A Research on Cross-sectional Return Dispersion and Volatility of US Stock Market during COVID-19

Jiawei Du
Abstract:

We studied the volatility and cross-sectional return dispersion effect of S&P Health Care Sector under the covid-19 epidemic. We innovatively used the Google index to proxy the impact of the epidemic and modeled the volatility. We also studied the influencing factors of the log-return of S&P Energy Sector and S&P Health Care Sector. We found that volatility is significantly affected by both the epidemic and cross-sectional return dispersion, and the coefficients in front of them are all positive, which means that the herding behaviour did not exist and as the cross-sectional return dispersion increases and the epidemic becomes more severe, the volatility of stock returns is also increasing. We also found that the epidemic has a significant negative impact on the return of the energy sector, and finally we provided our suggestions to investors.

Keywords: COVID-19, Cross-sectional return dispersion, Volatility, Health care sector, Energy sector, GJR-Garch(1,1)-X, Garch(1,1)-X, EGarch(1,1)-X
# Contents

1. Introduction .................................................................................................................. 4

2. Literature Review ......................................................................................................... 4

   2.1 Return Dispersion and Herding Behavior ............................................................... 4

   2.2 COVID-19 .................................................................................................................. 5

   2.3 Internet Search and Google Index ........................................................................... 6

3. Data ............................................................................................................................... 6

4. Methodology .................................................................................................................. 8

   4.1 ADF Test, Normal Distribution Test and Arch-LM Test .......................................... 8

   4.2 Model ........................................................................................................................ 9

5. Finding and Discussion ................................................................................................. 10

6. Conclusion and Suggestions ......................................................................................... 15

   6.1 Conclusions .............................................................................................................. 15

   6.2 Suggestions ............................................................................................................... 15

References: ....................................................................................................................... 16

Appendix A: Code .............................................................................................................. 18

   Stata code ....................................................................................................................... 18

   Python CSAD and CSSD Calculation code ................................................................. 20

   Python Plot .................................................................................................................... 22

Appendix B: Raw Data ...................................................................................................... 24

   Short horizon .................................................................................................................. 24

   Long horizon .................................................................................................................. 25
1. Introduction

Covid-19 epidemic swept the world in 2020, first in China, then in Japan, South Korea, Iran, Italy, Spain, the United Kingdom and the United States. The epidemic caused a sharp economic downturn. According to the information of the United States Department of Commerce, the real GDP of the United States fell by 4.8% in the first quarter.

Herding and return dispersion have been studied by many researchers. Many literatures showed that herding exists widely in stock market, especially in developing countries (Chiang, T. C., and Zheng, D., 2010). In this paper, we want to explore whether there is herding effect in American stock market, whether the epidemic has an impact on the return of American stock market, and what factors affect the stock volatility in this period. Different from the similar research, we not only use the traditional indicators such as the number of infected persons to represent the abstract variable of epidemic situation, but also use Google index to represent it, which enables us to better reflect the real impact of the epidemic situation when the data of COVID-19 in the United States is distorted in the early stage of the epidemic. In terms of sector selection, we selected S&P 500 health care sector and energy sector for comparison, and calculated their respective CSAD and CSSD indexes. Since the outbreak time in the United States is very close to now, and the epidemic is still developing, we chose two time periods, namely short time horizon from March 2 to May 29 (63 time observations), and long time horizon from February 20, 2020 to June 12, 2020 (80 time observations). The basis for this choice will be explained later. Finally, we will make investment suggestions to investors on our results.

2. Literature Review

2.1 Return Dispersion and Herding Behavior

Chang, E. C., Cheng, J. W., and Khorana, A. (2000) have made a remarkable contribution to the study of herd behaviour and herd behaviour. They have found that herd behaviour is more obvious in some emerging markets than in mature markets, especially during periods of rising market prices. Not only normal investors, Choi, N., and Sias, R. W. (2009) found that herding behaviour also existed in institutional investors. Fei, T., Liu, X., and Wen, C. (2019) regarded the cross-sectional stock return dispersion as a kind of herd behaviour indicator like previous researchers did, the cross-sectional return dispersion index in our later paper refers
Maio, P. (2016) studied the relation between stock return dispersion and the future stock return, he found that compared with other indicators, return dispersion have a stronger prediction, especially for the large-scale stocks and growth stocks. Prior to this, Stivers, C. T. (2003) conducted similar studies, but he used data on monthly returns dispersion. He also studied the relationship between returns dispersion and future volatility, and found that these relationships are very significant. Chortareas, G., Jiang, Y., and Nankervis, J. C. (2011) focus on the euro volatility, and they found that data with high-frequency and long-time horizon could improve volatility forecasting performance significantly.

In terms of the method for studying volatility, the conventional models are the ARCH model (Engle, R. F., 1982) and GARCH model (Bollerslev, T., 1986), but many researchers have used the improved model of the GARCH model. For example, Hwang*, S., and Satchell, S. E. (2005) have used the GARCHX model. Many researchers also used the GJR-Garch model proposed by Glosten et al. (1993) or the TGARCH model proposed by Zakoian (1994) or the EGARCH proposed by Nelson and Cao (1991). Fei, T., Liu, X., and Wen, C. (2019) chose the GARCH, GJR-GARCH, and HAR model, one of their findings is that the cross-sectional stock return dispersion did have significant influence on the forecast volatility. Onali, E. (2020) did a very updated research on the relation between the stock volatility and COVID-19 epidemic, he added the COVID-19 factor both into mean equation and volatility equation of the Garch(1,1) model, interestingly, he first regarded the increasing infections in Iran as a good index for American stock market, while in the end, he did not find the significance. Even, the infections and deaths cases from most of the countries (include America itself) he studied showed no influence on the US stocks returns, but showed significant influence on volatility.

### 2.2 COVID-19

In December 2019, an unexplained pneumonia virus outbreak occurred in Wuhan. The epidemic was first discovered in the Wuhan South China Seafood Wholesale Market. With the development of the epidemic, infected people also appeared in other provinces in China and in other
countries (Li et al. 2020). On February 11, 2020, WHO named the virus COVID-19. One month later, on March 12, 2020, the National Health Commission of China announced that the peak period of the covid-19 epidemic had passed, but other countries represented by US, Iran, Italy, and South Korea had an outbreak. At about the same time, WHO announced covid-19 Epidemic Global Pandemic.

Data from Johns Hopkins University, as of April 28, more than 1 million people in the United States were infected with COVID-19, and more than 2 million as of June 8.

2.3 Internet Search and Google Index

The data left behind by web searches for research and prediction has a long history. H. A. Johnson et al. (2004) collected the visit data of health-related websites and found that the number of visitors of flu-related articles and other flu-related data had a strong correlation with the influenza number data provided by the CDC. Later, J. Ginsberg et al. (2009) studied the relationship between Google search data and the incidence of influenza diseases, and found that through the modeling of the two, the outbreak of influenza could be predicted 1-2 weeks in advance of the traditional method. N. Askitas et al. (2009) took the unemployment rate of Germany as the research object. He conducted statistical analysis on the keywords related to unemployment rate in Google search and found that these keywords had a strong correlation with the actual unemployment rate. N. P. Lincoln (2011) studied the sales volume of apple's electronic products and empirically obtained the significant correlation between the sales volume of apple's electronic products and the Internet search data of Google by using the Internet search data provided by Google search.

We have reasons to believe that we can use Google index data to estimate the extent of the covid-19 epidemic.

3. Data

Stocks data we chose the constituent stocks in the S&P 500 Health Care Sector and S&P 500 Energy Sector from Yahoo Finance, time horizon from Feb 20, 2020 to June 12th, 2020 (80 time-observations). Also, we chose the time horizon from Mar 2nd to May 29th (63 time-observations), during which, Covid-19 in America is very severe, and after that, there was a large black march in the United States, which had a certain impact on our variables such as VIX and stock returns. It would be more scientific to select a narrower period.

In this paper, we focus on the three important factors. First is COVID-19. The herding behavior we study is in the context of the COVID-19 epidemic, thus, we need
to choose a proxy to represent the COVID-19. One is easy to understand, the infections in the US, we take the log of it (ln(1+x)), namely, ‘Inlg’. However, the lack of nucleic acid testing reagent and people's lack of awareness in the early stage of the epidemic in the United States caused distortion of the data, we should find another proxy to represent it, finally, we choose Google Index with keywords ‘COVID-19’. Why we choose it? The impact of the COVID-19 epidemic is twofold. On the one hand, it is the practical impact, such as the direct economic impact caused by the inability of infected people to work and the inability of factories or companies to function normally. On the other hand, the impact of the epidemic on people’s psychology. Once people panic about the COVID-19 epidemic, they will reduce their consumption and sell assets such as stocks, causing indirect economic losses. The increase in the number of covid-19 infections and people’s fear of the epidemic are both contributing to the rise of the Google index. Therefore, we believe that the Google index can better reflect the true impact of the epidemic with the distortion of the epidemic data. We take the log of it (ln(1+x)) as ‘Golg’. Also, we take log of (ln(1+x)) VIX Index, as another kind of panic indicator. In simple terms, see the table below.

| Table 1 | Proxy | Definition | Comment | Whether it can reflect the epidemic situation |
|---------|-------|------------|---------|---------------------------------------------|
| Inlg    | The log of Infections | Data distortion | Yes |
| Golg    | The log of Google Index | Reflect both the number of infections and investors’ panic about the epidemic | Yes |
| VIXlg   | The log of VIX Index | Reflect investors’ panic | Partial |

The second important factor is herding behavior, one of our targets. We followed Christie and Huang (1995) and use CSAD and CSSD to represent it.

\[
\text{CSSD}_t = \sqrt{\sum_{i=1}^{K} (r_{i,t} - r_{m,t})^2 / (K - 1)}
\]

(1)

\[
\text{CSAD}_t = \frac{\sum_{i=1}^{K} |r_{i,t} - r_{m,t}|}{K}
\]

(2)

Where \(r_m\) is the mean log-return.

The log-return, CSAD and CSSD plot see Figure 1 and 2.
The third important factor is the estimation of volatility. We use the Garch(1,1)-X model, EGarch(1,1)-X model and the GJR-arch(1)-X model. In next part, we will introduce it. For convenience, Table 2 showed the introduction of parameters.

| Parameter | Introduction |
|-----------|--------------|
| Splg      | The first order differential of S&P 500 Index log-return |
| Hclg      | The first order differential of S&P 500 HealthCare Sector Index log-return |
| Elg       | The first order differential of S&P 500 Index Energy log-return |
| Inlg      | The first order differential of log-infections |
| Golg      | The first order differential of log-Google-Index |
| Vixlg     | The first order differential of log-VIX-Index |
| Hccsad    | CSAD of S&P 500 HealthCare Sector |
| Hcssd     | CSSD of S&P 500 HealthCare Sector |
| Ecsad     | CSAD of S&P 500 Energy Sector |
| Ecssd     | CSSD of S&P 500 Energy Sector |

4. Methodology

4.1 ADF Test, Normal Distribution Test and Arch-LM Test

Before the establish of the model, we need to do the ADF Test and normal distribution test first, the Table 3 and 3 showed the test results. In Table 3, we see that except ‘Hccsad’ and ‘Hcssd’, all the variables showed the P-value less than 0.01, since the dataset is small, we still believed that the ‘Hccsad’ and ‘Hcssd’ (P-value
less than 0.1) do not exist unit root, they are stationary. Data with longer time horizon are more significant.

Table 4 showed the ‘sktest’ results, we regarded the distribution of the ‘hclg’ residual sequence as normal distribution while ‘elg’ as ‘t-distribution’ (later arch-LM test rejected the arch effect, we ended up abandoning it).

### Table 3. ADF Test

| Parameter            | t-statistic 63 time-observations | t-statistic 80 time-observations | 1% Critical | 5% Critical | 10% Critical | P-Value 63 time-obs. | P-Value 80 time-obs. |
|----------------------|----------------------------------|----------------------------------|-------------|-------------|-------------|---------------------|---------------------|
| Splg (S&P 500 Index) | -12.175                          | -13.212                          |             |             |             | 0.0000              | 0.0000              |
| Hclg                 | -12.324                          | -12.956                          |             |             |             | 0.0000              | 0.0000              |
| Elg                  | -10.012                          | -10.091                          |             |             |             | 0.0000              | 0.0000              |
| Inlg                 | -9.156                           | -8.677                           |             |             |             | 0.0000              | 0.0000              |
| Golg                 | -4.564                           | -7.376                           | -3.58       | -2.93       | -2.6        | 0.0002              | 0.0002              |
| Vixlg                | -10.439                          | -10.495                          |             |             |             | 0.0000              | 0.0000              |
| Hccsad               | -2.612                           | -3.191                           |             |             |             | 0.0905              | 0.0205              |
| Hccssd               | -2.612                           | -3.191                           |             |             |             | 0.0905              | 0.0205              |
| Ecsad                | -7.92                            | -4.978                           |             |             |             | 0.0000              | 0.0000              |
| Ecssd                | -7.92                            | -4.962                           |             |             |             | 0.0000              | 0.0000              |

### Table 4. Normal Distribution Test

| Parameter | P-Value (Skewness) | P-Value (Kurtosis) | Joint P-Value | P-Value (Skewness) | P-Value (Kurtosis) | Joint P-Value |
|-----------|--------------------|--------------------|---------------|--------------------|--------------------|---------------|
| Splg      | 0.0947             | 0.0253             | 0.0294        | 0.1022             | 0.0096             | 0.0160        |
| Hclg      | 0.3267             | 0.0961             | 0.1416        | 0.5139             | 0.0430             | 0.1015        |
| Elg       | 0.0061             | 0.0076             | 0.0025        | 0.0051             | 0.0042             | 0.0015        |
| Inlg      | 0.0001             | 0.0009             | 0.0000        | 0.1827             | 0.0000             | 0.0001        |
| Golg      | 0.0000             | 0.0016             | 0.0000        | 0.0000             | 0.0001             | 0.0000        |
| Vixlg     | 0.0014             | 0.0170             | 0.0016        | 0.0001             | 0.0111             | 0.0002        |
| Hccsad    | 0.0000             | 0.0272             | 0.0002        | 0.0000             | 0.0027             | 0.0000        |
| Hccssd    | 0.0000             | 0.0272             | 0.0002        | 0.0000             | 0.0027             | 0.0000        |
| Ecsad     | 0.0000             | 0.0000             | 0.0000        | 0.0000             | 0.0000             | 0.0000        |
| Ecssd     | 0.0000             | 0.0000             | 0.0000        | 0.0000             | 0.0000             | 0.0000        |

Another test is also needed, arch-LM test. We conduct it to verify if exist ‘arch’ effect. The p-value of the ‘hclg’ with short time horizon is 0.0039, long horizon is 0.0004. There exists ‘arch’ effect. For ‘elg’, the p-value are 0.8580 and 0.7475 respectively, ‘elg’ did not exist ‘arch’ effect.

### 4.2 Model

Garch (1,1)-X Model:
\[ y_t = \beta_0 + \epsilon_t \]  
\[ \sigma^2_t = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \beta_1 \sigma^2_{t-1} + \gamma_1 x_t \]

Where we use Inlg, Golg, VIXlg, HCCsad, HCCssd, LaggedHCCsad and LaggedHCCssd as \( x_t \) respectively.

Besides, for comparison, we would fit the data by two time-horizons. Time one from Mar 2nd to May 29th and time two from Feb 20th to June 12th.

**EGarch (1,1)-X Model:**
Mean equation is the same with the GARCH(1,1)-X model. And volatility term has been written as:

\[
\ln(\sigma^2_t) = \omega + \alpha_1 (\phi \eta_{t-1} + \gamma (|\eta_{t-1}| - E|\eta_{t-1}|)) + \beta_1 \ln(\sigma^2_{t-1}) + \lambda_1 x_t
\]

Where \( \eta_t \) is i.i.d. standardized random variables with \( \text{N}(0,1) \). Still, we would fit the data by two time-horizons.

**GJR-Garch (1,1)-X Model:**

\[
\sigma^2_t = \alpha_0 + \alpha_1 u^2_{t-1} + \beta_1 \sigma^2_{t-1} + \gamma u^2_{t-1} I_{t-1} + \lambda_1 x_t
\]

Where \( I_{t-1} = 1 \) if \( u_{t-1} < 0 \) and \( = 0 \) if \( u_{t-1} \geq 0 \)

In addition, we also want to study if the COVID-19 and lagged return dispersion affect the log return of Health Care Sector and Energy Sector, since we suspect that the epidemic will have a negative impact on energy stocks and a positive impact on health care stocks. Therefore, we establish the multiple linear regression model.

\[ y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \epsilon_t \]

Where \( x_{1,t} \) is the S&P 500 Index, \( x_{2,t} \) is the COVID-19 proxy, \( x_{3,t} \) is the return dispersion index (we use CSAD here due to the results are very closed).

### 5. Finding and Discussion

Table 5 and 6 showed the Garch-X(1,1) results, and table 7 and 8 showed EGarch-X(1,1) results. Both the models fit narrower time horizon data well, we think that is because the narrow time horizon is the most severe time period in the United States during COVID-19, after which, the stock market has digested many factors of the epidemic and there were protests against racism in the United States, and these protests may affect our proxies and stock returns. The time before march, COVID-19 infections are very low in the United
States, which also affect the model performance, that is why we could not run the results when using long time horizon.

### Table 5. Garch-(1,1) Health Care Sector
63 time-observations from Mar 2\(^{nd}\) to May 29\(^{th}\)

| Parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|-----------|---------|---------|---------|---------|---------|---------|---------|
| Garch-(1,1) with Het | Inlg | Golg | VIXlg | HCCsad | HCCssd | HCCssd | HCCssd |
| arch\_L1 | 0.4934783 (0.025) | 0.3115575 (0.016) | 0.1918693 (0.095) | 0.3071679 (0.024) | 0.3071679 (0.024) | 0.46644 (0.016) | 0.4663736 (0.016) |
| garch\_L1 | 0.4309016 (0.000) | 0.42564 (0.000) | 0.2839611 (0.03) | 0.3671006 (0.000) | 0.3071679 (0.024) | 0.46644 (0.016) | 0.3455894 (0.000) |
| cons | 0.0019067 227.8003 (0.000) | 0.0028897 | 0.0036610 | 0.0033123 | 0.0033123 | 0.0027997 | 0.0027973 |
| Inlg | / | / | / | / | / | / | / |
| Golg | / | / | / | / | / | / | / |
| VIXlg | / | / | / | 174.414 (0.000) | / | / | / |
| HCCsad | / | / | / | / | / | / | / |
| HCCssd | / | / | / | / | / | / | / |
| LaggedHCCsad | / | / | / | / | / | / | / |
| LaggedHCCssd | / | / | / | / | / | / | / |
| Het\_cons | -322.54790 | -10.10167 | 0.00366 | -12.42266 | -12.42266 | -12.49498 | -12.49653 |

Remark: Garch-(1,1) with Het Inlg, Golg, VIXlg: Stata cannot run the results.

### Table 6. Garch-(1,1) Health Care Sector
80 time-observations from Mar 2\(^{nd}\) to May 29\(^{th}\)

| Parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|-----------|---------|---------|---------|---------|---------|---------|---------|
| Garch-(1,1) with Het | Inlg | Golg | VIXlg | HCCsad | HCCssd | HCCssd | HCCssd |
| arch\_L1 | -0.0417931 (0.530) | -0.0417931 (0.530) | -0.0417931 (0.530) | 0.2325071 (0.002) | 0.2325070 (0.002) |
| garch\_L1 | 0.0437586 (0.764) | 0.0437586 (0.764) | 0.0437586 (0.764) | 0.6981627 (0.000) | 0.6981627 (0.000) |
| cons | -0.0010149 | -0.0010149 | -0.0010149 | 0.0004675 | 0.0004675 |
| Inlg | / | / | / | / | / |
| Golg | / | / | / | / | / |
| VIXlg | / | / | / | / | / |
| HCCsad | / | / | / | / | 158.7028 (0.000) |
| HCCssd | / | / | / | / | / |
| LaggedHCCsad | / | / | / | / | / |
| LaggedHCCssd | / | / | / | / | / |
| Het\_cons | -9.395639 | -9.395639 | -9.395639 | -10.69531 | -10.69531 |

Remark: Garch-(1,1) with Het Inlg, Golg, VIXlg: Stata cannot run the results.
In Table 5, the coefficients of the arch term and ‘Garch’ term are very significant in all the models from model 1 to model 7, we found that the coefficients of ‘Inlg’, ‘Golg’, ‘Vixlg’ are significantly positive, which implies that the COVID-19 affects the stock market volatility significantly, when COVID-19 more severe, investors more panic, and the stock volatility get higher. For herding behavior index, CSAD and CSSD, whether it is the current CSAD(CSSD) or the CSAD(CSSD) lagging one period, the coefficients in front of them are significantly positive, which means that there is no herding effect. But as far as these two indicators are concerned, the return dispersion has led to an increase in volatility. We believe that the reason could be that US dollar assets are safe-haven assets, and this COVID-19 epidemic is a global epidemic. When the global asset risk increases, Investors will increase the allocation of U.S. dollar assets, resulting in the negative factors caused by the epidemic and the positive factors brought about by risk aversion repeatedly affecting the market, resulting in return dispersion and rising volatility.

In Table 7, the coefficients of ‘egarch_a_L1’ and ‘egarch_L1’ are significant in model 1, 6 and 7, which indicate there exist the leverage effect, however, as for the whole, we cannot say that the leverage effect is significant. Table 8 showed part of the results. Some models Stata does not run the results, also, the ‘earch’ term of the model with current CSAD(CSSD) and vixlg are not significant. Meanwhile, these three models are not the focus of our discussion, so we will not discuss them here.

| Parameter                     | Model 1            | Model 2            | Model 3            | Model 4            | Model 5            | Model 6            | Model 7            |
|-------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| EGarch-X(1,1) with Het Inlg   | -0.1007196 (0.485) | -0.443223 (0.709)  | -0.327248 (0.294)  | -0.132522 (0.722)  | -0.132522 (0.722)  | -0.0962005 (0.575) | -0.096201 (0.575)  |
| egarch_a_L1                   | 0.4102835 (0.023)  | 0.4198592 (0.586)  | 0.2955902 (0.105)  | 0.5070969 (0.046)  | 0.5070969 (0.046)  | 0.5643479 (0.002)  | 0.5643479 (0.002)  |
| egarch_L1                     | 1.22968 (0.000)    | 0.1907936 (0.933)  | -0.046148 (0.907)  | 0.7805334 (0.105)  | 0.7805334 (0.105)  | 1.051382 (0.000)   | 1.051382 (0.000)   |
| cons                          | 0.0020182          | 0.0022444          | 0.0029944          | 0.0025576          | 0.0025576          | 0.0020959          | 0.0020959          |
| Inlg                          | 0.5259978 (0.216)  | /                  | /                  | /                  | /                  | /                  | /                  |
| Golg                          | /                  | 4.585414 (0.381)   | /                  | /                  | /                  | /                  | /                  |
| VIXLg                         | /                  | /                  | 0.0679617 (0.003)  | /                  | /                  | /                  | /                  |
| HCsad                         | /                  | /                  | /                  | 52.97203 (0.185)   | /                  | /                  | /                  |
| Hcssd                         | /                  | /                  | /                  | /                  | 6.569569 (0.185)   | /                  | /                  |
| Lagged HCsad                  | /                  | /                  | /                  | /                  | 21.23671 (0.462)   | /                  | /                  |
| Lagged HCSSD                  | /                  | /                  | /                  | /                  | /                  | 2.633768 (0.462)   | /                  |
| Het_cons                      | 1.33874 -5.96473   | -10.05605          | -2.77268           | -2.77268           | -0.2548682         | -0.2548682         | /                  |
Table 8. EGarch-X(1,1) Health Care Sector 80 obs.

| Parameter   | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|-------------|---------|---------|---------|---------|---------|---------|---------|
| EGarch-X(1,1) with Het | Inlg | Golg | VIXlg | HCcsad | HCcssd | HCcsad | HCcssd |
| earch_L1    | -0.2286454 (0.041) | -0.5837693 (0.001) | -0.5837692 (0.001) |         |         |         |         |
| egarch_a_L1 | 0.375154 (0.181)   | -0.026366 (0.903)  | -0.026366 (0.903)  |         |         |         |         |
| egarch_L1   | -0.6697288 (0.000) | 0.4099775 (0.180)  | 0.4099775 (0.180)  |         |         |         |         |
| cons        | 0.0072798         |         | -0.003212 | -0.003212 |         |         |         |
| Inlg        | /                  | /       |         |         |         |         |         |
| Golg        | /                  | /       |         |         |         |         |         |
| VIXlg       | 3.231587 (0.029)   |         |         |         |         |         |         |
| HCcsad      | /                  | 92.1735 (0.066)   |         |         |         |         |         |
| HCcssd      | /                  | /       |         |         |         |         | 14.53105 (0.002) |
| LaggedHCcsad| /                  | /       |         |         |         |         |         |
| LaggedHCcssd| /                  | /       |         |         |         |         |         |
| Het_cons    | -11.94937         | -5.589827 | -5.589826 |         |         |         |         |

Remark: EGarch-X(1,1) with Het Inlg, Golg, LaggedHCcsad, LaggedHCcssd: Stata cannot run the results.

Table 9 showed the GJR-Garch-X(1,1) results, compared with Egarch(1,1)-X, GJR model showed that the leverage effect (‘tgarch_a_L1’ term) is not significant, partly confirmed the results of the Egarch(1,1)-X model. We think the reason is that the COVID-19 epidemic is both good and bad effect for the stocks in health care. On the one hand, the epidemic makes the production of enterprises slow down and the economy under pressure. On the other hand, the market demand for the medical industry increases at this time, and finally the leverage effect is not significant. Nevertheless, of all the seven GJR-Garch-X(1,1) models, we are pleased that the $X_t$ term of volatility estimation are very significant, findings like Garch-X(1,1) model, Covid-19 has a positive impact on the volatility of the U.S. stock market. Herding effect is not significant during this period, but the dispersion of returns makes the volatility of stock returns increase.
Table 9. GJR-Garch-X(1,1) Health Care Sector 63 obs.

| Parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|-----------|---------|---------|---------|---------|---------|---------|---------|
| EGarch-X(1,1) with Het | Inlg | Golg | VIXlg | HCcsad | HCcssd | Lagged HCcsad | Lagged HCcssd |
| arch_L1 | 0.4948241 (0.012) | 0.5874224 (0.076) | 0.2504111 (0.145) | 0.3908686 (0.101) | 0.3908686 (0.101) | 0.4187954 (0.072) | 0.4187955 (0.072) |
| tarch_a_L1 | -0.0042664 (0.992) | -0.636893 (0.056) | 0.163735 (0.597) | -0.170801 (0.655) | -0.170801 (0.655) | 0.1276745 (0.809) | 0.1276746 (0.809) |
| garch_L1 | 0.4315761 (0.001) | 0.0074577 (0.887) | 0.389613 (0.000) | 0.3886578 (0.000) | 0.3886578 (0.000) | 0.3300482 (0.003) | 0.3300482 (0.003) |
| cons | 0.0018909 (0.000) | 0.0019388 (0.000) | 0.0029944 (0.000) | 0.0043292 (0.000) | 0.0043292 (0.000) | 0.0031582 (0.000) | 0.0031582 (0.000) |

Table 10. Multiple linear regression model

| Market factor | COVID-19 factor | Return dispersion factor |
|---------------|-----------------|-------------------------|
| Hclg (63 obs.) | SPlg | Golg | Inlg | LaggedHCcsad |
| Model 1 | 0.8371931 (0.000) | 0.0070321 (0.503) | / | -0.1129309 (0.482) |
| Model 2 | 0.8277126 (0.000) | / | -0.0008963 (0.836) | -0.1060166 (0.524) |
| Hclg (80 obs.) | | | | |
| Model 1 | 0.8386103 (0.000) | 0.0063166 (0.396) | / | -0.0027216 (0.838) |
| Model 2 | 0.8304563 (0.000) | / | -0.0001784 (0.957) | -0.0379182 (0.838) |
| Elg (63 obs.) | | | | |
| Model 1 | 1.261988 (0.000) | -0.0553695 (0.146) | / | 0.1986184 (0.685) |
| Model 2 | 1.334663 (0.000) | / | -0.0246747 (0.085) | 0.0944762 (0.837) |
| Elg (80 obs.) | | | | |
| Model 1 | 1.299324 (0.000) | -0.0510547 (0.037) | / | -0.022371 (0.965) |
| Model 2 | 1.366834 (0.000) | / | -0.0141673 (0.193) | 0.0429583 (0.935) |

Table 10 showed the results of multiple linear regression. From it, we found that except the market factor ‘splg’, both COVID-19 factor and return dispersion factor did not show the significant result on stocks in
Health Care sector. But for Energy Sector, COVID-19 factor ‘inlg’ in model 2 of 63 time-observations regression and ‘golg’ in model 1 of 80 time-observations showed the negative significant relation with the ‘elg’. Which partly confirm our conjectures. However, cross-sectional return dispersion(lagged) showed no relation with stock return.

6. Conclusion and Suggestions

6.1 Conclusions

We have the three main conclusion.

The first conclusion is that herding does not exist (in terms of S&P health care sector), however, the coefficients in front of CSAD(lagged) and CSSD(lagged) are significantly positive, which showed that cross-sectional return dispersion increased the volatility of stock returns.

The second conclusion is that the COVID-19 epidemic had a significant positive impact on the stocks return volatility of health care sector. The higher the number of infections and Google search volume, the higher the stocks return volatility.

The third conclusion is that the return-dispersion with lags one period has no significant impact on stock returns during the epidemic. The impact of the epidemic on different sectors is different, it has no significant impact on the health care sector, but has a significant negative impact on the energy sector.

6.2 Suggestions

We have three suggestions for investors. The first one is that investors can short the energy sector during the pandemic, and the specific operation basis can be the freely available Google index (with keywords COVID-19).

The second suggestion is that for risk-averse investors, investment in the stock market should be reduced during the COVID-19 outbreak, because the volatility will increase with the development of the epidemic.

The third suggestion is, for risk averse investors, if the return dispersion of the previous day is high, they can sell stocks or hedge stocks today to reduce the negative effects of high volatility. And for high-frequency traders, they can enter the market today to seek trading opportunities with high volatility.

4090 Words.
References:

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting.

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, 24(10), 1651-1679.

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases*, 49(10), 1557-1564.

Chiang, T. C., & Zheng, D. (2010). An empirical analysis of herd behavior in global stock markets. *Journal of Banking & Finance*, 34(8), 1911-1921.

Choi, N., & Sias, R. W. (2009). Institutional industry herding. *Journal of Financial Economics*, 94(3), 469-491.

Chortareas, G., Jiang, Y., & Nankervis, J. C. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different?. *International Journal of Forecasting*, 27(4), 1089-1107.

Christie, W. G., & Huang, R. D. (1995). Following the pied piper: Do individual returns herd around the market?. *Financial Analysts Journal*, 51(4), 31-37.

*COVID-19 United States Cases by County*, (2020). [Online]. Available from: [https://coronavirus.jhu.edu/map.html](https://coronavirus.jhu.edu/map.html) (Accessed: 14 June 2020).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Hwang*, S., & Satchell, S. E. (2005). GARCH model with cross-sectional volatility: GARCHX models. *Applied Financial Economics*, 15(3), 203-216.

Fei, T., Liu, X., & Wen, C. (2019). Cross-sectional return dispersion and volatility prediction. *Pacific-Basin Finance Journal*, 58, 101218.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5), 1779-1801.

Johnson, H. A., Wagner, M. M., Hogan, W. R., Chapman, W. W., Olszewski, R. T., Dowling, J. N., & Barnas, G. (2004). Analysis of web access logs for surveillance of Influenza. In *Medinfo* (pp. 1202-1206).

Lincoln, N. P. (2011). *The relationship between internet marketing, search volume, and product sales* (Doctoral dissertation, The Ohio State University).

Maio, P. (2016). Cross-sectional return dispersion and the equity premium. *Journal of Financial Markets*, 29, 87-109.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*. 59 (2): 347–370

Onali, E. (2020). Covid-19 and stock market volatility. *Available at SSRN 3571453*. 

16
Stivers, C. T. (2003). Firm-level return dispersion and the future volatility of aggregate stock market returns. *Journal of Financial Markets, 6*(3), 389-411.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control, 18*(5), 931-955.
Appendix A: Code

Stata code

gen date2=date(date,"YMD")
gen helg=hclg-elg
tset date2
*ADF Test
dfuller splg
dfuller hclg
dfuller elg
dfuller inlg
dfuller golg
dfuller vixlg
dfuller hccsad
dfuller hccssd
dfuller ecsad
dfuller ecssd
dfuller laggedhccsad
dfuller laggedhccssd
dfuller laggedecsad
dfuller laggedecssd

Normal-Distribution Test
sktest splg
sktest hclg
sktest elg
sktest inlg
sktest golg
sktest vixlg
sktest hccsad
sktest hccssd
sktest ecsad
sktest ecssd
sktest ecsad
sktest laggedhccsad
sktest laggedhccssd
sktest laggedecsad
sktest laggedecssd

*Garch(1,1)X
arch hclg, arch(1) garch(1) vce(robust) het(inlg)
arch hclg, arch(1) garch(1) vce(robust) het(golg)
arch hclg, arch(1) garch(1) vce(robust) het(vixlg)
arch hclg, arch(1) garch(1) vce(robust) het(hccsad)
arch hclg, arch(1) garch(1) vce(robust) het(hccssd)
arch hclg, arch(1) garch(1) vce(robust) het(laggedhccsad)
arch hclg, arch(1) garch(1) vce(robust) het(laggedhccssd)

*EGarch(1,1)X
arch hclg, earch(1) egarch(1) vce(robust)
arch hclg, earch(1) egarch(1) vce(robust) het(inlg)
arch hclg, earch(1) egarch(1) vce(robust) het(golg)
arch hclg, earch(1) egarch(1) vce(robust) het(vixlg)
arch hclg, earch(1) egarch(1) vce(robust) het(hccsad)
arch hclg, earch(1) egarch(1) vce(robust) het(hccssd)
arch hclg, earch(1) egarch(1) vce(robust) het(laggedhccsad)
arch hclg, earch(1) egarch(1) vce(robust) het(laggedhccssd)

*GJR- Garch(1,1)X
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(inlg)
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(golg)
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(vixlg)
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(hccsad)
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(hccssd)
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(laggedhccsad)
arch hclg, arch(1) tarch(1) garch(1) vce(robust) het(laggedhccssd)
* Multiple linear regression
  reg hclg splg golg laggedhccsad
  reg hclg splg inlg laggedhccsad
  reg elg splg golg laggedecsad
  reg elg splg inlg laggedecsad

Python CSAD and CSSD Calculation code

```python
# -*- coding: utf-8 -*-

Created on Fri Jun  5 15:22:09 2020

@author: GavinDu

from basicFun import getLag
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from pandas_datareader import data as pdr
import fix_yahoo_finance as yf
import math

# S&P 500 health care csad and cssd
# Download
L1 = pd.read_csv('c:/Users/Gavin/Desktop/s&p-sectors---health-care-06-05-2020.csv')
df = pd.DataFrame(columns = L1['Symbol'])
for i in L1['Symbol']:
    data = yf.download(i, start="2020-02-16", end="2020-06-15")
    df[i] = np.log(data['Adj Close']).diff().dropna()

df=df.drop(columns=['Downloaded from Barchart.com as of 06-05-2020 04:37am CDT'])
```
#Calculation
CSAD=[]
CSSD=[]
list1=[]
list2=[]
for r in range(df.iloc[:,0].size):
    for c in range(df.columns.size):
        list1.append(abs((df.iloc[r][c] - df.iloc[r][:].mean())/df.iloc[:,0].size))
        list2.append(((df.iloc[r][c] - df.iloc[r][:].mean())**2/(df.iloc[:,0].size-1))**0.5)
value = sum(list1)
value2 = sum(list2)
CSAD.append(value)
CSSD.append(value2)
list1=[]
list2=[]

hccsad=CSAD
hccssd=CSSD

#S&P 500 Energy csad and cssd
#download
L2 = pd.read_csv('c:/Users/Gavin/Desktop/s&p-sectors---energies-06-05-2020.csv')
df2 = pd.DataFrame(columns = L2['Symbol'])
for i in L2['Symbol']:
    data = yf.download(i, start="2020-02-16", end="2020-06-15")
    df2[i] = np.log(data['Adj Close']).diff().dropna()
df2=df2.drop(columns=['Downloaded from Barchart.com as of 06-05-2020 09:03am CDT'])

#Calculation
CSAD2=[]
CSSD2=[]
list4=[]
list44 = []
for r in range(df2.iloc[:, 0].size):
    for c in range(df2.columns.size):
        list4.append(abs((df2.iloc[r][c] - df2.iloc[r][:].mean())) / df2.iloc[:, 0].size)
        list44.append(((df2.iloc[r][c] - df2.iloc[r][:].mean()) ** 2 / (df2.iloc[:, 0].size - 1)) ** 0.5)
value = sum(list4)
value2 = sum(list44)
CSAD2.append(value)
CSSD2.append(value2)
list4 = []
list44 = []
ecsad = CSAD2
ecssd = CSSD2

# Save
df.to_csv('df0011.csv', index = False)
df2.to_csv('df0022.csv', index = False)
# Read
df = pd.read_csv('c:/Users/Gavin/Desktop/df0011.csv')
df2 = pd.read_csv('c:/Users/Gavin/Desktop/df0022.csv')

**Python Plot**

# plot
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.axes_grid1 import make_axes_locatable

# some random data
xxx = np.arange(len(df))
x = np.tile(xxx, (1, df.columns.size))
x = x.T
```python
yyy=[]
for i in range(len(df)):
    for j in range(df.columns.size):
        yyy.append(df.iloc[i][j])
y = np.array(yyy)

xxx2=np.arange(len(df2))
x2 = np.tile(xxx2, (1,df2.columns.size))
x2 = x2.T

yyy2=[]
for i in range(len(df2)):
    for j in range(df2.columns.size):
        yyy2.append(df2.iloc[i][j])
y2 = np.array(yyy2)

cssd_1=np.array(list33)/10
plt.plot(x, y, "o", label = "Actual data")
plt.plot(xxx, list3, color="yellow", label = "CSAD")
plt.plot(xxx, cssd_1, color="red", label = "CSSD*0.1")
plt.legend(loc = "best")
plt.title('Health Care Sector')
plt.show()

plt.close()

plt.figure()
cssd_2=np.array(list55)/10
plt.plot(x2, y2, "o", label = "Actual data")
plt.plot(xxx2, list5, color="yellow", label = "CSAD")
plt.plot(xxx2, cssd_2, color="red", label = "CSSD*0.1")
plt.legend(loc = "best")
plt.title('Energy Sector')
plt.show()
```

| Date       | Open | High  | Low  | Close | Volume |
|------------|------|-------|------|-------|--------|
| 10/02/2023 | 180  | 185   | 178  | 181   | 1200   |
| 10/03/2023 | 182  | 186   | 180  | 183   | 1500   |
| 10/04/2023 | 185  | 190   | 182  | 187   | 2000   |
| 10/05/2023 | 188  | 192   | 185  | 189   | 1800   |
| 10/06/2023 | 190  | 194   | 187  | 192   | 1600   |
| 10/07/2023 | 192  | 196   | 190  | 194   | 1700   |
| 10/08/2023 | 194  | 198   | 192  | 196   | 1800   |
| 10/09/2023 | 196  | 200   | 194  | 198   | 1900   |

*Note: The table above represents the stock price movements for a specific company over the period from 10/02/2023 to 10/09/2023.*

**Appendix B: Raw Data**

**Short horizon**

**Date       | Open | High  | Low  | Close | Volume**
**10/02/2023 | 180  | 185   | 178  | 181   | 1200   |
**10/03/2023 | 182  | 186   | 180  | 183   | 1500   |
**10/04/2023 | 185  | 190   | 182  | 187   | 2000   |
**10/05/2023 | 188  | 192   | 185  | 189   | 1800   |
**10/06/2023 | 190  | 194   | 187  | 192   | 1600   |
**10/07/2023 | 192  | 196   | 190  | 194   | 1700   |
**10/08/2023 | 194  | 198   | 192  | 196   | 1800   |
**10/09/2023 | 196  | 200   | 194  | 198   | 1900   |

*Note: The table above represents the stock price movements for a specific company over the period from 10/02/2023 to 10/09/2023.*

**Appendix B: Raw Data**

**Short horizon**

**Date       | Open | High  | Low  | Close | Volume**
**10/02/2023 | 180  | 185   | 178  | 181   | 1200   |
**10/03/2023 | 182  | 186   | 180  | 183   | 1500   |
**10/04/2023 | 185  | 190   | 182  | 187   | 2000   |
**10/05/2023 | 188  | 192   | 185  | 189   | 1800   |
**10/06/2023 | 190  | 194   | 187  | 192   | 1600   |
**10/07/2023 | 192  | 196   | 190  | 194   | 1700   |
**10/08/2023 | 194  | 198   | 192  | 196   | 1800   |
**10/09/2023 | 196  | 200   | 194  | 198   | 1900   |

*Note: The table above represents the stock price movements for a specific company over the period from 10/02/2023 to 10/09/2023.*
