Strike Independent Implied Volatility Analysis in the Q-Alpha Model

Wen Qi Chin, Ankur Goel, Jianwei Low, Seihout Sok

June 10, 2009

Abstract

We examined the short term (30 days to maturity) and long term (180 days to maturity) strike-independent volatility changes derived from the q-alpha model and applied Principal Components Analysis (PCA) to identify the main factors that affect the implied as well as the historical volatilities. We then performed the same analysis on the stock space and compared the factors found in the implied and historical volatility space with the stock space. To find common factors that affected movement of the volatility surface, and also to tie economic meaning into these factors, we isolated stocks from the financial industry and picked out several financial indices (e.g: S&P500, Nikkei). We used these data and again performed PCA and correlation analysis among historical volatility, implied volatility, financial stocks and financial indices.

1 INTRODUCTION

The Black-Scholes model takes in 5 parameters (option price, strike price, stock price, time to maturity and interest rate) and assumes a constant volatility for the underlying asset when predicting option prices. However, implied volatilities are observed to vary in strike price and time to maturity, making the Black-Scholes model inaccurate in its predictions. As such, several models have been developed in an effort to better model the volatility changes. The q-alpha-sigma is one such model. It replaces the Brownian motion with a different stochastic process, and by doing so, does a respectable job of modeling the volatility smile. Hence, the two-dimensional volatility surface is effectively reduced to a one-dimensional term structure.

Much work has been done analyzing the surface changes of the Black-Scholes model, and how these changes might relate to economic factors. In Mixon's study (2002), it explored the relationship of changes in the S&P500 index implied volatility as derived from the Black-Scholes model to economic state variables and found that the majority of the explanatory power came from index returns.[i] Another example of past work done with the Black-Scholes model is in Cont's and Fonseca's 2002 study, where they used a time series of option prices on the SP500 and FTSE indices to study the deformation of the surface and found that it may be represented as a randomly fluctuating surface driven by a small number of orthogonal random factors.[ii] Although this paper will specifically look at the implied volatility surface generated by the q-alpha-sigma model, both of these studies done on the Black-Scholes model provide a good starting point on how we approached our analysis. We used Mixon's results to organize our data into short-term and long-term volatilities, and modeled part of our analysis process after Cont's and Fonseca's study. The little work that has been done on the Q-Alpha-Sigma model only focused on studying factors that affected the implied volatility surface, and how best to model this surface (see

Dixon et al and Abmruster et al).[iii] However, no work has been done yet to put economic meaning behind these factors that affect the q-alpha-sigma implied volatility surface.

To explore this issue, we first performed PCA on the implied volatilities, found clusters, and then carried out correlation analysis on the economic factors with financial stocks in an effort to put economic meaning to the factors.

2 DATA

A data set consisting of consecutive data points from January 2003 to December 2005 was given to us. This data had a constant alpha value of 1.5. We parsed the data and picked out 81 unique stock IDs. For each of these IDs, we kept a matrix that contained the different start dates, end dates, time to maturity and implied volatility. For each start date, each stock had about 4-6 different options with varying time to maturity. The time to maturity ranged from 5 to 751 days.

To investigate how does implied volatility varies in time space in relation to the stock space, we decided to build two matrices – one that represented short term (30 days to maturity) volatilities, and another that represented long term (180 days to maturity) volatilities for a consecutive three year time span (2003 to 2005).

The motivation for organizing the data in this fashion came from Mixon's study. In his study on *Factors Explaining Movements in the Implied Volatility Surface* (2002), three principal components were found that explained 97.8% of the total variance in the S&P500 volatility index.[iv] He found that each of these components explained different portions of the surface depending on their time to maturity – the first principal component accounted for most of the variance in options with maturities of up to 12months, while the second principal component explained 30-50% of the variance in the 3-5year maturity variation, while the third principal component had most of its explanatory power at the 1 month horizon. Based on this study, it seems that options with the same time to maturity are affected by the same common factors. Accordingly, we decided to organize our data into these two short-term and long-term matrices so as to isolate the common factors for each of these different times to maturities.

For both of these matrices, the columns are organized by stock IDs, while each row contained the implied volatility represented different start dates. We performed a couple of calculations to arrive at the final data matrix. Firstly, as not every ID had a 30 or 180 days time to maturity for each particular date in the three year time span, we used linear interpolation to calculate the missing data. Secondly, we also discarded matrix rows where there was absolutely no data across all the IDs (e.g: weekends, public holidays). Thirdly, we calculated the log difference (log $\frac{today\ implied\ vol}{yesterday\ implied\ vol}$) for each data point. And lastly, we used a moving average to smoothen out the data.

We also gathered corresponding stock prices and historical volatilities (for both 30 days and 182 days to maturity) for each respective date in the three-year time frame from OptionMetrics, Wharton Business School and performed the same calculations on them to arrive at the final stock price matrix, 30-day historical volatilities matrix and 182-day historical volatilities matrix. It is important to note that we used a 182-day historical volatilities matrix instead of a 180-day one because no data could be found for historical volatilities with 180 days to maturity. However, in our analysis, we viewed the 182-days historical volatilities the same as a 180-days one because our group thought that the difference would be negligible for the purpose of comparing short-term versus long-term volatilities.

3 ANALYSIS

<u>*PCA Analysis*</u>

Figure 1 shows the ten largest eigenvalue percentages for the PCA analysis of the short-term and long-term, implied and historical volatility matrix. The largest percentage for all of these volatility matrices ranges from 20% to 40%, while the rest of the ten eigenvalue percentages each constitute about 5 to 8%. Indeed, the first principle component explains about 20% to 40% of the fluctuation whereas the rest of the ten principle components only each explain about 5 to 8%. These ten principle components explain about 80% of the fluctuation of the volatility.

Figure 2 shows the ten largest eigenvalue percentages for the PCA analysis of the stock return matrix. For the stock PCA analysis, the largest eigenvalue is 62%, which is much more than that of the volatility matrices. The combination of ten principle components explains more than 90% of the fluctuation.

PCA results show that while one main factor can be used to explain the fluctuation in stock return, the fluctuation of volatility of an option is driven by many small factors. Although the main factor is about 20 to 40%, this factor might be hard to pinpoint since it only constitutes a very small percentage of the total fluctuation. Furthermore, considering that PCA analysis is done on all 76 stock options representing different companies from many diverse industries, it is difficult to find a common principle component or factor affecting these stock options. Therefore, in the following section, we will perform cluster analysis to narrow down our data.

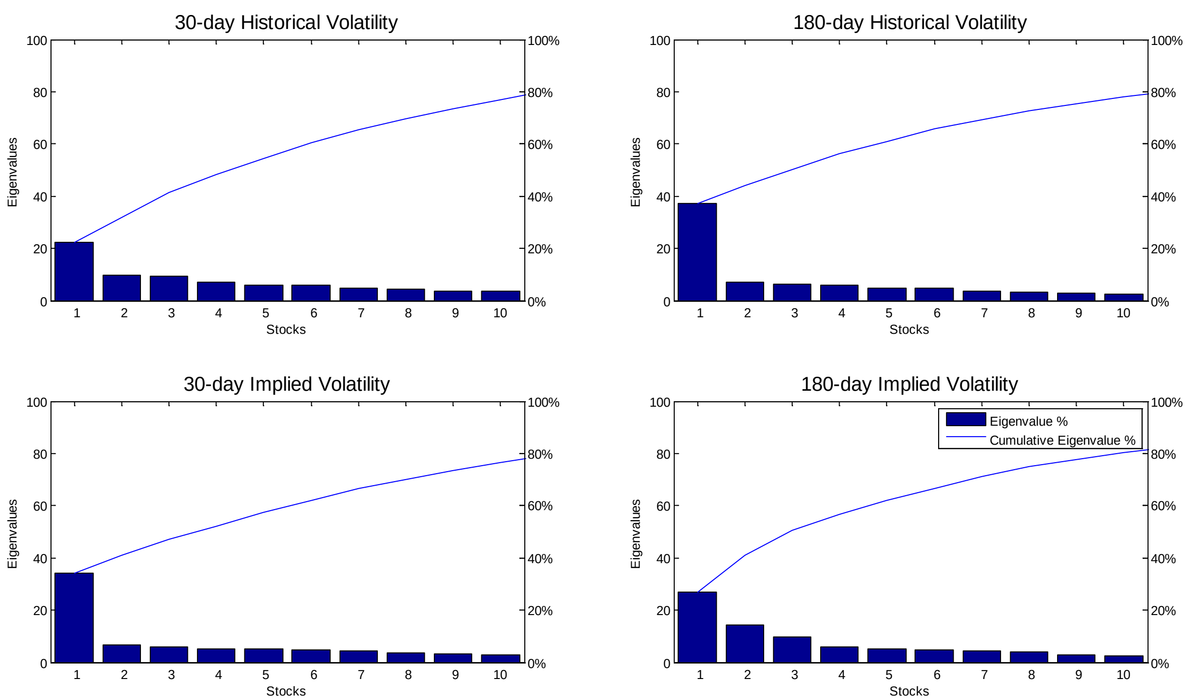


Figure 1: PCA Results from Short-term and Long-term, Historical and Implied Volatility.

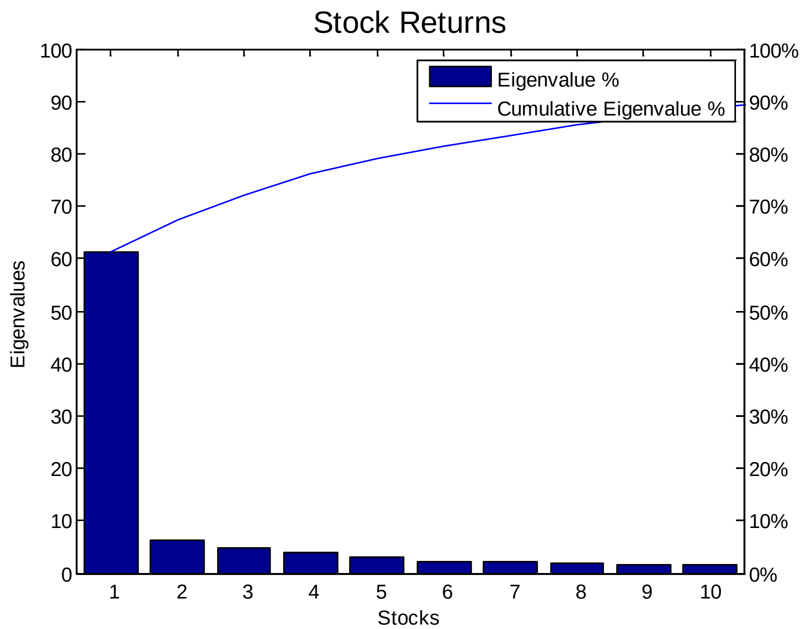


Figure 2: PCA Result from Stock Returns

Cluster Analysis

Cluster analysis was only performed on the long-term implied volatility and stock return data because its data quality is better than that of short-term implied volatility data. We found that short-term implied volatility data was not as normally distributed as the other data. As shown in Figure 3, the long-term implied volatility data of American Express Company follows the red line closely where a normally distributed data lies. We used this method and determined that long-term option and stock return data are suitable for the clustering.

Two types of cluster analysis were performed: hierarchical and k-means clustering[v]. The results of these analyses are shown in Figure 4. Long-term options are clustered well with a total of 11 clusters for both hierarchical and k-means clustering. Two to three clusters have a very dense group of about 15 options. The stock returns data does not have as good clustering result as that of the long-term option and it has different result for different clustering analysis. Hierarchical clustering results in 17 stock groupings whereas the k-means analysis only results in 12 groupings.

However, these clustering methods do not give us any consistent and good cluster we need to narrow down our analysis. In addition, we want to also cluster our data based some economic meaning. Therefore, we will cluster these options based on industry instead. In the following section, we will perform correlation analysis first before repeating the PCA analysis on the financial industry.

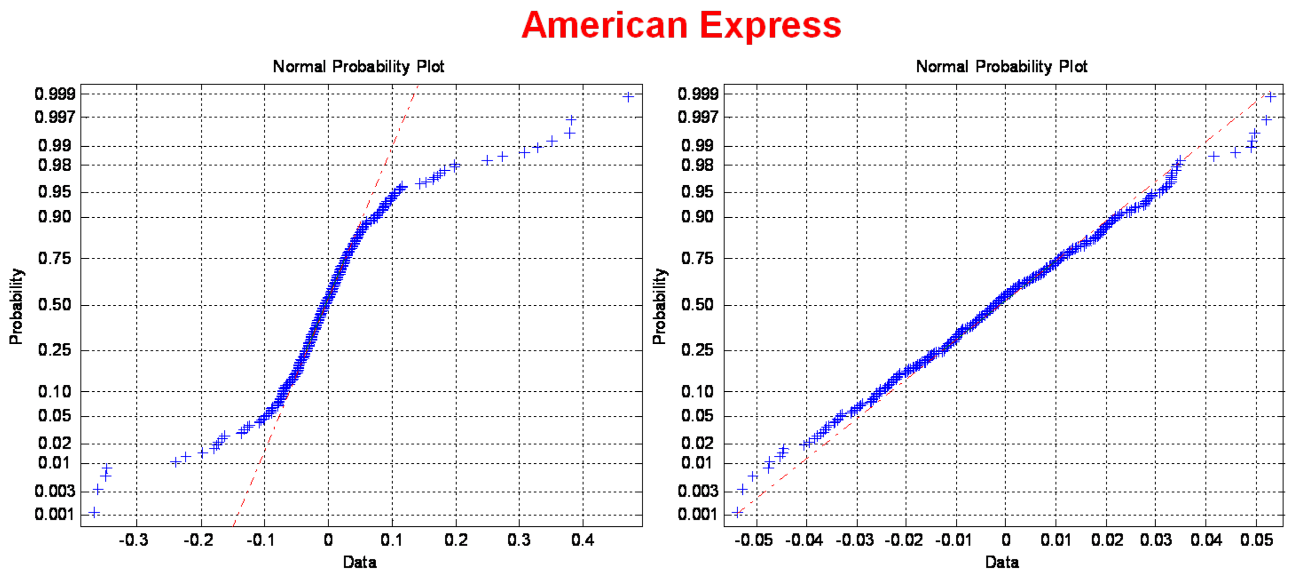


Figure 3: Data Quality a) Implied Volatility-Short Term b) Implied Volatility-Long Term

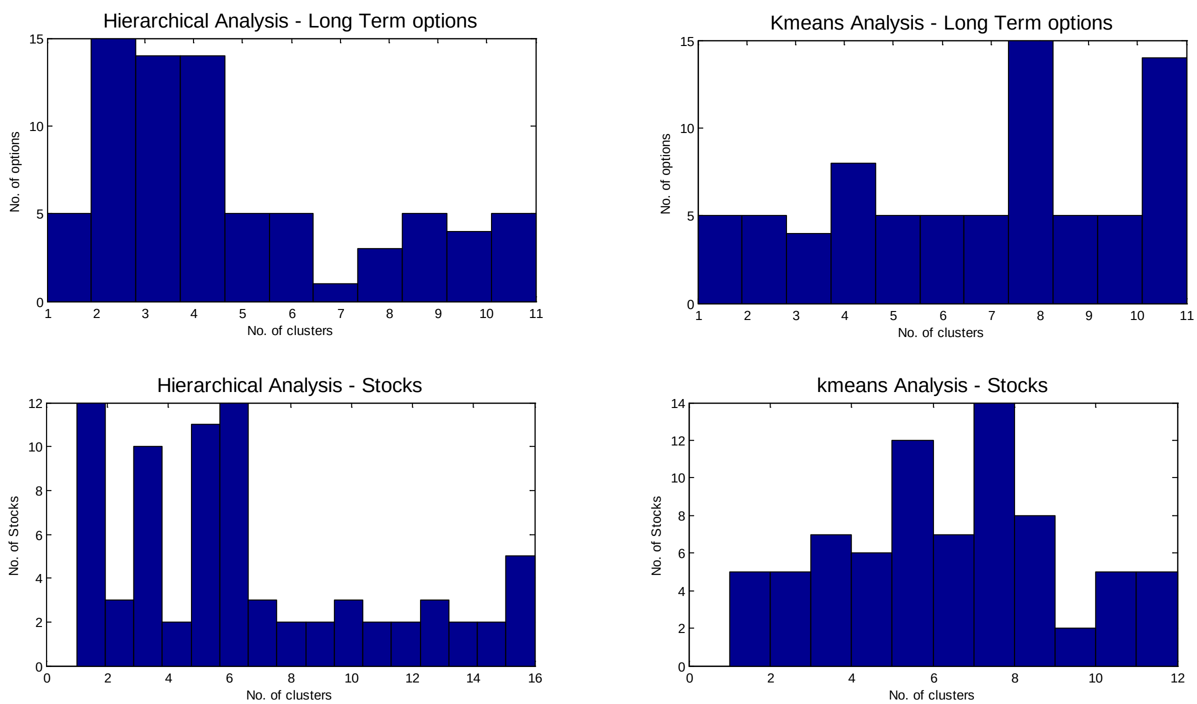


Figure 4: Cluster Analysis

Economic Factors

We focused on the financial industry for our analysis. The following stocks were grouped under the financial industry:

Ticker	Industry	Company	
101966	410	BAC	Bank of America
102936	410	JPM	JP Morgan
111953	410	WFC	Wells Fargo
104866	414	USB	U.S. Bancorp
106893	420	LEH	Lehman
107455	420	MER	Merrill Lynch
105329	423	GS	Goldman Sachs
107704	423	MS	Morgan Stanley
101375	424	AXP	American Express Company
101273	432	ALL	The Allstate Corporation
101397	432	AIG	AIG
105581	432	HIG	Hartford Financial Services

Table 1: Details of Financial Stocks analyzed from data

The economic factors that we chose were

- S&P Index
- Nikkei Index
- US Libor 3 months
- US Libor 1 year
- Treasury bonds: 4 weeks, 3 months, 6 months
- Constant Maturity Treasury (CMT): 1 month, 6 months
- AAA Corporate Bonds

Both implied volatility of the options and economic factors were expressed in the $X_t = \log\left(\frac{r_t}{r_{t-1}}\right)$ where r_t is the return on time t. We then performed a correlation analysis between 180 days implied volatility and the economic factors as shown in the tables below. We chose 180 days implied volatility because it has a more normal distribution and has a better PCA analysis.

PCA on Financial Volatility

PCA analysis was done on the 180 days implied volatility (Fig 5) and based on the results we see that 3 factors account for ~95% of volatility variation.

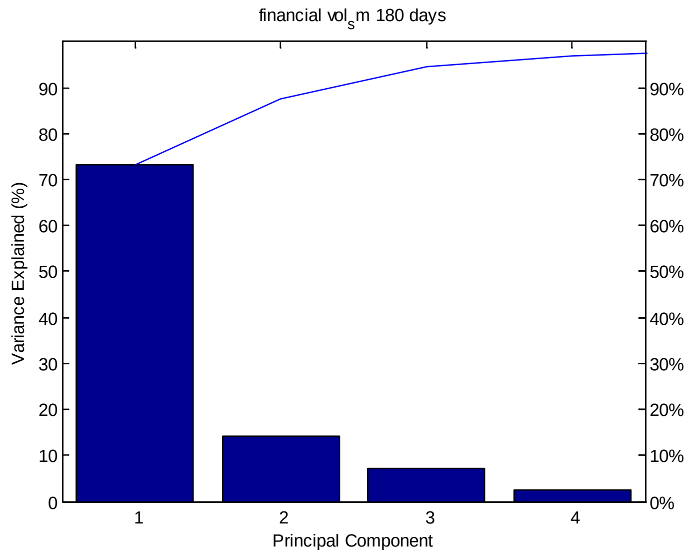


Figure 5: PCA on Implied Volatilities (Financial Stocks)

We plotted the implied volatility on the coordinates of the principal components (Fig 6) and concluded that there are few outliers and most of the data are clustered at zero mean. This verifies that the data is suitable for further analysis and that the principal components can rightly account for the variations.

Analysis

Based on the correlations shown in Tables 2, 3 and 4, we concluded that the possible economic factors that affect the movements in the 180 days implied volatility are AAA corporate bonds, S&P Returns, Nikkei Index and the stock returns or respective companies. Possible reasons for the strong correlations are:

- S&P 500: Represents the general market sentiment in the primary market Nikkei
- index: Represents the international market sentiment
- Stock return: Caused by the leverage effect
- AAA bonds: Represents the general market sentiment in the secondary market

As we would then try to trace the movement of the implied volatility, we chose assets from multiple asset classes that span the largest independent space. We identified this by performing singular value decomposition on the combination of assets and determined 3 assets that span the optimum space – AAA bonds (fixed income), S&P 500 (equities) and stock return (equities).

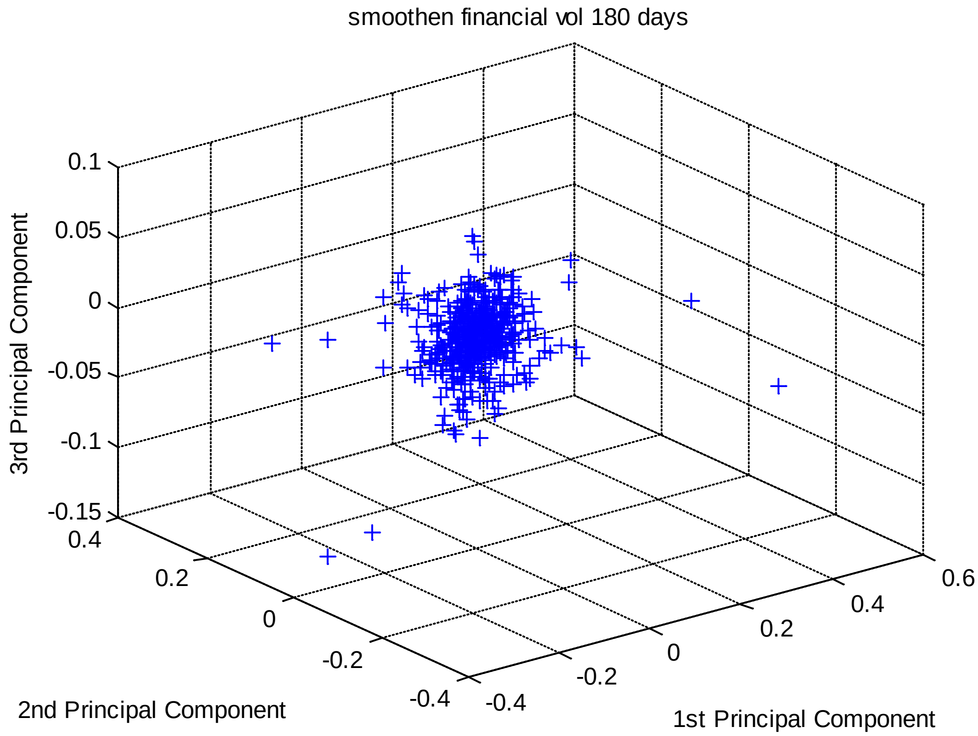


Figure 6: Implied Volatility on Principal Component Space (Financial Stocks)

		Economic factors									
		T-Bill 4wk	T-Bill 3m	T-Bill 6m	CMT 1m	CMT 6m	AAA bond	S&P Return	Nikkei Index	Libor 3m	Libor 1yr
Stock name	Bank of America	-0.00885	-0.01059	-0.01961	-0.03090	-0.12099	-0.22223	-0.28780	-0.17806	-0.09369	0.00584
	JP Morgan	-0.01860	-0.01801	-0.02705	-0.03595	-0.10855	-0.18409	-0.24186	-0.13264	-0.07253	-0.01212
	Wells Fargo	-0.05098	-0.05518	-0.05838	-0.05621	-0.13394	-0.23142	-0.28918	-0.17677	-0.09942	-0.02052
	U.S. Bancorp	-0.07661	-0.08429	-0.08182	-0.07043	-0.14381	-0.24771	-0.29120	-0.17540	-0.09513	-0.01038
	ehman Brother	-0.07982	-0.08468	-0.08413	-0.07435	-0.14728	-0.25038	-0.30279	-0.19515	-0.11125	-0.02428
	Merrill Lynch	-0.08421	-0.09081	-0.08886	-0.07612	-0.14843	-0.25194	-0.29794	-0.19423	-0.11028	-0.02010
	Goldman Sachs	-0.11266	-0.12586	-0.11673	-0.09128	-0.17278	-0.30586	-0.35229	-0.24965	-0.13965	-0.00905
	Morgan Stanley	-0.09395	-0.10371	-0.10012	-0.08455	-0.17839	-0.31379	-0.37206	-0.26004	-0.14513	-0.00504
	American Express Co.	-0.06834	-0.07248	-0.07625	-0.07618	-0.17743	-0.31263	-0.41252	-0.29585	-0.17336	-0.02948
	The Allstate Corporation	-0.08133	-0.08833	-0.08859	-0.08700	-0.18893	-0.32648	-0.40687	-0.27272	-0.15174	-0.01178
	AIG	-0.07589	-0.08369	-0.08239	-0.07957	-0.17288	-0.30338	-0.39162	-0.26643	-0.15329	-0.02348
	Hartford Financial Services	-0.06230	-0.06701	-0.06853	-0.07433	-0.16389	-0.28623	-0.37390	-0.26075	-0.15181	-0.02896

Table 2: Correlation between 180 days implied volatility (Financial Stocks) and economic factors

Stock name		Historical volatility											
		Bank of America	JP Morgan	Wells Fargo	U.S. Bancorp	Lehman Brothers	Merrill Lynch	Goldman Sachs	Morgan Stanley	American Express Co.	The Allstate Corporation	AIG	Hartford Financial Services
	Bank of America	-0.00586	0.00904	-0.00104	0.00690	0.00913	0.01312	-0.00993	-0.02104	-0.01751	-0.02066	-0.20294	-0.41183
	JP Morgan	-0.02263	-0.09092	-0.13124	-0.11837	-0.11422	-0.10822	-0.06249	-0.05201	-0.05621	-0.05455	-0.17975	-0.32284
	Wells Fargo	-0.02907	-0.07408	-0.10551	-0.09634	-0.09191	-0.08304	-0.05005	-0.03788	-0.04029	-0.03859	-0.17527	-0.35155
	U.S. Bancorp	-0.01555	-0.05925	-0.09572	-0.08426	-0.08586	-0.08109	-0.05191	-0.03206	-0.03599	-0.02754	-0.13643	-0.31233
	ehman Brothe	-0.02707	-0.08557	-0.11561	-0.10589	-0.10781	-0.10556	-0.06670	-0.04632	-0.04985	-0.04396	-0.14408	-0.30143
	Merrill Lynch	-0.01862	-0.07355	-0.10395	-0.09509	-0.10027	-0.10008	-0.06985	-0.04765	-0.05303	-0.05241	-0.15854	-0.31222
	Goldman Sachs	-0.00048	0.02021	0.01759	0.02244	0.00891	0.00157	-0.01973	-0.01042	-0.00749	-0.00772	-0.13260	-0.33916
	Morgan Stanley	0.01188	0.01545	0.01735	0.02520	0.00880	-0.00287	-0.01896	-0.01428	-0.00739	-0.01001	-0.15009	-0.36970
	American Express Co.	-0.00700	0.02262	0.03975	0.04428	0.03264	0.02681	0.00251	-0.00032	0.00820	0.00314	-0.14429	-0.38964
	The Allstate Corporation	0.01002	0.04056	0.04619	0.04517	0.03440	0.03041	0.00415	0.00831	0.02460	0.01513	-0.14289	-0.38996
	AIG	-0.00462	0.03668	0.04961	0.04567	0.03526	0.03295	0.00963	0.00707	0.03007	0.02787	-0.10320	-0.34506
	Hartford Fin. Services	-0.00925	0.02763	0.04597	0.03649	0.02964	0.03229	-0.00184	-0.00243	0.02009	0.01300	-0.11747	-0.34301

Table 3: Correlation between 180 days implied volatility (Financial Stocks) and historical volatility

Stock name		Stock return											
		Bank of America	JP Morgan	Wells Fargo	U.S. Bancorp	Lehman Brothers	Merrill Lynch	Goldman Sachs	Morgan Stanley	American Express Co.	The Allstate Corporation	AIG	Hartford Financial Services
	Bank of America	-0.53076	-0.45827	-0.49013	-0.45739	-0.44271	-0.43008	-0.48511	-0.4777	-0.48159	-0.49992	-0.47527	-0.47031
	JP Morgan	-0.41384	-0.38327	-0.40224	-0.38057	-0.37024	-0.35389	-0.36321	-0.35463	-0.34801	-0.36267	-0.34612	-0.34649
	Wells Fargo	-0.45412	-0.39585	-0.43751	-0.42204	-0.41059	-0.40168	-0.44873	-0.42989	-0.4218	-0.43835	-0.41948	-0.41549
	U.S. Bancorp	-0.41927	-0.37033	-0.42037	-0.42934	-0.41789	-0.40816	-0.45772	-0.43766	-0.41883	-0.43368	-0.42145	-0.42167
	Lehman Brothe	-0.41104	-0.36016	-0.41363	-0.42402	-0.42976	-0.41837	-0.48016	-0.46225	-0.44773	-0.45595	-0.44993	-0.44731
	Merrill Lynch	-0.41412	-0.36971	-0.41911	-0.42732	-0.4327	-0.43268	-0.48656	-0.46874	-0.45191	-0.45896	-0.44257	-0.43977
	Goldman Sachs	-0.46266	-0.40171	-0.47314	-0.49081	-0.50313	-0.51437	-0.62631	-0.60724	-0.59577	-0.59836	-0.58368	-0.56274
	Morgan Stanley	-0.51256	-0.47706	-0.53365	-0.5506	-0.5699	-0.56906	-0.66067	-0.66304	-0.65679	-0.65383	-0.64773	-0.63303
	American Express Co.	-0.53757	-0.52191	-0.57708	-0.58253	-0.60344	-0.60447	-0.68748	-0.69362	-0.70702	-0.70275	-0.69714	-0.67835
	The Allstate Corporation	-0.54533	-0.53342	-0.57702	-0.58486	-0.59003	-0.58831	-0.64922	-0.65307	-0.65985	-0.67494	-0.66315	-0.65805
	AIG	-0.48201	-0.48048	-0.52609	-0.54342	-0.55045	-0.53966	-0.60398	-0.61006	-0.62166	-0.63446	-0.64587	-0.63963
	Hartford Fin. Services	-0.47443	-0.46235	-0.50273	-0.51995	-0.52433	-0.5103	-0.55646	-0.56527	-0.57781	-0.59529	-0.60414	-0.6196

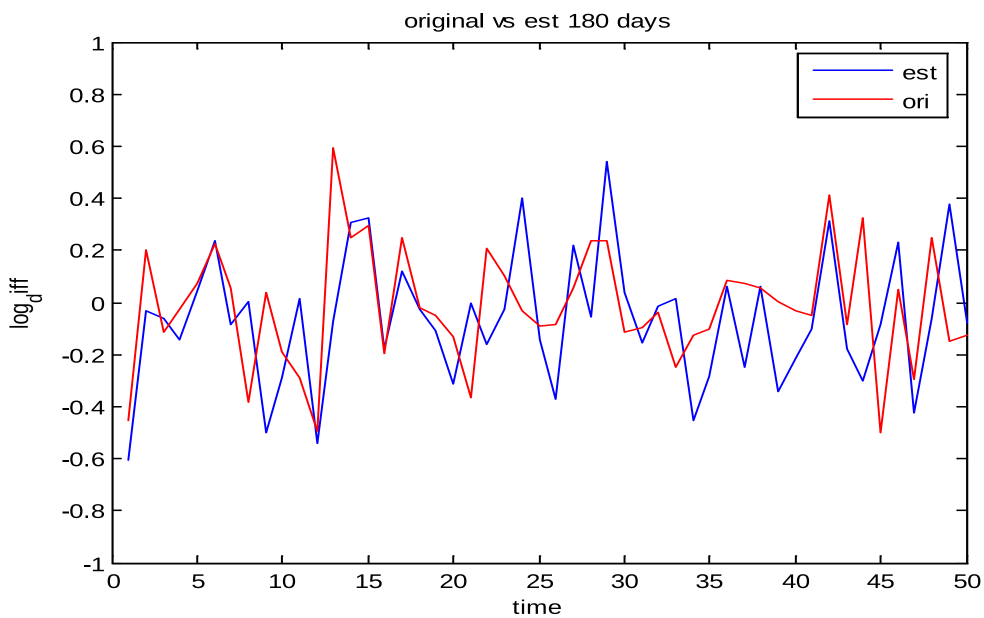
Table 4: Correlation between 180 days implied volatility (Financial Stocks) and stock return

Solving the Least-Square Problem

We will try to predict the 180 days implied volatility movements for American Express Co. from here. Using the 3 factors mentioned above, we solve for the least square problem of $\sigma_{ori} = LF$, where L is the loading matrix while F is the matrix of containing daily data of the AAA bonds, S&P 500 return and stock return. We then use Moore-Penrose pseudoinverse to minimize the tracing error such that estimated implied volatility matches the actual implied volatility, $\sigma_{est} \approx \sigma_{ori}$

Approach 1: Fix Window

We used all 436 historical data points to determine a loading factor that best approximates the impact on the 3 selected factors on the implied volatility movements



Correlation = 0.4534

Figure 7: Fixed Window Approach

Approach 2: Sliding window

We used a sliding window to identify a loading matrix. In our case, we used a 50 days sliding window length to solve for the loading coefficient:

$$L_t = \sigma_{ori(t-50:t-1)} F^{+}_{(t-50:t-1)}$$

then, we solve for the estimated implied volatility for each day iteratively:

$$\sigma_{est@t} = L_t F_t$$

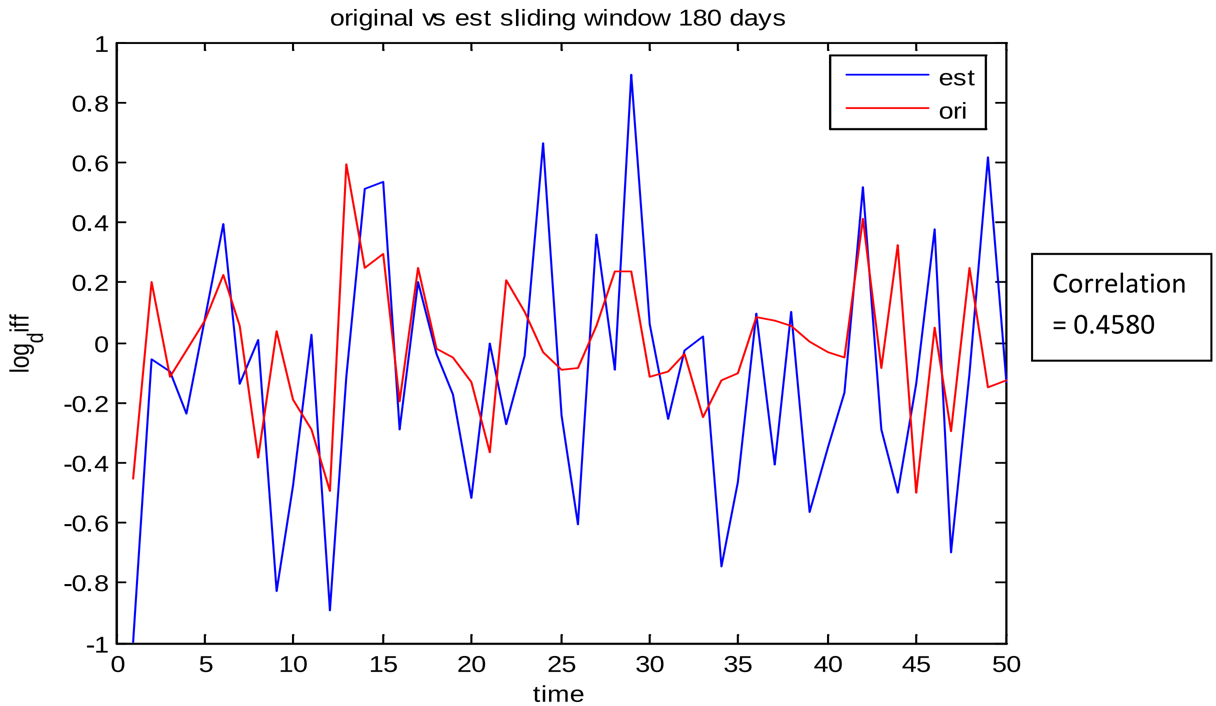


Figure 8: Sliding Window Approach

Analysis

Both the fix and sliding window (Fig 7 and Fig 8) produce the same shape or ability to predict the direction. However the fix window is able to better track the magnitude of the original data. Nonetheless both the sliding and fix window produces an estimated implied volatility that is perfectly correlated with each other. In terms of predicting a long/short decision regardless of the target value, our estimator is able to correctly predict 67.89% of the time.

Conclusions and Future Work

Based on the above analysis, we found that statistically significant negative correlation exists between the implied volatilities and the stock returns. No such correlation, on the other hand was found between the historical volatilities and stock returns. Note that we have not taken into account of any lag between historical volatilities and stocks and this might have an impact on the veracity of above observations.

Next we were able to find better clustering for long term implied volatility options as compared to the stock returns. The long term options data is again found to be much more normal distributed than the short term options. Data quality has an impact on the results given by the PCA and clustering.

Finally we isolated the stocks from financial industry and attempted to explain their behavior vis a vis few economic factors and treasury indices. We found statistically significant correlation between long term implied volatility and stock returns, S&P 500 Index, Nikkei and AAA bonds. Though by no means conclusive, this analysis should give the reader an understanding of the methodologies that can be employed in doing similar analysis for other indices and industries.

In future, there is a scope for doing the above analysis with other economic factors that might explain the movements in volatility index better. The analysis can also be extended to different industries like technology, consumer goods etc. One can also look at the residuals in more detail to understand the tracing of the volatility movements better.

References

[i] Mixon, Scott (2002). *Factors Explaning Movements in the Implied Volatility Surface.* The Journal of Future Markets. Vol 22, 915-937.

[ii] Cont, Rama and Jose da Fonseca (2002). *Dynamics of Implied Volatility Surfaces.* Quantitative Finance. Vol 2, 45-60.

[iii] Dixon, Matthew et al (2008). *Implied Volatility Term Structure and Vega Hedging in the Q-Alpha-Sigma Model.*
Abmruster, Benjamin et al (2008). *Implied Volatility Modeling.*

[iv] Mixon, Scott (2002). *Factors Explaning Movements in the Implied Volatility Surface.* The Journal of Future Markets. Vol 22, 915-937.

[v] Look Up Matlab StatisticsToolbox