

# **Venture Capital:**

A study of robust features that influence change in industry  
specific deals

Andrew Koe and Harrison Zhao  
EID300A

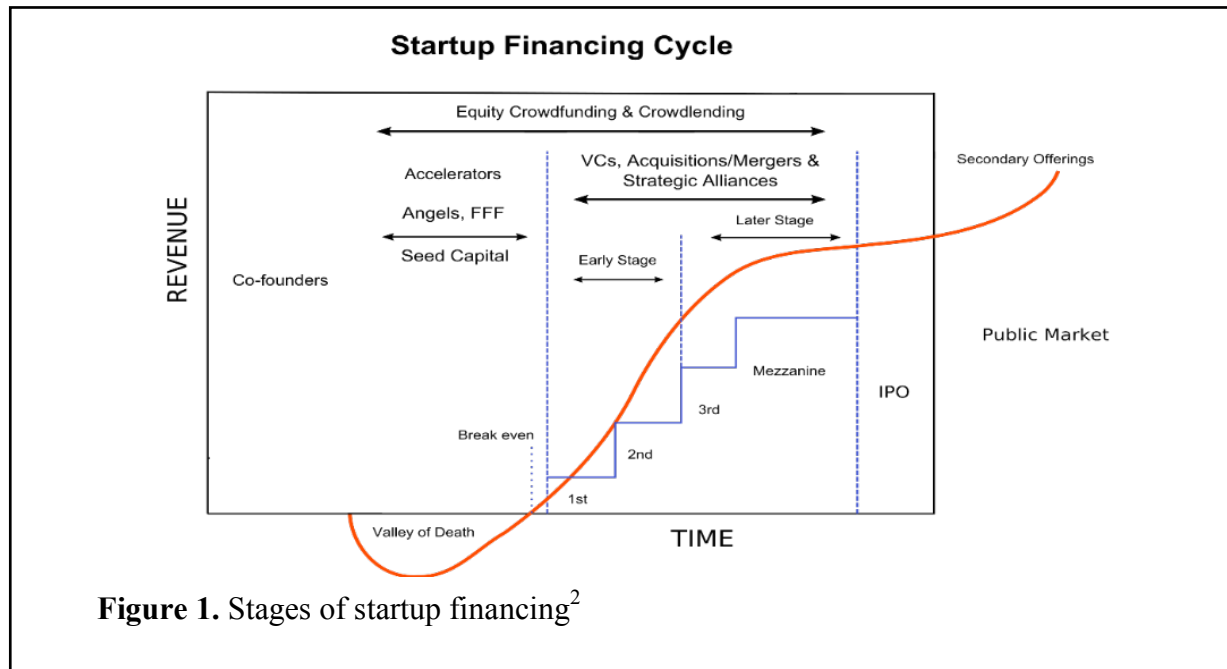
## **Introduction**

Stock indices such as the S&P 500, NASDAQ, and Dow Jones Industrial Average are used as a metric to study how the overall stock market is performing. These indices essentially take weighted moving averages of stock prices of select publicly traded companies whose performance is thought to be indicative of the greater stock market. Traders aiming to use statistical arbitrage techniques use the performance of stock indices to predict which way certain stocks will move in the near future. Analysts also use index performance as a way to study investor sentiments and as a way to infer economic health.

In this paper, we present a novel use of stock index data. We use stock indices for various sectors such software, biotechnology, industrial, and finance and study how they affect funding and investment in startups and smaller private companies. By performing regressions on stock index data we were able to closely model investment trends in startups of various sectors. We demonstrate the causal relationship between changes in sector specific index performance and private investments in those sectors.

## Background

### *Startup Funding Stages*



There is a long journey between the inception of a company and the initial public offering. Figure 1 shows the different stages of the startup up financing cycle along with the typical revenue of the company at each stage. The typical stages are:

- 1) The first stage is labeled co-founders. At this stage, the company is simply at the inception stage and the founders are aiming for proof of concepts and prototypes. Funding at this stage is most often provided by friends and family.
- 2) The next stage is called the seed funding stage. Companies that make it to this stage begin to be funded by angel investors. These investors are usually wealthy individuals who invest their own money in companies in return for convertible debt or equity. As Figure 1 shows, the seed funding stage is a high risk stage for investors. At this stage, companies are most likely not generating revenue rather, as the graph shows, companies may be losing money while they are hiring and trying to scale.

Because of the risk associated with this stage of investment and the likelihood of equity stakes becoming diluted in future stages, angel investors typically demand a higher return.<sup>1</sup>

- 3) The next stage is the early stage. Companies that make it to this stage attract the interest of venture capitalists. Venture capitalists invest with funds pooled from many investors. They are still taking on significant risk at this stage and will also invest in exchange for equity. Only 3% of companies in the venture capital universe generate 95% of the industry's returns.<sup>1</sup> Venture capitalists are also expected to bring business development and management expertise and advice to the companies they invest in.
- 4) At the late stages, companies that make it through the early stage have the potential to be acquired by a larger company or be taken public in an initial public offering. Investment banks typically enter at this stage to handle mergers and acquisitions or to act as bookrunners.

Statistics show only a small minority of startup companies make it through all the stages of the financing cycle. As a result, investments in startups are incredibly high risk investments and as angel investors and venture capitalists are unlikely to make a return on most of their investments.

### *Stock Indices*

Stock indices are measurements of the value of a section of the stock market. Indices are mathematical constructs and various indices use different metrics and techniques to calculate the value. The most commonly used indexes are composite indices that include stocks from various sectors of the economy. These indices are used to gain insight into the overall performance of

the stock market. However, there are also sector specific indices that use similar methodologies. For example, the Dow Jones Industrial Average is a price weighted index. The index only takes into account the prices of the 30 component stocks when calculating the value. This has the benefit of closely tracking movements of all the component stocks since the price movement of a single security can affect the value of the index. However, this index does not take into account the relative sizes of the companies. This can lead to misleading index values if the larger companies are performing well but a small company sees its stock value fall.

Another set of widely used indices are the NASDAQ indices. These indices are market-capitalization weighted. This weighting technique allows the index to better reflect the health of the market since smaller companies will not be able to move the value of the index as much as larger companies. The formula for the index value is:

$$\frac{\text{Aggregate Adjusted Market Value}}{\text{Divisor}}$$

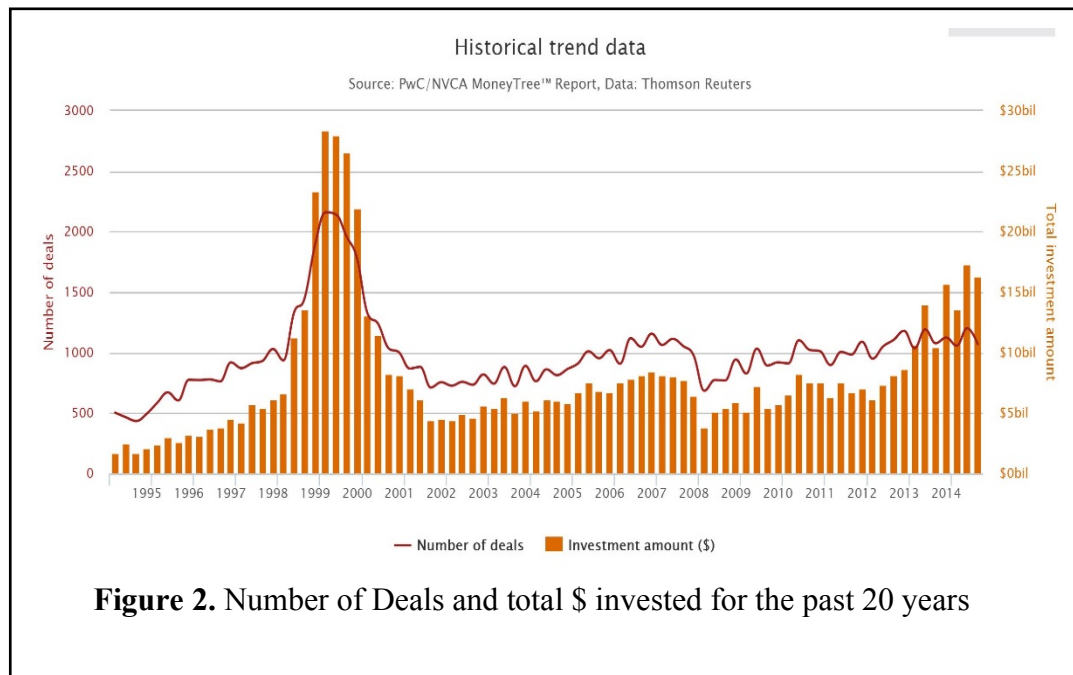
where the divisor is defined as:

$$\frac{\text{Market Value after Adjustments}}{\text{Market Value before Adjustments}} \times \text{Divisor before Adjustments}$$

The limitation with the NASDAQ indices are they only use components that are exclusively listed on the NASDAQ exchange.<sup>3</sup> The NASDAQ set of indices were used in this paper for the regression studies.

## Data

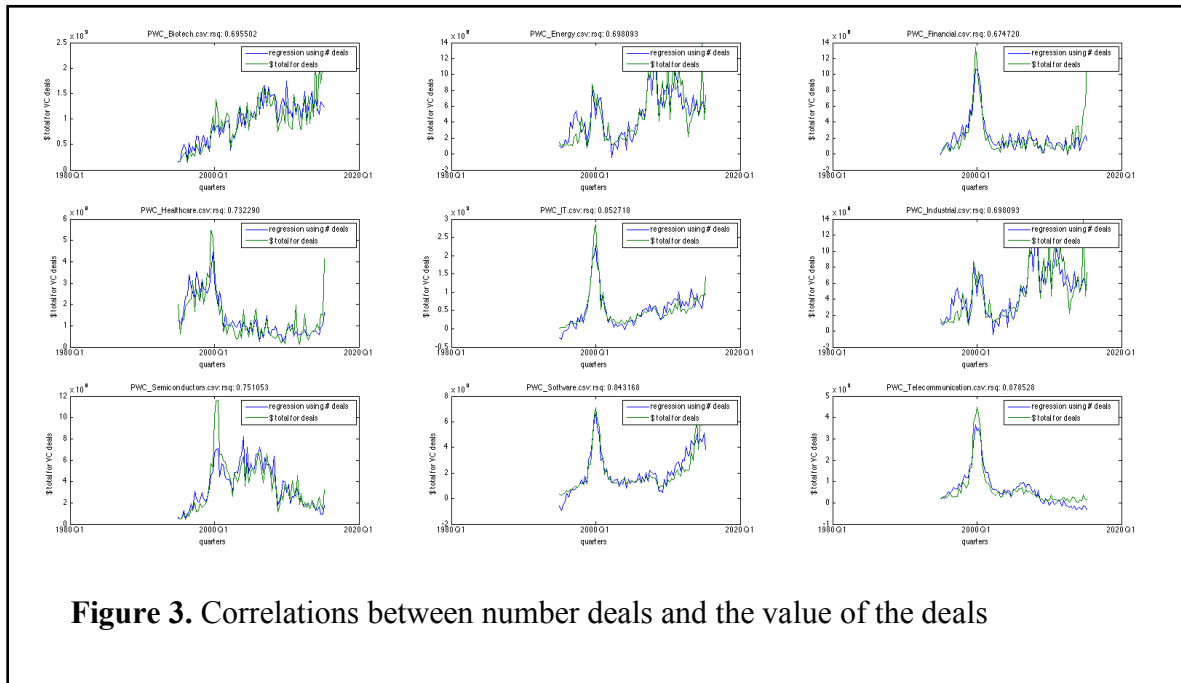
In order to perform our study, we obtained data regarding deals and investments in startups from PriceWaterhouseCoopers. Figure 2 shows an example of the data we were able to obtain. The line represents the total number of deals and the bars represent the total amount invested. Each bar corresponds to one quarter. Figure 2 shows the total for across all sectors. It



should be noted that the data is slightly biased towards software companies. This is because the majority of PwC deals are in the software industry. This explains the drop in financial mortgage crisis of 2008 as well as the drop observed during the dotcom bubble. During the financial mortgage crisis of 2008, the software industry was not hit as hard as other industries. That is why, even though the crisis of 2008 was arguably worse than the dotcom bust, venture capital investments did not drop that much during that period of time. PwC also can break down the data into various industry sectors such as software, biotechnology, energy, and industrial. We were pulled PwC data for biotechnology, energy, financial, healthcare, industrial, IT, semiconductors, software, and telecommunication. Corresponding indices NASDAQ indices

were available on Yahoo finance. Specifically, we obtained historical data for the NASDAQ bank, biotechnology, computer, industrial, and insurance, other-finance, and telecommunication indices. Since the some of the NASDAQ historical data only dated back to quarter one 1997, we only used data corresponding to dates after that quarter. All data was converted to quarterly data and transformed into units of percent change from last quarter before analysis.

The PWC data contains both number of deals and total value of the deals in a given quarter. In our study, we only used the number of deals from the PWC. Figure 3 shows the correlation between the PWC number of deals and the corresponding amounts invested.



For all of the sectors, there is a high R-squared value of approximately 0.7 or greater. This suggests that there is a high correlation between the number of deals and the total value of all deals, allowing us to simply use the number of deals in our regression.

The NASDAQ data is initially formatted as monthly data. There are several features associated with the raw data: opening, high, low, and closing indices. In order to format the data

into quarterly data, the opening index is mapped to the opening index of a three-month period. Similarly, the closing index is set to the closing index of a three-month period. The high index is set to the maximum of all the high indices in the three-month period, and the low index is set to the minimum of all the low indices in the three-month period. Another feature extracted was index swing, which is calculated by subtracting the opening index from the closing index in a quarter. Momentum is also derived from the raw data and used to measure how many consecutive quarters has the index been overall increasing. This is calculated by checking if the swing is positive or negative. If the swing is positive, the momentum for the current quarter is set equal to the momentum of the previous quarter plus one. If it is negative, the momentum for the current quarter is set to be equal to the momentum of the previous quarter minus one. Before any analysis was performed, all the features other than momentum were then converted into percent change from the previous quarter. In total, there are six features extracted from the raw NASDAQ data: swing, open, high, low, close, and momentum.



## **Analysis and Results**

Initial analysis was done on two time periods, the dotcom era (Q1 1998 – Q1 2004) and the subprime mortgage crisis era (Q1 2006 – Q1 2013). We suspected that during the dotcom era, software venture capital investments were driven in part by the movements of related industry stock market indices. In order to test our hypothesis, we chose three NASDAQ indices for the dotcom era that were involved in software: industrial, computer, and telecommunications. Since each NASDAQ index corresponds to 6 features, there were potentially 18 independent variables for regression analysis. We ran a granger causality test, which told us whether or not each of the variables were useful in predicting the future value of the percent change in number of PWC venture capital deals for software companies. 17 of the 18 independent variables were valuable for regression. The one feature that was not important was the momentum for NASDAQ computer indices. After filtering out the features that were not valuable in the regression and lagging every independent variable by one quarter, we performed regression. Figure 4 shows the results of the regression. An R-squared value of 0.8731 was observed, suggesting that there is a high correlation between the change in previous quarter stock market data for the chosen indices and change in number of PWC venture capital deals for software companies during the dotcom era.

We suspected industrial company venture capital investments were heavily influenced by related industry stock market indices during the subprime mortgage crisis. We ran a similar regression similar to the one run on the dotcom era PWC software data. The NASDAQ industrial, computer, bank, and telecommunications indices were chosen as independent variables. After running the 24 independent variables through a granger causality test, 19 were valuable in the analysis. The remaining independent variables were lagged by one quarter.

Regression was then performed and an R-squared value of 0.8623 was observed. Figure 5 shows the results. As we expected, there is a high correlation between the change in previous quarter stock market data for the chosen indices and change in number of PWC venture capital deals for industrial companies during the subprime mortgage crisis era.

After performing the initial analysis, we performed a rolling window regression analysis using all of our data in an attempt to see which important stock market indicators could be valuable in predicting venture capital investments. Rolling window regression analysis involves taking a window of data and running regressions within that window of data. By observing the regression coefficients and how they change as the window rolls forward in time, one can identify valuable independent variables that would be useful in predicting the dependent variable. The independent variables that have stable regression coefficients would be the ones that would be useful in forecasting future dependent variables.

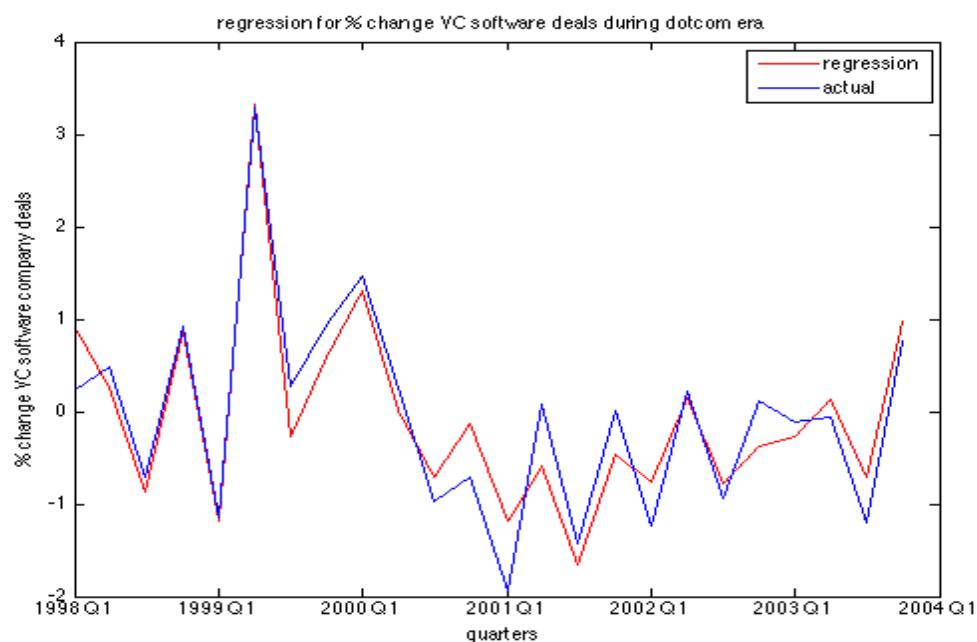
Since there are 7 NASDAQ indices extracted and each NASDAQ dataset has 6 features, there are in total 42 different features that can be used as independent variables in regression analysis. There are a total of 9 different PWC industries. Before any analysis is done, independent variables are filtered down by the granger causality test and lagged by one quarter. The analysis starts in Q1 1998 and windows corresponding to 3, 4, or 5 years were used as the timeframe for a regression.

In order for a rolling window analysis to be considered useful, an R-Squared threshold and threshold for number of regressions surpassing the R-Squared threshold is used to filter out bad results. Results are displayed for a 4-year rolling window analysis on PWC software deals. Figure 6 shows how the R-squared value changes over time as the window rolls forward. Figure 7 shows the variables and the associated normalized regression coefficients for each variable as

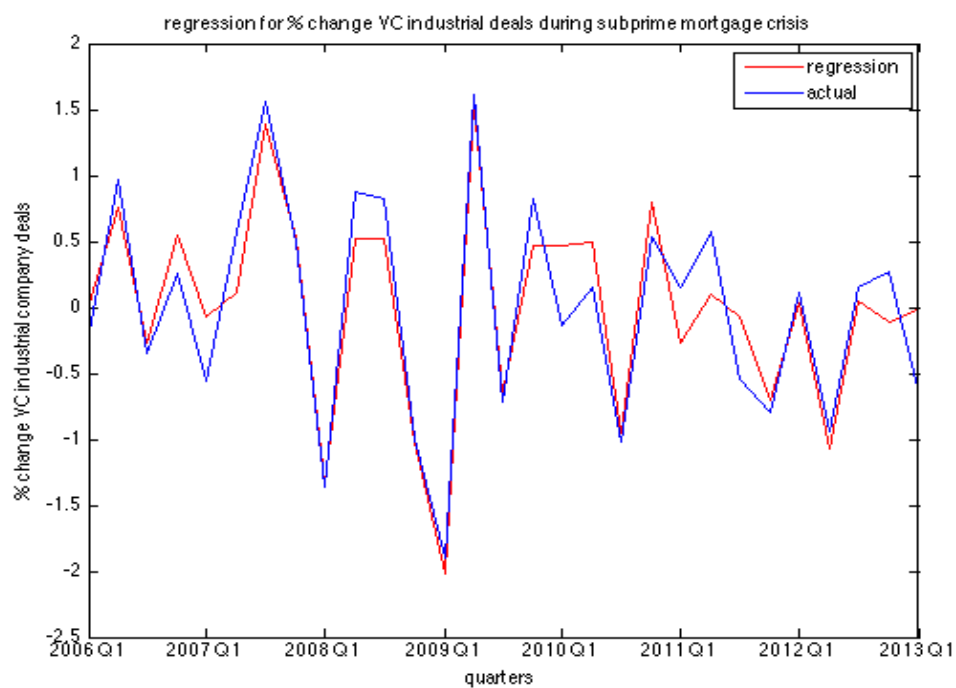
the window rolls forward in time. Figure 8 shows the filtered down variables that are valuable in forecasting future values of PWC software deals because of their low variance regression coefficients. Figure 9 shows one filtered down variable with low variance regression coefficients plotted against its mean. The stable weights for the 5-year regression software window analysis share common significant features with the 4-year regression software window analysis. Four of the five features (bank momentum, biotech open, insurance momentum, other-finance close index) in the 5-year analysis appear in the 4-year analysis. The results obtained suggest that there are several stock market indicators that are useful in predicting venture capital behavior.

Figure 11 shows the relationship between the number of deals and the value of the NASDAQ index. The plot shows that NASDAQ leads the number of deals, which confirms our hypothesis that the venture capital private investments are influenced by the performance of NASDAQ.

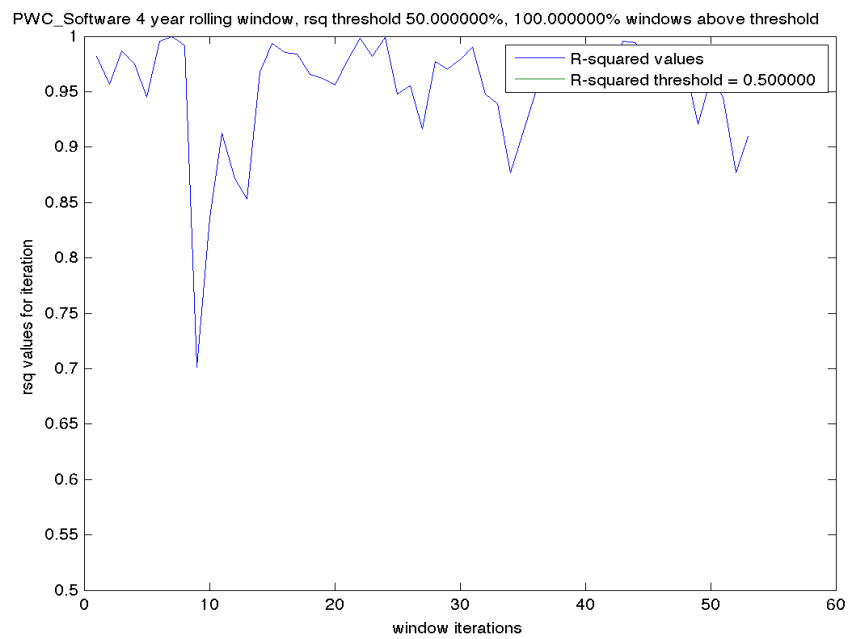
Overall the rolling window analysis yielded good results for 3-year and 4-year windows for the biotechnology, energy, and industrial industries. For software, there were too many independent variables to fit a regression model for 3-years. This is because it was trying to regress on 12 data points with more than 20 independent variables. The 4-year and 5-year rolling window analysis for software yielded good results. The other PWC industries (financial, healthcare, IT, semiconductors, telecommunication) did not yield good rolling window results. This was due to the fact that the selected NASDAQ sectors were not very close to the PWC industries that yielded bad results. For example, the companies in NASDAQ computer are probably very similar to the companies under the PWC software industry category.



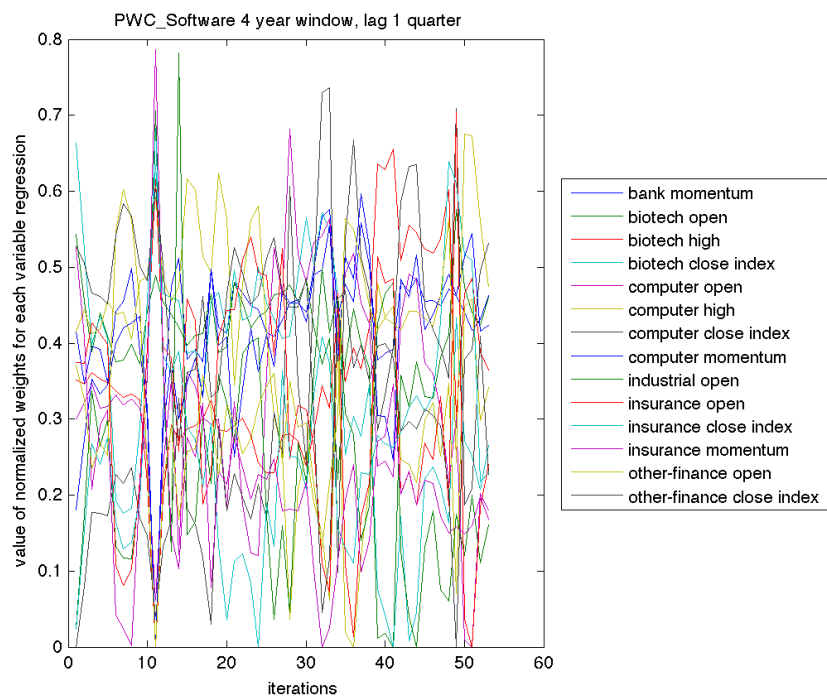
**Figure 4.** Regression during the dotcom era



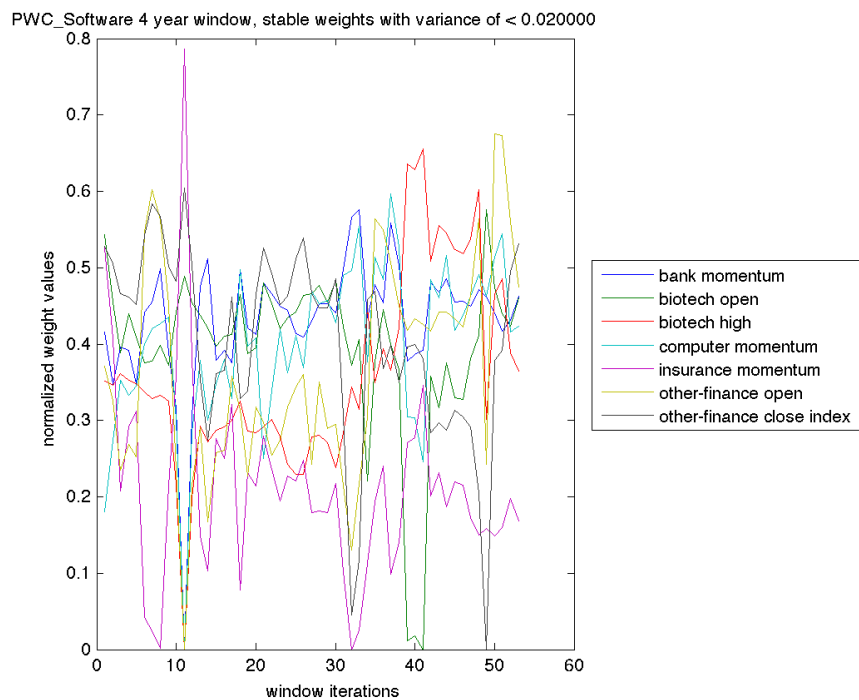
**Figure 5.** Regression during the subprime mortgage crisis



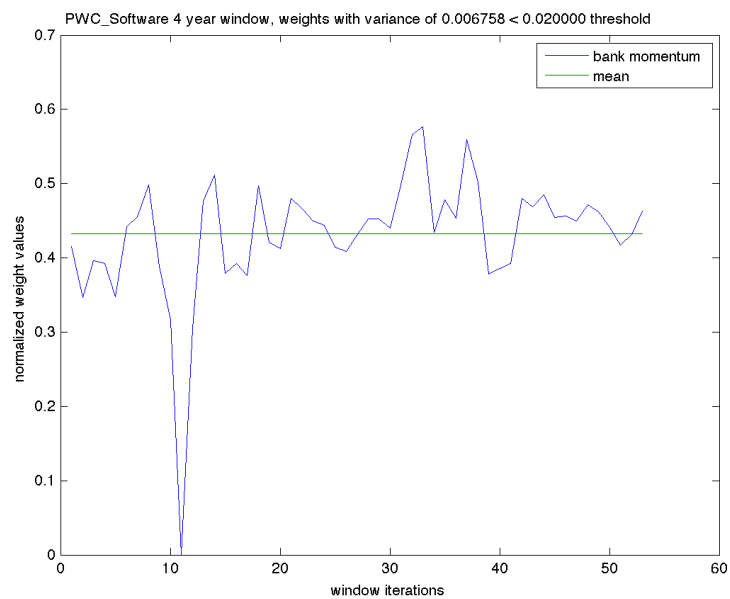
**Figure 6.** 4 year rolling window



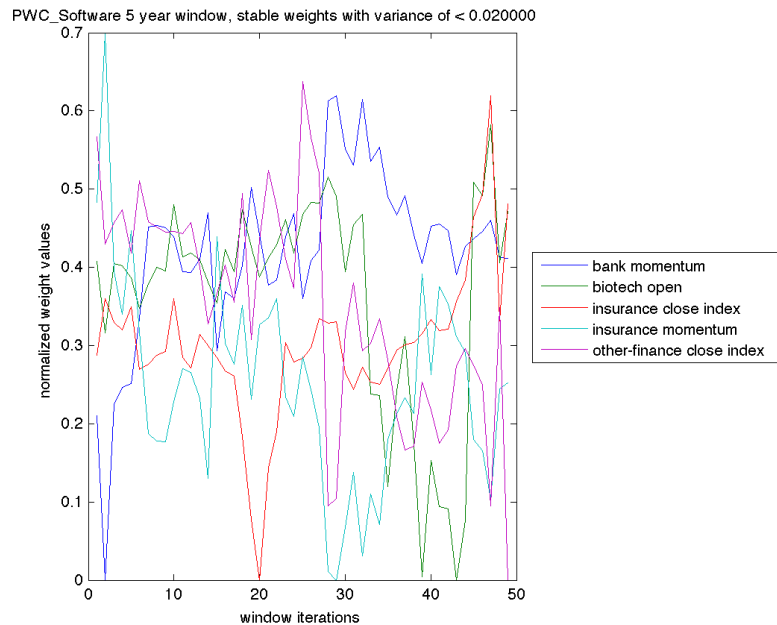
**Figure 7.** Software, 4 year rolling window



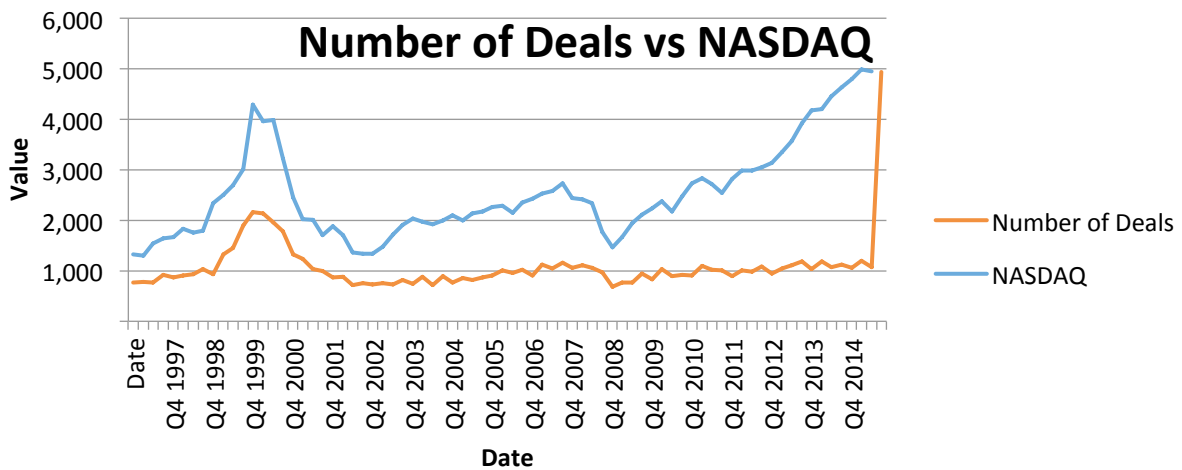
**Figure 8.** Software, 4 year stable weights



**Figure 9.** Lowest Variance Weights



**Figure 10.** Software, 5 year stable weights



**Figure 11.** Number of venture Deals plotted along with the NASDAQ Composite

## **Conclusions**

### **Andrew's Conclusion**

Our results show that the performance of publically traded companies is strongly correlated with the willingness of investors to invest in startup companies. Using the stock market indices as a metric for investor enthusiasm, we can see that as investors are more willing to trade and hold stocks in a certain sector, they are also willing to take on riskier investments such as startups. We also found that performance of related sectors are also strong indicators of investors' willingness to invest in startups.

The relationship between the public market and the private equity space is apparent during atypical periods as well. During the dotcom bust and the financial crisis, we see that the market volatility and investor uncertainty was also reflected in private investments. This shows that investors who are moving out of a turbulent public market are also not willing to reinvest in startup companies. This may be explained by the uncertainty of a return from initial public offerings or buyouts.

### **Harrison's Conclusion**

We used the PWC dataset and the NASDAQ dataset because we hypothesized that there was a relationship between public markets and private investments. Our results indicate that public market investment enthusiasm translates to enthusiasm in venture capital investments. We found that performance of closely related sectors acts as a better predictor of the venture capital investments than the performance of less closely related sectors. This relationship holds true even in atypical periods of market activity, such as during the dotcom bubble and the financial mortgage crisis. If we were to redo the project, we would want to get more data. This is because the current data is only on quarterly data. It would be great if there were a source for monthly



data. This would triple the number of data points that we would be able to regress against.

Another thing that would help combine the NASDAQ data with other external data sources such as the S&P. This is because the NASDAQ data and S&P cover different sets of companies. All the code can be found at <https://github.com/harrisonzhao/eid-300-financial-regression>. The graphs for all the analyses performed can also be found there.

## References

[1] Rachleff, Andy. "Why Angel Investors Don't Make Money ... And Advice For People Who Are Going To Become Angels Anyway." TechCrunch.

[2] "Startup Financing Cycle" by Kmuehmel, VC20 -  
[https://commons.wikimedia.org/wiki/File:Startup\\_financing\\_cycle.svg](https://commons.wikimedia.org/wiki/File:Startup_financing_cycle.svg). Licensed under CC BY-SA 3.0 via Commons -  
[https://commons.wikimedia.org/wiki/File:Startup\\_Financing\\_Cycle.png#/media/File:Startup\\_Financing\\_Cycle.png](https://commons.wikimedia.org/wiki/File:Startup_Financing_Cycle.png#/media/File:Startup_Financing_Cycle.png)

[3] [https://indexes.nasdaqomx.com/docs/methodology\\_COMP.pdf](https://indexes.nasdaqomx.com/docs/methodology_COMP.pdf)