Peter Manning

Econ 191

"How does the release of a new drug by a firm tend to impact the firm's share value?"

Pharmaceutical companies do not have the reputation for outperforming, let alone keeping up with average stock performance of something like the S&P 500. In fact, many biotech and pharmaceutical corporations face stagnation or huge losses during times when top finance and tech stocks are trending upwards with leaps and bounds. The reason isn't surprising, afterall, most projects taken on in this sector are considered money pits in that tey garner 0 revenue and provide nothing to show for years of work and dedication before a drug release, let alone a successful one. But the one-in-a-hundred that pull through and release a drug into the market are usually in for sizable uptick in demand for their stock, or at least one would assume. My curiosity lays in the entirety of this process, however. I want to look into how announcements and discussion generated from these affect prices, and if the release of the product tends to live up to expectations or not. I am curious as to situations where the release of a new drug could actually be damaging to share value, as in scenarios where people react negatively to a substance and see if this is reflected in the stock. I will attempt to perform regression analysis of the effect on releasing drugs or possibly a difference in difference approach for drug companies side by side to evaluate outcomes. As I learn more about the nuances of these situations I will likely find new variables to control for, but I will have to give this more thought. Additionally, it seems some people have beat me to researching this topic, so I will look into their literature and articles surrounding this. Many finance and consulting firms like McKinsey, Harvest, and Deloitte have conducted some data exploration into these subjects and found results for interesting ideas such as first-to-market advantage in the pharmaceutical industry, and how often being first to release a new kind of product or an innovation via a better alternative product leads to long-term market domination. I will use commonly available stock market datasets such as yahoo and google finance to source price data for the stocks i'm interested in. Through my early preliminary browsing, it seems stock market data isn't hard to find.

Market watch has csv files for every single stock on the nyse as well. All this goes to show my question is in a field with lots of readily available and quality datasets. The challenge I will face is limiting the sope of the data to a manageable size in a format I can do something insightful with, as my approach will be much more technical than some of my peers who opt to just grab a ready-made dataset off the web.

I also want to look into how US government subsidizing the use of certain drugs through medicaid (reimbursements for hospitals, etc) affects success of a company, which I can do via this dataset "State Drug Utilization Data 2023", (source: https://catalog.data.gov/dataset/state-drug-utilization-data-2023)

and then cross reference the company names that appear in this data set with the overall stock performance ones and look at average growth of those companies and the ones not used by the government and received no payments from medicaid. This is another area where I may choose to use a difference-in-difference approach and show parallel trends, selecting two stocks that had similar performance up to the time one of them released a new drug and compare the aftershocks.

Current literature gets too specific on drug categories and nuanced nit-picking of data, such as in this article: https://jamanetwork.com/journals/jama/fullarticle/2792986
My goal is to provide a more comprehensive look at the big picture that the average person would more readily be able to internalize, as opposed to controlling for launch prices for orally administered drugs vs other, etc. If the research process permits I'd prefer to keep things simple.

Problems with the research process include having to cross reference information from multiple data sets, alongside there not being a great comprehensive data set I could find regarding pharmaceutical company drug announcements/ launch dates & info, meaning i'd have to do manual digging on a case-by-case basis. I expect there to be bumps in the road like this, though. Oxford university put out an article on this exact topic and without going into detail about what I didn't think was done perfectly, the article is from 2011, and a lot has changed since in both medicine and finance.

Being still in the early stages of formulating my approach to conducting research on the effects of drug releases and announcements on respective pharma stocks and looking for signs of ripple effects across industry stock values, I am still testing out various ways of looking at stock performance over time. My first dataset is yfinance, an incredibly comprehensive dataset containing stock tickers and associated information such as open, close, highs and lows, volume, and Dividends for some companies as far back as 1962. I may consult other datasets if my research diverges from just stock pricing and volume, but for the time being these will be metrics I'm most interested in surrounding announcements. This is in particular one issue that I am trying to find a way to automate. Robinhood.com lists 458 pharmaceutical companies available to trade, and obviously I do not have the time, patience nor desire to find and confirm announcements and releases of drugs and products by these companies, so my solution will either involve some shortcutting if I can't find a way to consolidate a list of dates when announcements and releases occurred in history. There is a dataset called Drugs@FDA that has information on drug releases but I have had difficulty accessing it and getting it into my python environment.

On the successes side of things, I have constructed a few time series objects and have familiarized myself with the process of filtering the data frame to specific columns and rows, and what the granularity of the dataset is, a single day for one stock. Each of these days are formatted as year-month-day-00:00:00-05:00, meaning there is probably a way to get the granularity down to the second, but I have not figured out if its possible yet. It doesn't matter much though because I'm more interested in daily movement. As for the structure of the yFinance dataset, the shape is 15658x7 for $KO/$COKE, meaning that there have been 15658 active market days, and each day has 7 different data points of which i mentioned earlier (volume, daily highs/lows, open & close values) as well as dividend and splits which are usually 0 since these events are infrequent.

As for a more in depth look at my research design, I have decided to Implement a couple different models to compare and contrast, firstly analyzing the abnormal returns estimate, a classic economic indicator of a stock's value differing from what the formula expects it to be solely on price data, meaning world events and sentiments were a confounding variable at play.

The AR formula is given by $AR_{it} = P_{it} - PE_{it}$

Where ARit is the difference between observed price P_{it} of stock index i at time t and expected price PE_{it} of the same stock at time t based on an expectation formula- Capital Assets Pricing Model or the market model, which I will weigh and try out both. CAPM, given by

$E(R_i) = R_f + \beta_i (E(R_m) - R_f)$

estimates return of an investment using bond return rate, market rate, and some algebra to find a beta term. The first term R_f denotes the risk free rate, which is the typical yield of “risk free” Treasury bills with a 3 month or 1-year maturity. We then find beta, which denotes the investment’s sensitivity to changes in the market at large. Beta > 1 tells us that the stock is more volatile than the market, and $0<\beta<1$ means it is less volatile than the market. The quantity beta is a coefficient of is a difference between the aforementioned Risk-free rate and Expected market return, which we choose arbitrarily or use data to assess over a certain period of time of our choice. I think I will choose to assess the returns 12 months before vs 12 months after, and take differences between these figures and the same methods applied to other pharma stocks which did not release a drug in these time frames. Using a 12 month time frame will allow control for routine seasonal dips in the market whereas showing the difference between 3 months before and after would be capturing assumed systematic differences in the market, between, say, spring and winter.

The market model, on the other hand, denoted by:

$R_i = \alpha_i + \beta_i R_m + \epsilon_i$,

where return of asset i = intercept term alpha + beta coefficient * return of market portfolio R_m added to an error term which captures the portion of the asset’s return not explained by market returns. Looking at this, it is your common least squares regression, predicting stock change given market fluctuation over a certain time period. This formula feels more comfortable to me than the CAPM model but in actuality may more difficult to implement. The use of these approaches will revolve around manually finding dates of drug releases and conducting the experiements within a fixed before-and-after window.

A second approach I want to use to evaluate potential increases in stock price after release of a drug is using a Difference-in-Differences approach. Using a formula such as:

$$Y_{it} = \alpha + \beta DrugRelease_{it} + \gamma Time + \delta(DrugRelease_{it} * Time) + \epsilon$$

Which has a similar idea of capturing the difference between two or more stocks' difference before and after a given release if one is present. the formula is a bit redundant in its terms due to the fact that we need two dummy variables which encode a true or false for drug release and if we are in the before or after period. S a company which released the drug by time t will have a DrugRelease value of 1 which will allow us to solve for beta and delta using regression or a different analytical approach potentially. If we are analyzing a before period, and are talking about a company which did not release a drug, we will encode these two variables to 0 in which case the error term and alpha values will be absorbing the effects. This system allows us to analyze several coefficients simultaneously and differentiate between the effects and their differences in two time periods given two different starting states. To employ this approach, I will have to find two stock following similar trends up to the release date/time, and then I can manually and analytically determine the DiD (by eyeballing change in time series data graphs' lines, and using an ordinary least squares regression through ssci-kit learn or stats models, two stat libraries I have used a bit in python.

In order to use these models, I will need appropriate variables and their data or find/ create proxies for them. I don't think I will be encountering any issues with this, as i will be estimating price or price as a percentage of growth in a time period, data we have, estimated using functions of price, dummy variables such as in the DiD model (i can create a column such as "DrugRelease" filled with 1s or just encode the dummy variables as true if ==stock that had a release (which i will have to know previously through research) and 0 if it is another stock. Lastly, I will simply fetch values for the risk free rates and market return rates for the CAPM formula and market model formula to be used in the AR formula's PEit term.

Findings - Overview of Data And Initial Results

To conduct my analysis into whether the drug release affects a pharmaceutical/ medicinal company's stock price, I first engaged yahoo's stock market information accessible via python in the package "yfinance". My next step after importing it was to obtain a list of stock ticker symbols I could regress on, given that they were in the pharmaceutical industry. I found a list from topforeignstocks.com, and copied them into my jupyter notebook. I constructed two primary functions to input the data, one that creates a data frame of multiple stocks and their weekly average prices over a ~7 month window (chosen arbitrarily at 2 months before drug release date and 5 months after), and another function that takes in the dataframe, using feature engineering and difference in difference regression with one stock as the treatment and any number (I chose one, but will be testing a larger quantity in my next research period) to select as the control. The function outputs a summary table of the regression results, a residuals plot, and a line plot comparing the two stocks' prices to ensure they follow similar trends before the treatment date. The following table is the output of the OLS summary, but to provide context before interpreting the output of the OLS summary, I'll briefly detail the specific code and variables I used to achieve these results. My function I detailed above takes in a list of stock symbols, and one must be used as treatment. I chose to compare $PFE (pfizer), and GSK's stock, which is also $GSK. I fitted the difference is difference formula ($Yit = \alpha + \beta DrugRelease_{it} + \gamma Time + \delta(DrugRelease_{it} * Time) + \epsilon$) to my dataframe which, for sample followed the format:

	Date	WeekAvgPrice	AfterDate	Treatment	TreatmentxAfter
PFE	2020-10-12	32.230046	0	1	0
PFE	2020-10-19	32.291018	0	1	0
PFE	2020-10-26	32.338687	0	1	0
PFE	2020-11-02	32.431104	0	1	0
PFE	2020-11-09	32.519475	0	1	0

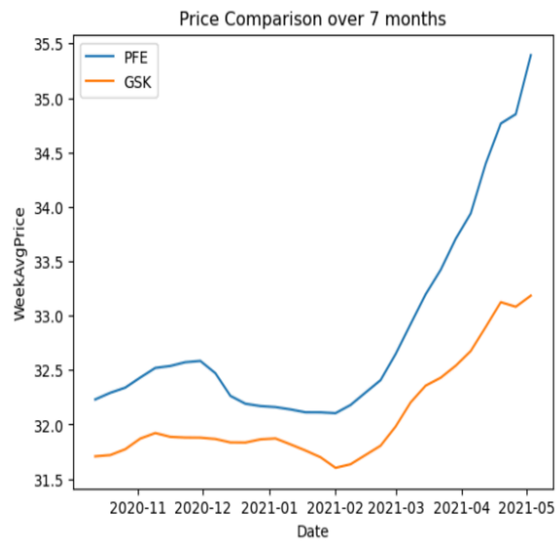
Y_{it} was assigned to the Weekly Average Price column "WeekAvgPrice", and is the dependent variable we are regressing on, to be predicted by the following components of the formula. I set $\beta DrugRelease_{it}$ to the binary indicator column of "Treatment" which naturally defined which indices equalled the treatment stock, and with Pfizer being my treatment, the values for the column were 1. $\gamma Time$ was the label for indicating if the corresponding row's date was before, or after the set release date. Lastly, $\delta(DrugRelease_{it} * Time)$ is the interactive term which only takes value of 1 in the case that the corresponding row is the treatment stock in the after release date period, which will help us determine the effect of the treatment, whether it affected stock prices positively or negatively.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           WeekAvgPrice   R-squared:                       0.286
Model:                            OLS   Adj. R-squared:                  0.248
Method:                 Least Squares   F-statistic:                     7.475
Date:                Tue, 23 Apr 2024   Prob (F-statistic):           0.000271
Time:                        02:41:44   Log-Likelihood:                -63.381
No. Observations:                  60   AIC:                             134.8
Df Residuals:                      56   BIC:                             143.1
Df Model:                           3
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept          31.8336      0.240    132.585      0.000      31.353      32.315
Treatment           0.6080      0.340      1.791      0.079      -0.072       1.288
AfterDate           0.3528      0.287      1.229      0.224      -0.222       0.928
TreatmentxAfter     0.2231      0.406      0.550      0.585      -0.590       1.036
==============================================================================
```

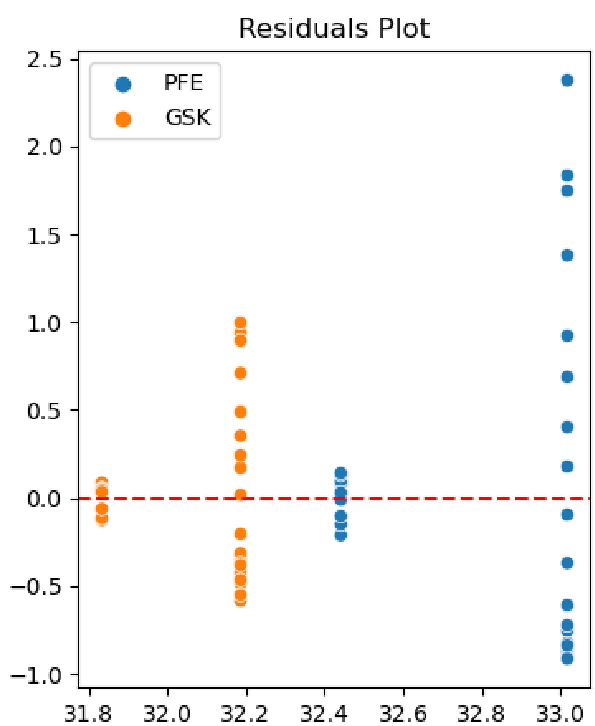
In the output of the OLS Table, we can interpret the coefficient's values and their implications. Intercept, which corresponds to α in my formula, sits at 31.8336. Given both stocks were roughly centered around price of $32 at time of treatment (defined as December 11, 2020, when Pfizer received FDA Approval for their covid vaccine), the magnitude of this coefficient makes sense, and the standard error (0.24) being so much lower than the coefficient gives us a large t-stat and therefore a very statistically significant interpretation. β = 0.608 implies that our model determined that the treatment lead PFE stock to increase $0.60, although more research and causal inference would have to go into determining if the release of the covid vaccine was a total or partial cause of the stock price jumping $0.608. All we can really infer here is that PFE's predicted price in the time frame before treatment is $\alpha + \beta$, and $\alpha + \beta + \gamma + \delta$ in the after period. β's standard error is a bit lower than we expect to see for strong significance (coefficient / std. error >= 2) and the t-stat is 1.7, meaning our coefficient result is 1.7 standard deviations away from the mean. At a p value of 0.05, which is standard in statistics, we would need our coefficient to have a t-stat of at least 2 in order to reject the null hypothesis that the treatment had no effect on stock price. However, the t-stat being >= 2 requirement is a bit arbitrary, and 1.7 is still a relatively strong result- the strongest out of all of our coefficients besides intercept. Next, interpreting the coefficient "AfterDate" (γ) follows the same procedure. The coefficient value 0.3528 tells us that irrespective of the treatment group, the average price increased $0.35 from the before treatment time period, inclusive of both stocks. Its std error being 0.287 leaves us with a less than significant t-stat of 1.229, meaning that in the time frame window I've chosen of ~5 months after time of treatment, we cant be too statistically certain that the result we've gotten of a $0.35 increase over time is due to random chance or not. If our t-stat was a higher value, we'd know that there is at least a 95% chance that the time period of "afterDate" increased our average price by this much due to time alone instead of random chance, but in this case the t-stat informs us that there is not a very significant time trend across both groups.

Lastly, we must interpret our Treatment * AfterDate interaction term. The coefficient for this indicator column is 0.2231 with a standard error of 0.406, and without repeating the prior reasoning we can infer that this event is not very significant- the t-stat is around 0.5, which is the weakest yet. However, the variable is not totally useless, even given this information. We can still learn that the treatment effect had a slightly positive effect on the treatment group based on our model. The term in general displays the differential effect of the treatment over time. A positive coefficient suggests that the treatment had a more significant positive effect on the outcome variable for the treatment group compared to the control group over time.

Now that we've cleared the statistical analysis for this specific instance of treatment vs control given these specific stocks and this date, we can look at some data visualization I've employed to further my understanding of the results. This first graph was helpful in my understanding of why the coefficients took their values by plotting the two stock's price change over the defined timeframe. I also used it as a precursor to the regression itself to ensure that the similar trends requirement was met.



Next, I outputted a residual plot to understand where the actual values lay in relation to where the predicted values were. Even though we aren't inherently predicting price with our model, the residual plot is still useful for showing the spread of the values

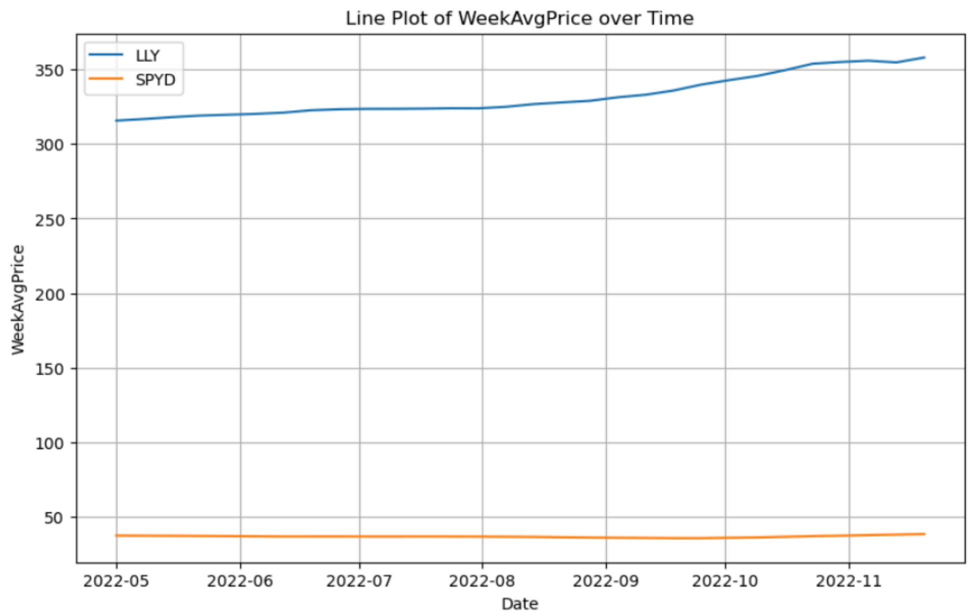


for both stocks in both time periods. The graph is a bit difficult to interpret so I will provide my observations. The GSK labels in orange indicate that in the before treatment category (The cluster to the left) exhibit well behaved residuals given their tight spread and lack of outliers. In both stocks, the after category, which is the right-hand-side cluster for each group, displays comparatively more variance which could be due to the model being less than perfect, or the data being more volatile. The residual plot also tells us that PFE's spread is very large in the after treatment period, considering the large range of y values it occupies as well as the presence of outliers. I attempted to adjust the parameters of the model with log transformation, higher order polynomial transformations and trying different solver types in the smf.ols regression's ".fit" parameter, but it did not yield different results. I think the data for this time period was just more volatile, which might also explain some of the lower significance of the coefficients from earlier analysis.

Lastly, I constructed a QQ plot, which compares the values in the dataset to a version that the machine generates which follows the normal distribution, and it is a way to ensure that my regression is an optimal choice under the assumption of normality in the residuals. In the case of perfect normality, the points will fall along the 45 degree line, which these did to greater extent than expected based on the less than ideal results of the coefficients' significance and the spread of the residuals plot. The qq plot's shape informs us that our data is quite normally distributed upon normalization, with a bit of skew at the tails (points curving upwards at the end end imply heavier tails- more likelihood of extreme values compared to the normal distribution. Here's the plot:

While I am satisfied with the findings of this stage of my research process, my next step will be to conduct repeated experiments with different stocks and time periods, and see if there is truly a correlation between drug release and stock price. I also intend to implement the market model comparison and construct a few additional metrics beyond what I currently have.

The next case study I wanted to look at was Eli Lilly's drug for the generic Tirzepatide, called Mounjaro, released May/June 2022 (I opted to place the treatment on June 30 so as to give some leeway for the drug release to hit the market). The drug was Intended to treat blood sugar regulation for patients with Diabetes, but subsequently enhanced weight loss in these individuals, leading to its explosive popularity. To perform my regression analysis once again, I needed a control stock with similar trends up to the release date. I picked $PRGO, due to their similar trend line up to the date I picked, and whether by cause or coincidence, the PRGO line stayed roughly flat with a bit of decrease over the time frame, while Eli Lilly shot up around the time of release I picked. As a demonstration to compare trend lines pre-treatment, here is a time series analysis of their side by side prices throughout 2022:



Here is the Ordinary Least squares regression results for Stock price being regressed on:

*treatment + time + treatment * time:*

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           WeekAvgPrice   R-squared:                       0.997
Model:                            OLS   Adj. R-squared:                  0.997
Method:                 Least Squares   F-statistic:                     7243.
Date:                Fri, 10 May 2024   Prob (F-statistic):           1.83e-72
Time:                        19:41:47   Log-Likelihood:                -206.06
No. Observations:                  60   AIC:                             420.1
Df Residuals:                      56   BIC:                             428.5
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         36.9454      2.589     14.270      0.000      31.759      42.132
Treatment        282.7231      3.661     77.218      0.000     275.388     290.058
AfterDate         -0.4047      3.094     -0.131      0.896      -6.604       5.794
TreatmentxAfter   18.0963      4.376      4.135      0.000       9.330      26.863
==============================================================================
Omnibus:                        8.262   Durbin-Watson:                   0.246
Prob(Omnibus):                  0.016   Jarque-Bera (JB):                8.310
Skew:                           0.642   Prob(JB):                       0.0157
Kurtosis:                       4.294   Cond. No.                         8.95
==============================================================================
```

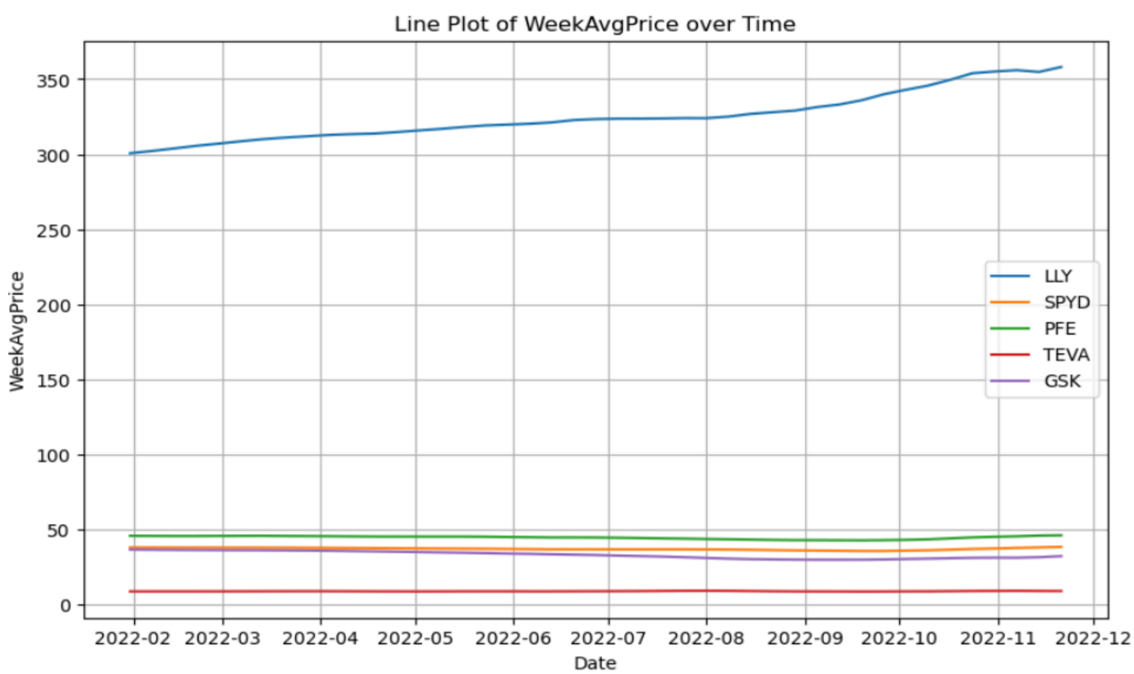
The results in this report display higher statistical significance overall with the t-stats being greater than 2 in all variables except the Treatment Time period indicator. While performing a causal inference to determine effect of the treatment here would be out of scope of my abilities and technical practicality for this project, I think by looking at our results there is something to interpret here and the release of this drug did positively impact price. Confounding variables when talking about stock prices and their influences could be infinite in quantity and scale, but to name a few, we would have to consider state of the market, other drug releases, public sentiment, and what is and isnt “priced in” to the stocks already, although this isn’t exactly a measurable metric.

To backtrack to some of the strategies I was going to use to determine impact of the release including the abnormal returns equation that incorporates either CAPM formula or market model formula, became a bit of a trivial and insightful metric in hindsight. My intuition that told me to use it stemmed from the fact that one could verify how much the stock deviated from the rest of the market and interpret that coefficient how one would, or compare it with a control stock using the same formula. The function of this resulted in just a simple value that provided no real insight into the statistics behind the variables at play as my regression did, and I also realized that this formula was better applied in a live setting, where predicting this value in the short term could help determine whether an asset was over or under priced and execute a sell or buy using this information. So, I opted not to show the results of this as they werent insightful beyond what you could get by comparing the price lines in hindsight, and they also weren’t your comprehensive, traditional econometric models.

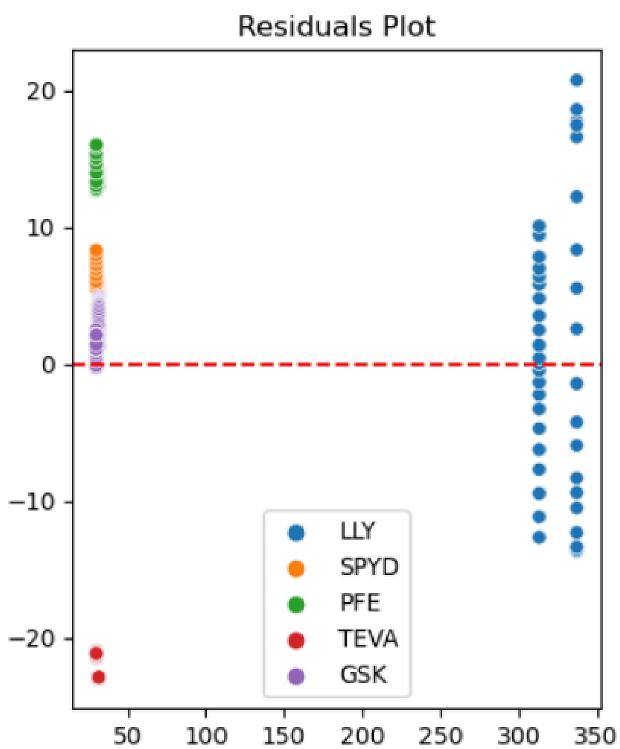
Before I go on to rapid fire some more regression results under different stocks, timeframes, and conditions, I want to explore using my regression model to utilize multiple regression in the form of multiple levels of the control variable (ie, using more than one stock as the control group, but keeping the treatment variable at one, naturally). To accomplish this, I built a customized Pipeline function that takes in x number of stocks as a list, with one of the elements being the treatment, and a date parameter, and the function iterates over the list, converting the strings of stock tickers into yFinance ticker Objects, and creating a dataframe of the weekly average prices for each stock in the list over a period of time that I chose to be 150 days before/after the treatment date this time. I reran the LLY regression with three stocks as control, being SPYD (the s&p 500 index fund), GSK, PFE (Pfizer), and TEVA, and I obtained this regression output:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           WeekAvgPrice   R-squared:                       0.988
Model:                            OLS   Adj. R-squared:                  0.988
Method:                 Least Squares   F-statistic:                     5936.
Date:                Sat, 11 May 2024   Prob (F-statistic):         1.96e-203
Time:                        01:00:49   Log-Likelihood:                -853.61
No. Observations:                 215   AIC:                             1715.
Df Residuals:                     211   BIC:                             1729.
Df Model:                           3
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept          31.6094      1.380     22.906      0.000      28.889      34.330
Treatment         281.6879      3.086     91.289      0.000     275.605     287.771
AfterDate          -1.6156      1.975     -0.818      0.414      -5.508       2.277
TreatmentxAfter    25.6784      4.415      5.816      0.000      16.974      34.382
==============================================================================
Omnibus:                       30.741   Durbin-Watson:                   0.089
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               24.154
Skew:                          -0.720   Prob(JB):                     5.69e-06
Kurtosis:                       2.209   Cond. No.                         6.79
==============================================================================
```

The t-stats overall display high statistical significance in all except the after-date, which is the one coefficient in difference in difference regression that in theory has no connection to the treatment effect- The time indicator is irrespective of treatment/ control stock and therefore makes the assumption of no treatment. It's magnitude informs us how the passage of time on average affected the dependent variable, and given that it is negative and statistically insignificant, we can still infer the treatment effect without issue from the other coefficients. I wont go much more into detail on the inference to avoid redundancy, but an important takeaway is that the addition of many control variables helped the strength and significance of all of the variables except time indicator, which could be explained by the fact that the market took a downturn in 2022, which I remember all too well as I started investing not long before the crash. Taking a look at the price trends for both the treatment and control group informs us that, counter intuitively, prices did not take any dramatic hits, as the prices of all except the treatment remained quite steady. But recall that the magnitude of the Time indicator coefficient



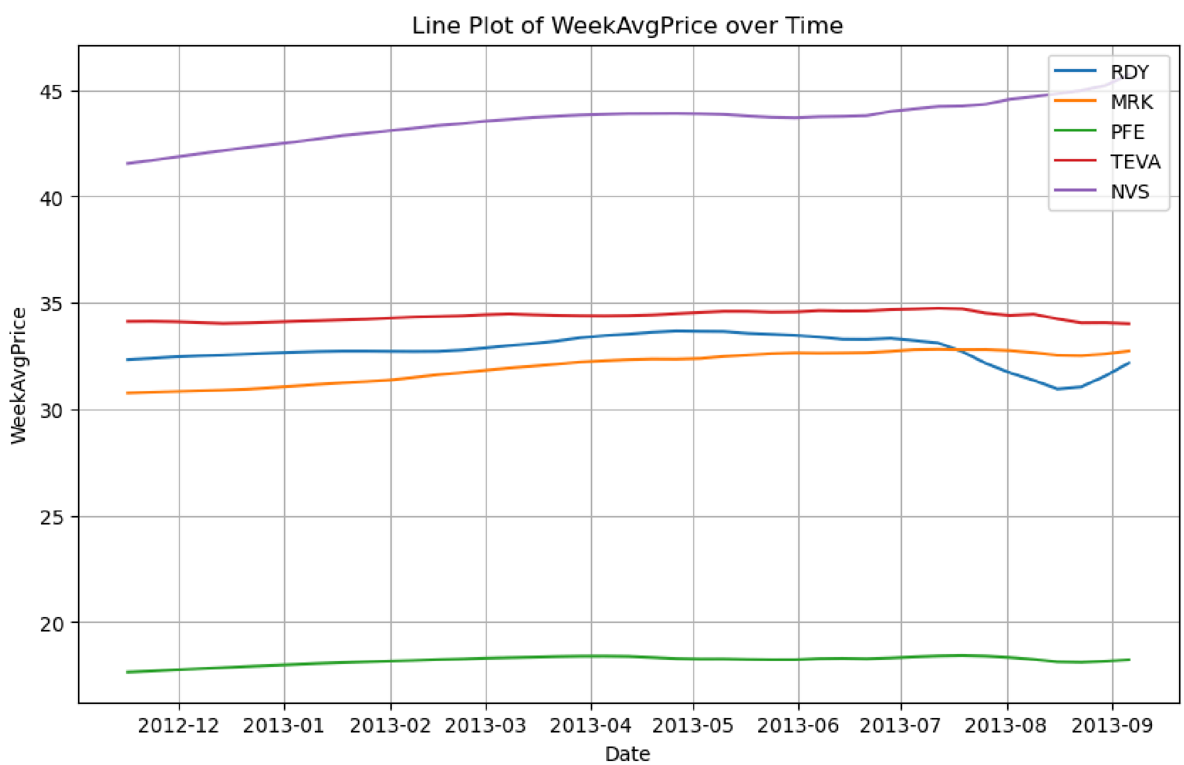
was only -1.615, meaning that the passage of time only tended to impact prices by -$1.62, which was certainly held up by LLY's huge price increases alone throughout 2022. When thinking about causality once again, It becomes important to weigh other confounding variables and factors into what made LLY rise so much in price in 2022- was it other drug releases? Why was Eli Lilly so resistant to a general state of market stagnation/ deflation in 2022? Why did customers think Eli Lilly was worth so much more than what it had been at in 2021 in a time when many customers were selling their other assets? Was a drug release that impactful on prices, or was LLY just riding the waves of success it has been experiencing since 2018, where it started seeing rapid increases in rate of return. I'll leave this discussion with this nicely colorful



residual plot:

I think the plot demonstrates precisely why this date was a good pick for LLY as a treatment variable: The before and after time periods of the LLY stock in blue are more distinctly separated in their average price by period, meaning that its price change over time was rather significant, while the other groups are noticeably compact in the absolute distance between their two groups, partially due to lack of movement and partially due to the scale that the price of LLY has imposed- its $300+ price point forces the other groups close together due to proportional change in price compared to the max of the graph at 350 on the x axis.

One final comparison I performed compared NVS (Novartis) against PFE, MRK (Merck & Co.), RDY (Dr. Reddy's Laboratories), and TEVA. I performed my testing centered around april 2013, when Novartis released Tafinlar, a "kinase inhibitor" for the treatment of melanoma and lung cancer. To perform the sanity check of trends similarity before treatment, we have this timeseries graph of prices once again:



which demonstrates relatively similar price changes before April by eyeballing it- unfortunately I did not think to implement a metric of average slope, but I may if I revisit this project in the future, which would help mathematically ensure that trends were similar enough within some margin of error that I think is not really defined in econometrics as a hard rule, but more of something you tend to eyeball. The regression report for this instance was as follows:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           WeekAvgPrice   R-squared:                       0.488
Model:                            OLS   Adj. R-squared:                  0.481
Method:                 Least Squares   F-statistic:                     67.10
Date:                Sat, 11 May 2024   Prob (F-statistic):           1.66e-30
Time:                        02:16:12   Log-Likelihood:                -684.27
No. Observations:                 215   AIC:                             1377.
Df Residuals:                     211   BIC:                             1390.
Df Model:                           3
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept          29.1515      0.628     46.435      0.000      27.914      30.389
Treatment          13.7706      1.404      9.810      0.000      11.003      16.538
AfterDate           0.3894      0.898      0.433      0.665      -1.381       2.160
TreatmentxAfter     0.9214      2.009      0.459      0.647      -3.038       4.881
==============================================================================
Omnibus:                       36.648   Durbin-Watson:                   0.071
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               52.183
Skew:                          -1.204   Prob(JB):                     4.66e-12
Kurtosis:                       2.833   Cond. No.                         6.79
==============================================================================
```

The reader can interpret this result table using the process of the preceding examples, which I believe I have covered extensively enough. However, one important takeaway that the results exhibit is that in this example, both the interaction term and time indicator term are weak in significance and magnitude compared to treatment/intercept terms. This example is therefore indicative of a case where the drug release over time was not particularly and comparatively impactful with respect to the coefficient of intercept (the average price of each stock with a 0 value in time indicator and treatment stock- ie the control stocks in the before treatment period).

Holistically, the findings of this study underscore the influence of pharmaceutical companies' drug releases on their respective stock prices. Through a randomized selection process, technical analysis revealed a semi-consistent pattern wherein the announcement/ release date of a new drug tends to positively correlate with an increase in the company's stock price- there were instances in my research where this was not the case, but in these few instances it was difficult to match one or more control groups with the treatment, so attempting difference in difference regression would have been unfeasible. The empirical distribution of my procured results suggests that investors perceive the introduction of a new pharmaceutical product as a positive indicator of future financial performance and market strength for the firm given that the increase in price was at least partially in response to real world events including the actual release. Such reactions may be derived from expectations regarding the drug's ground-breaking or otherwise distinct efficacy and ability, market demand, and revenue potential, all of which contribute to an investor's confidence and hence company valuation. Moreover, the observed market response implies a convoluted but real relationship between pharmaceutical innovation and financial markets, which is part of understanding how healthcare advancements influence our world beyond the benefits (or detriments) to patients alone. However, it is essential to acknowledge the limitations of this study, including the potential for confounding variables and the generalizability of findings across different market conditions and contexts. Future research endeavors may benefit from exploring additional factors that could influence stock price reactions to pharma events, such as drug bans, clinical trial outcomes, and how funding and sponsorships may influence these factors as well as price. Ultimately, I believe my findings are an interesting look using real world data at the interplay between scientific innovation, market dynamics, and investor behavior relating to the pharmaceutical industry.

Sources:

yFinance Dataset by Yahoo Finance

https://topforeignstocks.com/stock-lists/the-complete-list-of-major-pharmaceutical-stocks-on-the-nyse/

fda.gov/drugs