and industry day-to-day index returns, with Fama-French five factors. We aim to show that unanticipated changes in predicted infections based on daily estimations obtained from the Deep Learning Model forecasts next-day stock returns. We adopted multiple regression formulas in order to find the most accurate and reliable regression, landing on a two-stage regression.

However, the paper would like to briefly show how the idea developed with the example of the Computer Science Industry Index in the Chinese equities market.

Regression 1:

$$\Delta \ln(\text{Industry Index}_{(t)}) = \beta_0 + \beta_1 * \text{shock}_{(t)} + \varepsilon_t$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.053
Model:                            OLS   Adj. R-squared:                  0.040
Method:                 Least Squares   F-statistic:                     4.056
Date:                Wed, 07 Apr 2021   Prob (F-statistic):             0.0477
Time:                        07:37:31   Log-Likelihood:                -180.25
No. Observations:                  75   AIC:                             364.5
Df Residuals:                      73   BIC:                             369.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1625      0.313      0.519      0.606      -0.462       0.787
x            -10.1691      5.049     -2.014      0.048     -20.232      -0.106
==============================================================================
Omnibus:                        1.988   Durbin-Watson:                   2.240
Prob(Omnibus):                  0.370   Jarque-Bera (JB):                1.966
Skew:                          -0.372   Prob(JB):                        0.374
Kurtosis:                       2.726   Cond. No.                         16.1
==============================================================================
```

Figure 5: Regression Results for Method 1

With adjusted r-squared being 4%, and coefficient of the shock function as -10.16.

Regression 2:

$$\Delta \ln(\text{Industry Index}_{(t)}) - Rf_{(t)} = \beta_0 + \beta_1 * \text{shock}_{(t)} + \beta_2 * SMB_{(t)} + \beta_3 * HML_{(t)} + \varepsilon_t$$

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | y | **R−squared (uncentered):** | 0.641 |
| **Model:** | OLS | **Adj. R−squared (uncentered):** | 0.626 |
| **Method:** | Least Squares | **F−statistic:** | 42.86 |
| **Date:** | Wed, 07 Apr 2021 | **Prob (F−statistic):** | 5.28e−16 |
| **Time:** | 07:38:04 | **Log−Likelihood:** | −144.00 |
| **No. Observations:** | 75 | **AIC:** | 294.0 |
| **Df Residuals:** | 72 | **BIC:** | 300.9 |
| **Df Model:** | 3 | | |
| **Covariance Type:** | nonrobust | | |

| | **coef** | **std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **shock** | −5.8501 | 3.229 | −1.812 | 0.074 | −12.287 | 0.587 |
| **SMB** | 0.7837 | 0.244 | 3.209 | 0.002 | 0.297 | 1.270 |
| **HML** | −2.3933 | 0.284 | −8.431 | 0.000 | −2.959 | −1.827 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 0.080 | **Durbin−Watson:** | 2.342 |
| **Prob(Omnibus):** | 0.961 | **Jarque−Bera (JB):** | 0.100 |
| **Skew:** | 0.067 | **Prob(JB):** | 0.951 |
| **Kurtosis:** | 2.882 | **Cond. No.** | 16.1 |

Figure 6: Regression Results for Method 2

With adjusted r-squared being 62.6%, and coefficient of the shock function as -5.85.

Regression 3:

$$\Delta \ln (\text{Industry Index}_{(t)}) - Rf_{(t)}$$

$$= \beta_0 + \beta_1 * \text{shock}_{(t)} + \beta_2 * SMB_{(t)} + \beta_3 * HML_{(t)} + \beta_4 * MKT_{(t)} + \varepsilon_t$$

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | y | **R-squared (uncentered):** | 0.875 |
| **Model:** | OLS | **Adj. R-squared (uncentered):** | 0.868 |
| **Method:** | Least Squares | **F-statistic:** | 123.8 |
| **Date:** | Wed, 07 Apr 2021 | **Prob (F-statistic):** | 3.15e-31 |
| **Time:** | 07:38:29 | **Log-Likelihood:** | -104.56 |
| **No. Observations:** | 75 | **AIC:** | 217.1 |
| **Df Residuals:** | 71 | **BIC:** | 226.4 |
| **Df Model:** | 4 | | |
| **Covariance Type:** | nonrobust | | |

| | **coef** | **std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **MKT** | 1.1319 | 0.098 | 11.498 | 0.000 | 0.936 | 1.328 |
| **shock** | -1.4304 | 1.960 | -0.730 | 0.468 | -5.339 | 2.478 |
| **SMB** | 0.6942 | 0.146 | 4.769 | 0.000 | 0.404 | 0.984 |
| **HML** | -1.2660 | 0.195 | -6.480 | 0.000 | -1.656 | -0.876 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 0.988 | **Durbin-Watson:** | 1.760 |
| **Prob(Omnibus):** | 0.610 | **Jarque-Bera (JB):** | 0.637 |
| **Skew:** | 0.221 | **Prob(JB):** | 0.727 |
| **Kurtosis:** | 3.093 | **Cond. No.** | 26.0 |

Figure 7: Regression Results for Method 3

With adjusted r-squared being 86.8%, and coefficient of the shock function as -1.43.

Regression 4:

Step 1:

$$MKT_{(t)} = \beta_0 + \beta_1 * shock_{(t)} + \varepsilon_t$$

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | MKT | **R-squared (uncentered):** | 0.045 |
| **Model:** | OLS | **Adj. R-squared (uncentered):** | 0.032 |
| **Method:** | Least Squares | **F-statistic:** | 3.457 |
| **Date:** | Wed, 07 Apr 2021 | **Prob (F-statistic):** | 0.0670 |
| **Time:** | 07:31:06 | **Log-Likelihood:** | −131.77 |
| **No. Observations:** | 75 | **AIC:** | 265.5 |
| **Df Residuals:** | 74 | **BIC:** | 267.9 |
| **Df Model:** | 1 | | |
| **Covariance Type:** | nonrobust | | |

| | **coef** | **std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **shock** | −4.8840 | 2.627 | −1.859 | 0.067 | −10.118 | 0.350 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 3.753 | **Durbin-Watson:** | 2.199 |
| **Prob(Omnibus):** | 0.153 | **Jarque-Bera (JB):** | 3.117 |
| **Skew:** | −0.320 | **Prob(JB):** | 0.211 |
| **Kurtosis:** | 3.767 | **Cond. No.** | 1.00 |

Figure 8: Regression Results for Method 4 – Step 1

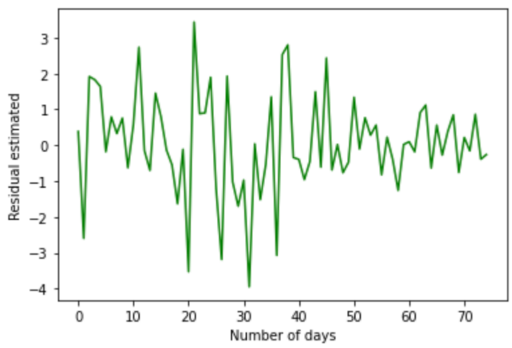


Figure 9: Residual Plot for Method 4 – Step 1

Step 2:

$$\Delta \ln(\text{Industry Index}_{(t)})$$

$$= \beta_0+\beta_1 * \text{shock}_{(t)} + \beta_2 * \text{SMB}_{(t)} + \beta_3 * \text{HML}_{(t)} + \beta_3 * \text{Est. Residuals}_{(t)} + \varepsilon_t$$

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | Computer return | R-squared (uncentered): | 0.875 |
| Model: | OLS | Adj. R-squared (uncentered): | 0.868 |
| Method: | Least Squares | F-statistic: | 123.9 |
| Date: | Wed, 07 Apr 2021 | Prob (F-statistic): | 3.08e-31 |
| Time: | 07:35:46 | Log-Likelihood: | -104.55 |
| No. Observations: | 75 | AIC: | 217.1 |
| Df Residuals: | 71 | BIC: | 226.4 |
| Df Model: | 4 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| shock | -6.9581 | 1.924 | -3.616 | 0.001 | -10.795 | -3.121 |
| SMB | 0.6944 | 0.146 | 4.771 | 0.000 | 0.404 | 0.985 |
| HML | -1.2672 | 0.195 | -6.487 | 0.000 | -1.657 | -0.878 |
| res | 1.1317 | 0.098 | 11.498 | 0.000 | 0.935 | 1.328 |

| | | | |
|---|---|---|---|
| Omnibus: | 0.991 | Durbin-Watson: | 1.760 |
| Prob(Omnibus): | 0.609 | Jarque-Bera (JB): | 0.641 |
| Skew: | 0.222 | Prob(JB): | 0.726 |
| Kurtosis: | 3.092 | Cond. No. | 24.9 |

Figure 10: Regression Results for Method 4 – Step 2

With adjusted r-squared being 87.5%, and coefficient of the shock function as -6.95.

From the four regression analyses above, we can clearly see Regression 4 offers significantly superior explainability, with very high R-squared. The intuition and the logic behind such result is that because the overall market return is highly correlated with the unanticipated shock variable, if we were to put both variables into the right-hand-side of the same equation, there would be significant problem of multicollinearity. This refers to a situation in which two explanatory variables in a multiple regression model are highly linearly related. Upon testing with the Variance Inflation Factor, this assumption is further proven.

Therefore, the paper employs a 2-stage approach where step 1 strips away the correlated factor and we obtain the residual, which is the movement in market returns that cannot be explained by the unanticipated shock. We then run the consecutive regression taking into account only the shock and the residual, but not the market return. From the results above, our model is able to explain over 85% of the variations in the Computer Science Industry Index in the Chinese market, with the coefficient of the shock variable being -6.95. In summary, we finalize our regression formula as method 4 where we run two consecutive regressions.

## 5. Generalizing Regression Analysis to All Industries in China

The paper then runs regression analysis on all 28 industries as defined by the ShenWan Level One industry segments. Their subsequent R-squared and coefficients can be found in the table below.

| | R-squared | Coefficient of the shock variable |
|---|---|---|
| Farming | 0.624 | -5.2355 |
| Mining | 0.588 | -7.3347 |
| Chemical | 0.484 | -4.2397 |
| Iron and Steel | 0.564 | -3.8572 |
| Colored Metals | 0.648 | -1.2358 |
| Electronics | 0.759 | -2.3593 |
| Household Appliances | 0.693 | -3.5821 |
| Food and Drinks | 0.889 | -1.4924 |
| Clothing | 0.769 | -4.1824 |
| Light Industry | 0.495 | -2.4512 |
| Pharmaceutical | 0.598 | 0.4147 |
| Public Affairs | 0.802 | -5.2349 |
| Transportation | 0.768 | -8.8481 |
| Realty | 0.729 | -3.8294 |
| Business and Commerce | 0.889 | -6.2843 |
| Leisure | 0.603 | -26.4824 |
| General | 0.562 | -5.2033 |
| Industrial Materials | 0.49 | -4.2854 |
| Industrial Decorations | 0.623 | -2.5911 |
| Electrical Equipments | 0.582 | -4.2854 |
| National Defense and Military | 0.782 | -2.3593 |
| Computer Science | 0.875 | -6.9581 |
| Media | 0.885 | -6.0059 |
| Communications | 0.893 | -1.2452 |
| Banking | 0.857 | -5.192 |
| Non-bank Finance | 0.903 | -5.7586 |
| Automobile | 0.571 | -6.3912 |
| Mechanical Equipment | 0.294 | -4.5912 |

Figure 11: Key Regression Results for All Industry Indices

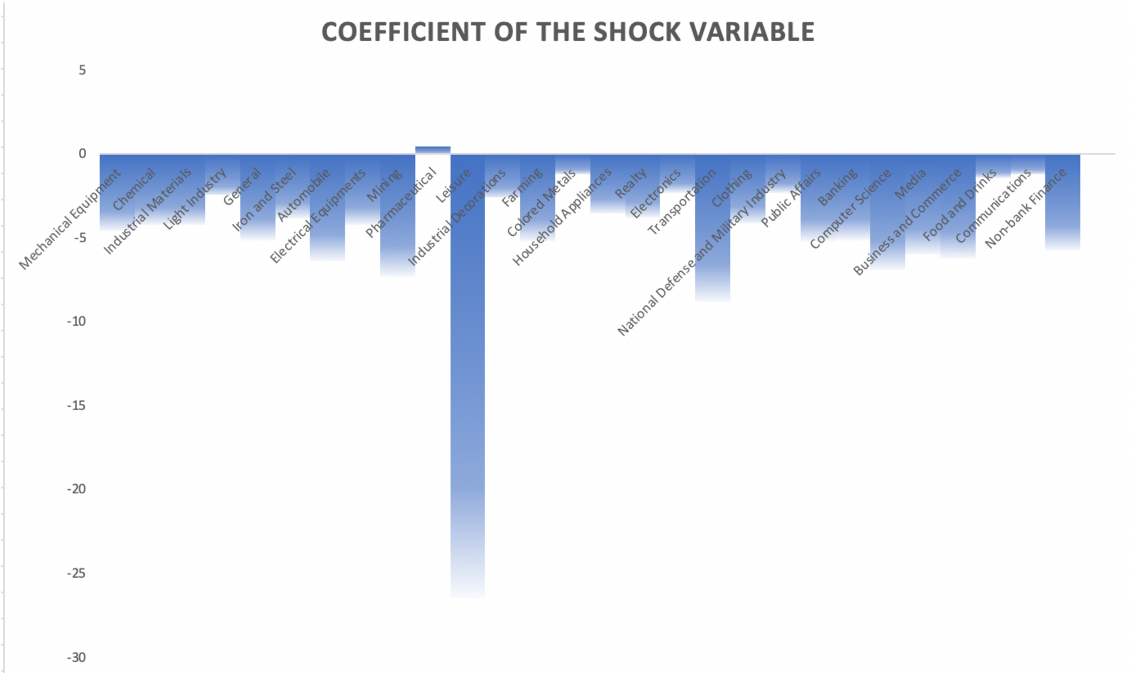


Figure 12: Graphic Illustration of Coefficients on Shock Variable - All Industry Indices

There are several industries worth noting in more details:

### 5.1 The Most Negatively Impacted Industry – Leisure

With a coefficient as low as -26.4824, the Leisure Industry is undoubtedly the most impacted in China. The interpretation for the coefficient is that a one unit change in unanticipated shock results in a -26.4824 change in the expected value of the logarithm of the industry index returns while all the predictors are held constant. The intuition behind such observation is obvious. The Chinese government quickly implemented emergency actions to fight the spread of this virus, appealing to individuals and businesses to not conduct unnecessary tourism, face-to-face meetings, events, and so on. Such suspension from a policy-making level largely cut down people's leisure activities. Additionally, the panic of infection and the quarantine procedures enforced by the government led many people to abandoning their tourism plans. Both factors contribute to the large negative influence.

### 5.2 The Most Positively Impacted Industry – Pharmaceutical

On the other side of the equation, we see pharmaceutical industry as the single positively benefited industry by COVID-19 in China. The coefficient of +0.4147 was to our surprise as though it is anticipated that pharmaceutical would have the least negative impact by COVID-19, the fact that COVID-19 in fact benefited the industry was not expected.

There are mainly four contributing reasons:

(1) societal awareness for disease control.

(2) prevention and healthcare were enhanced throughout the period of the pandemic.

(3) the government has laid more emphasis on driving the development of hierarchical medical systems, strengthening the competence of community medical intuitions.

(4) the wide development and commercialization of medicine, vaccines and medical devices are enhanced.

### 5.3 The Industry Performance with the Least Explaining Power – Mechanical Equipment

In general, industries such as mechanical equipment, industrial materials, light industries, iron and steel, automobile are all not well-explained by our unanticipated shock estimates. From our regression summary, we see that the R-Squared for such industries are relatively around 0.5, indicating a relatively weak and low-effect model. The reason is that such industries are mainly infrastructural, therefore making it less volatile. Their industry index performances largely depend on other factors that cover their supply and demand in their chain of business.

### 5.4 The Industry Performance with the Most Explaining Power – Non-bank Finance

The performance of index returns of the finance industry in general shows large explainability from COVID-19, with R-Squared close to 0.9. One intuition includes the fact that finance is largely people-dependent hence the inability to conduct business in person largely compromises the effectiveness of the industry. Additionally, the finance industry shows the market sentiment, where the panicking mindset of the public is conveyed and reflected.

## 6. Analyzing the Whole Timeframe Versus Two Subsample Periods

To further our research on how the impact of shock varies with the development of COVID-19 spreading patterns, the paper will also break down the initial 75-day period into a 45-day period and the consecutive 30-day period for the Chinese market. In this section, we will only employ the 2-step regression model as in method 4 with the Computer Science industry as an example for the general market, as we have developed it as the model offering most explaining ability.

$$\text{Step 1: } MKT_{(t)} = \beta_0+\beta_1 * shock_{(t)} + \varepsilon_t$$

OLS Regression Results (Days 0 – 45)

| | | | |
|---|---|---|---|
| Dep. Variable: | MKT | R-squared (uncentered): | 0.024 |
| Model: | OLS | Adj. R-squared (uncentered): | 0.002 |
| Method: | Least Squares | F-statistic: | 1.104 |
| Date: | Thu, 08 Apr 2021 | Prob (F-statistic): | 0.299 |
| Time: | 06:24:02 | Log-Likelihood: | −87.652 |
| No. Observations: | 45 | AIC: | 177.3 |
| Df Residuals: | 44 | BIC: | 179.1 |
| Df Model: | 1 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| shock | −4.7588 | 4.529 | −1.051 | 0.299 | −13.886 | 4.369 |

| | | | |
|---|---|---|---|
| Omnibus: | 0.786 | Durbin-Watson: | 2.150 |
| Prob(Omnibus): | 0.675 | Jarque-Bera (JB): | 0.736 |
| Skew: | −0.291 | Prob(JB): | 0.692 |
| Kurtosis: | 2.769 | Cond. No. | 1.00 |

OLS Regression Results (Days 46 – 75)

| | | | |
|---|---|---|---|
| Dep. Variable: | MKT | R-squared (uncentered): | 0.169 |
| Model: | OLS | Adj. R-squared (uncentered): | 0.141 |
| Method: | Least Squares | F-statistic: | 5.915 |
| Date: | Thu, 08 Apr 2021 | Prob (F-statistic): | 0.0214 |
| Time: | 07:03:03 | Log-Likelihood: | −34.776 |
| No. Observations: | 30 | AIC: | 71.55 |
| Df Residuals: | 29 | BIC: | 72.95 |
| Df Model: | 1 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| shock | −5.0079 | 2.059 | −2.432 | 0.021 | −9.219 | −0.797 |

| | | | |
|---|---|---|---|
| Omnibus: | 6.124 | Durbin-Watson: | 2.054 |
| Prob(Omnibus): | 0.047 | Jarque-Bera (JB): | 4.359 |
| Skew: | 0.806 | Prob(JB): | 0.113 |
| Kurtosis: | 3.942 | Cond. No. | 1.00 |

Figure 13: Regression Results for Days 0 – 45 (left), Days 46 – 75 (right)

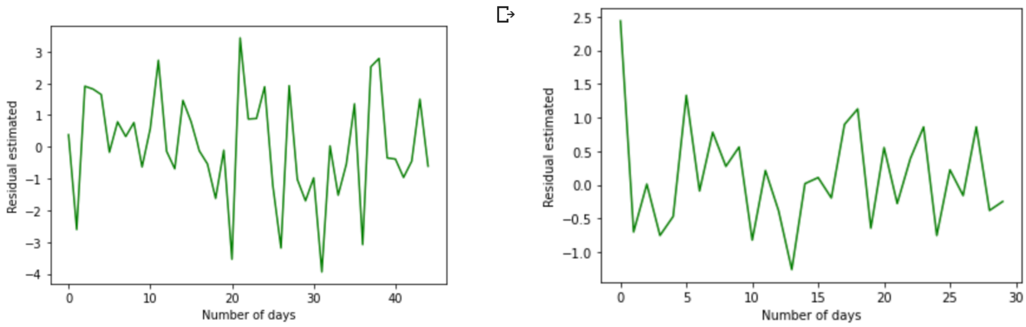


Figure 14: Residual Plot for Days 0 – 45 (left), Days 46 – 75 (right)

## Step 2: $\Delta\ln$ (Industry Index $_{(t)}$)

$$= \beta_0 + \beta_1 * \text{shock}_{(t)} + \beta_2 * \text{SMB}_{(t)} + \beta_3 * \text{HML}_{(t)} + \beta_3 * \text{Est. Residuals}_{(t)} + \varepsilon_t$$

OLS Regression Results (Days 0 – 45)

| | | | |
|---|---|---|---|
| Dep. Variable: | Computer return | R-squared (uncentered): | 0.885 |
| Model: | OLS | Adj. R-squared (uncentered): | 0.873 |
| Method: | Least Squares | F-statistic: | 78.54 |
| Date: | Thu, 08 Apr 2021 | Prob (F-statistic): | 1.15e−18 |
| Time: | 07:08:41 | Log-Likelihood: | −67.270 |
| No. Observations: | 45 | AIC: | 142.5 |
| Df Residuals: | 41 | BIC: | 149.8 |
| Df Model: | 4 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| shock | −6.3899 | 3.136 | −2.038 | 0.048 | −12.722 | −0.057 |
| SMB | 0.6383 | 0.203 | 3.138 | 0.003 | 0.228 | 1.049 |
| HML | −1.4559 | 0.256 | −5.678 | 0.000 | −1.974 | −0.938 |
| res | 1.0892 | 0.117 | 9.308 | 0.000 | 0.853 | 1.326 |

| | | | |
|---|---|---|---|
| Omnibus: | 0.115 | Durbin-Watson: | 1.679 |
| Prob(Omnibus): | 0.944 | Jarque-Bera (JB): | 0.150 |
| Skew: | 0.104 | Prob(JB): | 0.928 |
| Kurtosis: | 2.809 | Cond. No. | 33.2 |

OLS Regression Results (Days 46 – 75)

| | | | |
|---|---|---|---|
| Dep. Variable: | Computer return | R-squared (uncentered): | 0.861 |
| Model: | OLS | Adj. R-squared (uncentered): | 0.840 |
| Method: | Least Squares | F-statistic: | 40.28 |
| Date: | Thu, 08 Apr 2021 | Prob (F-statistic): | 8.77e−11 |
| Time: | 07:09:13 | Log-Likelihood: | −33.191 |
| No. Observations: | 30 | AIC: | 74.38 |
| Df Residuals: | 26 | BIC: | 79.99 |
| Df Model: | 4 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| shock | −7.3336 | 2.094 | −3.502 | 0.002 | −11.638 | −3.029 |
| SMB | 0.7711 | 0.200 | 3.856 | 0.001 | 0.360 | 1.182 |
| HML | −0.6603 | 0.324 | −2.035 | 0.052 | −1.327 | 0.007 |
| res | 1.4204 | 0.255 | 5.571 | 0.000 | 0.896 | 1.944 |

| | | | |
|---|---|---|---|
| Omnibus: | 0.326 | Durbin-Watson: | 1.950 |
| Prob(Omnibus): | 0.849 | Jarque-Bera (JB): | 0.410 |
| Skew: | 0.218 | Prob(JB): | 0.815 |
| Kurtosis: | 2.629 | Cond. No. | 14.6 |

Figure 10: Regression Results for Days 0 – 45 (left), Days 46 – 75 (right)

Compared with regressing on the entire 75 days, we can see the r-squared in the first stage is significantly lower in both subsamples. Moreover, the second stage suggests that in days 46-75, unanticipated shock has more negative effect on index performance than the initial 45 days, which seems to contradict our intuition. Upon generalizing to other industry indices in China, we come to a conclusion that dividing into subsamples does not provide tremendous justification abilities to our studies.

## 7. Conclusion

This paper shows that day-to-day changes in the Deep Learning model predictions of the COVID-19 pandemic forecast changes in industry index returns in the Chinese equities market. In future updates to this paper, we plan to extend the analysis to other countries, and to investigate the similarities and differences between index performances of the same industry across countries. Future studies may also be conducted on the examination of more granular state-wise and province-wise index performances.

**References**

Laura Alfaro, Anusha Chari, Andrew N. Greenland & Peter K. Schott. 2020. "Aggregate and Firm-Level Stock Returns During Pandemics, in Real Time." *National Bureau of Economic Research*. (26950). http://www.nber.org/papers/w26950

Alan, Nazli Sila and Engle, Robert F. and Karagozoglu, Ahmet K. 2020. "Multi-regime Forecasting Model for the Impact of COVID-19 Pandemic on Volatility in Global Equity Markets." *NYU Stern School of Business.* (June 15, 2020) https://ssrn.com/abstract=3646520 or http://dx.doi.org/10.2139/ssrn.3646520