# Energy Sector Stock Returns

IEOR 4307

Andrew Chen, Leon An, Cathy Wang

## WRITTEN REPORT

For our project, we selected 10 stocks in the energy sector whose corresponding companies were primarily involved in oil. We hoped to gain new investing insights into this field through statistical tests. First, we observed and analyzed the log-returns of each individual stock. Then we compared pairs of stocks to determine whether they had similar log-returns and observed their correlation.

## Data Set

Our data set consists of the daily closing stock prices of ten S&P 500® energy companies every trading day from January 29, 2014 to April 30, 2015. A total of 315 data points were collected for each stock and the daily closing stock prices were taken from Google Finance. The stocks and the companies they represent in our data set are:

SLB – Schlumberger Limited
RIG – Transocean LTD
MPC – Marathon Petroleum Corporation
HAL – Halliburton Company
XOM – Exxon Mobil Corporation
HES – Hess Corporation
DVN – Devon Energy Corporation
OXY – Occidental Petroleum Corporation
CNX – CONSOL Energy Inc.
TSO – Tesoro Corporation

The aforementioned companies were drawn from a pool of S&P 500® energy companies specializing in all aspects of oil and gas production from oil and gas exploration to refining and processing.

## Project Goals

We chose to focus on large companies within the oil subsector of the energy sector because the stock prices of these powerful oil conglomerates have recently been through many fluctuations due a precipitous drop in crude oil prices starting from the peak of about $105 a barrel in Jun 2014 and bottoming out around $47 a barrel in late January 2015. Prices have since risen to around $60 a barrel for crude oil, but they are nowhere near what they were merely a year ago. The price drop in crude oil is attributed to an increased oil supply especially through growing domestic oil production from shale deposits as well as a decrease in global oil demand as the economies of fast-growing countries like China slow down (Krauss). According to Friedman, “Nationwide, crude-oil supplies are at the highest in more than 80 years”.

Despite expert claims that oil prices will remain low for a while, we believe the oil sector to be an exciting industry with numerous potential investment opportunities for those who seek them. Therefore we hope that conducting these statistical tests could lead to a more informed decision regarding whether or not to invest in the oil industry along with which stocks and at what time to invest in.

To assist our decision making, we investigated a series of questions involving the log-returns of one or two stocks at a time.

For a single stock, the first question we asked was whether or not the log-returns were consistent with a random sample. We then tested to see whether the returns were normally distributed. Although both the previous questions were required in order to perform the rest of the statistical tests on our data, we performed the remaining tests even if the necessary criteria were not met. Next, we asked the question how well a stock performed in the past year and a half given its risk or volatility by looking at the mean return and variance of the stock's log-returns. Finally, we observed how the stock performed over time and ultimately if we could identify any positive or negative trends in its log-returns.

After observing all the stocks individually, we wanted to know whether there were any general trends or observations within the oil subsector of the energy sector that we could identify from our tests. This could be determining whether oil company stocks performed positively overall or if they were highly volatile.

For two stocks, we first asked whether or not the mean of their log-returns were equal. Then we performed a regression of one log-return on the other to see whether the two stocks were correlated or if they moved in similar patterns.

# Analysis

## Single Stock Analyses

In order to see if the stock log-returns were consistent with a random sample, we utilized a runs test for determining non-randomness. The null hypothesis is the elements in the data set are mutually independent and therefore the data is consistent with a random sample. The alternative hypothesis is the elements in the data set are not mutually independent and therefore the data isn't consistent with a random sample.

**Runs Test for Determining Non-Randomness**

| Stock | P-Value |
|---|---|
| **SLB** | 0.334379 |
| **RIG** | 0.690819 |
| **MPC** | 0.23292 |
| **HAL** | 0.334379 |
| **XOM** | 0.07833 |
| **HES** | 0.690819 |
| **DVN** | 0.394358 |
| **OXY** | 0.334379 |
| **CNX** | 0.00539 |
| **TSO** | 0.690819 |

After conducting the runs test on all of the log-returns, it was shown that all of the stocks except for CNX had a P-Value greater than 0.05. Hence we fail to reject the null hypothesis for all of the stocks except for CNX and can thus conclude that the log-

returns of all of the stocks except for CNX are consistent with a random sample. The log-returns of CNX are not consistent with a random sample.

Next we tested the assumption that every stock log-returns data came from a normal distribution. A histogram of the log-returns presented a frequency distribution that looked to be approximately normal, but it is hard to ascertain for sure with only a histogram. For that reason a normal probability plot was produced to better answer the question. The normal probability plots hinted at the possibility that the log-returns of some of the stocks such as RIG, HAL, and HES could come from a distribution with slightly lighter tails than a normal distribution. However for the most part, the normal probability plots of the log-returns showed that there are not any strong deviations from linearity and thus we can conclude that the stock log-returns came from a normal distribution.

Once the assumptions of a random sample and normality have been satisfied, we can proceed with statistical testing. The first thing we do is create confidence intervals for stock log-returns' mean, $\mu$, and the variance, $\sigma^2$. We do not know the variances of the normal distributions from which the stock log-returns came from, so we substituted in the sample standard deviation and utilized the t-distribution with $n-1$ degrees of freedom in order to find the confidence interval for $\mu$. To find the confidence interval for $\sigma^2$, we utilized a chi-squared distribution with $n-1$ degrees of freedom.

**95% Confidence Interval for $\mu$ and $\sigma^2$**

| Stock | Confidence Interval for $\mu$ | Confidence Interval for $\sigma^2$ |
|---|---|---|
| **SLB** | (-0.0019 , 0.0014) | (3e-05 , 0.00043) |
| **RIG** | (-4e-04 , 0.0057) | (0.00011 , 0.00142) |
| **MPC** | (-0.0024 , 0.0016) | (5e-05 , 0.00061) |
| **HAL** | (-0.0023 , 0.0022) | (6e-05 , 0.00076) |
| **XOM** | (-9e-04 , 0.0015) | (2e-05 , 0.00022) |
| **HES** | (-0.0019 , 0.0019) | (4e-05 , 0.00054) |
| **DVN** | (-0.0025 , 0.0016) | (5e-05 , 0.00065) |
| **OXY** | (-0.0013 , 0.0019) | (3e-05 , 0.00038) |
| **CNX** | (-0.0018 , 0.0027) | (6e-05 , 0.00076) |
| **TSO** | (-0.0038 , 6e-04) | (5e-05 , 7e-04) |

The 95% confidence interval for $\mu$ were small and showed that the mean log-return of all of the stocks were around zero with the upper and lower bounds not exceeding $\pm 1$ cent. Similarly, the 95% confidence interval for $\sigma^2$ displayed an interval that was extremely small, suggesting that the log-returns of all of the stocks possessed little variability.

We then performed a linear regression of the log-returns plotted against time for each stock.

**Linear Regression of Log-returns Against Time**

| Stock | $\widehat{\beta_0}$ | $\widehat{\beta_1}$ | $R^2$ |
|---|---|---|---|
| **SLB** | 0.00063596 | -5.64e-06 | 0.00114228 |
| **RIG** | 0.00290702 | -1.4e-06 | 2.148e-05 |
| **MPC** | -0.00101081 | 3.78e-06 | 0.00036188 |
| **HAL** | 0.00199851 | -1.296e-05 | 0.00341115 |
| **XOM** | 0.00099081 | -4.57e-06 | 0.00146166 |
| **HES** | 0.00144368 | -9.1e-06 | 0.00237158 |
| **DVN** | 0.00035325 | -5.15e-06 | 0.00062849 |
| **OXY** | 0.00155647 | -8e-06 | 0.00258259 |
| **CNX** | 0.00162239 | -7.35e-06 | 0.00110412 |
| **TSO** | -0.00199113 | 2.48e-06 | 0.00013449 |

The results indicated there doesn't appear to be a linear between these ten stock's log-returns and time as the $R^2$ values for every linear regression was pretty close to zero. A look at the scatter plots of log-returns vs. time confirms the lack of a correlation between the two parameters. The residual plots do not demonstrate any pattern either. Thus it appears that a linear regression method cannot be utilize to provide an estimation of future log-returns for these ten energy stocks, which is unfortunate from an investor's point of view.

## Two Stock Analyses

When comparing two stocks at a time, we used a confidence level of 95% and the two-sample t-test with variance unknown and assumed to be unequal to determine whether the stock returns were equivalent or not. We then used a linear regression to determine any sort of correlation or similarities in the returns. For the sake of the report, we will compare SLB with the nine other stocks.

**SLB Two Sample T-test and Linear Regression**

| Stock | P-Value | Slope | $R^2$ |
|---|---|---|---|
| **RIG** | 0.098 | 1.116 | 0.378 |
| **MPC** | 0.905 | 0.371 | 0.097 |
| **HAL** | 0.885 | 1.011 | 0.579 |
| **XOM** | 0.618 | 0.476 | 0.443 |
| **HES** | 0.838 | 0.881 | 0.620 |
| **DVN** | 0.880 | 0.909 | 0.547 |
| **OXY** | 0.642 | 0.696 | 0.545 |
| **CNX** | 0.615 | 0.649 | 0.239 |
| **TSO** | 0.334 | 0.271 | 0.045 |

After performing a series of tests on our stocks, we were surprised to find that we failed to reject our null hypothesis that the stocks had the same log returns. Stocks such as SLB and HAL possessed a p-value of 0.885 from the two-sample t-test. Thus we failed to reject the null hypothesis that the returns were significantly different. However, this is somewhat expected when comparing the companies themselves. SLB and HAL are the two largest oilfield services companies in the world, and are competitors in the same industry. Thus we would assume that macro events such as the shale boom would have similar effects on the two companies. In addition, neither company experienced any substantial or influential internal problems, creating the effect that their stock prices are mostly in sync with one another.

However, this was not the case for most of our stocks such as with TSO and RIG, which had a p-value of 0.025; forcing us to rejected the null hypothesis and concluded that the two stocks had very different log returns. Unlike SLB and HAL, TSO and RIG don't operate in the same exact sector within energy, thus we don't expect them to perform the same over a given amount of time.

Further analysis of why SLB demonstrated similar log-returns as the rest of our stocks may rest in the fact that SLB is one of the largest companies in our data set and thus is more likely to be less volatile than the rest of the stocks. In addition, as the largest oilfield services company in the world, SLB has a very diversified service and product lineup, which allows it to be the most reflective of the energy sector overall out of all of our companies.

In order to determine the correlation of the returns between two companies, we performed a linear regression between the two log-returns and look at the slope and $R^2$ value for the regression. If the slope is close to 1, then we expect to see the returns to be approximately the same over the time period. This would conclude that the two stocks returned relatively similar amounts. The higher the $R^2$ value is, the more correlated the returns were. Thus we would expect two very similar stocks to have a slope of approximately 1 and a $R^2$ value near 1.

In our actual results, we found few pairings that had very high $R^2$ values but some that had slopes close to 1. In a typical linear relationship, we would expect the $R^2$ value to be at least 0.9, which we did not observe. However, some pairings such as SLB and HAL had slopes for their regression relatively close to one, and we can see from the actual stock charts, that these two stocks ended up with around the same returns. This was expected given the volatility of stocks and the various company-by-company factors that contribute to the movement of a stock on any given day. Although as mentioned before, that macro activity may cause the stocks within the energy sector to move in the same direction over a long period, the day-to-day returns showed little correlation.

Observing our results collectively, we can see that stocks that had low P-values also tended to have low $R^2$ values and slopes that were not close to one, and stocks with higher P-values tended to have higher $R^2$ values and slopes closer to one. This is also

expected as stocks that are highly correlated would have a stronger linear relationship and move at the same rate.

In addition, the residual plots against time and the residual plots against log-returns demonstrate no obvious patterns, which suggest that a linear regression is the appropriate type of regression for the data.

## Conclusion

From our tests, we could first conclude that, for the most part, our log returns were both consistent with a random sample, and also normally distributed, thus allowing us to conduct further hypothesis testing on the data.

As for our tests regarding the log returns, it was very difficult to generalize our data. From the tests, we could see that the log returns and thus the actual returns varied significantly from stock to stock. There are also weak signs of linearity between the log returns and time, thus it is difficult to use any sort of linear prediction to select long or short investments.

Comparing two stocks at a time, at a confidence level of 95%, we could not significantly prove any differences in log returns when comparing SLB to the remaining 9 stocks. The slopes and $R^2$ values for our two stock regressions also showed varied results, with some slopes ranging from 0.271 to 1.116, and $R^2$ values ranging from 0.045 to 0.620. Once again, it was hard to generalize any trends for the energy sector as a whole using these tests and our low $R^2$ values rejected the possibility of any linear relationship between log returns of any two given stocks.

Overall, our stocks proved to be much more inconsistent than we first imagined. Even though all the companies operated in the same industry, their log returns were very different over the past year and a half. Unfortunately, because of the variability of the stocks themselves, it is difficult to recognize any significant patterns in our results and thus it would be difficult to use our tests to make any investment decisions regarding the individual stocks.

# References

Friedman, Nicole. "Oil Prices Slip on Strengthening Dollar, Ample U.S. Crude Supplies." WSJ. Dow Jones & Company, Inc., 04 May 2015. Web. 04 May 2015. <http://www.wsj.com/articles/oil-prices-move-to-five-month-highs-on-european-factory-data-1430735189>.

Krauss, Clifford. "Oil Prices: What's Behind the Drop? Simple Economics." *The New York Times*. The New York Times, 12 Jan. 2015. Web. 04 May 2015. <http://www.nytimes.com/2015/01/13/business/energy-environment/oil-prices.html?_r=0>.