19 MARCH 2015

DATAMINR SIGNALS ANALYSIS

Financial Network Analytics Ltd.

EXECUTIVE SUMMARY

In this study we analyze asset returns around the time of Dataminr signals. In particular, our study carries out the following analyses:

1) Comparison of end of day returns with Dataminr signals including lagged effects
2) Comparison of intraday pre-signal and post-signal returns
3) Enhancement of equity short-term reversal strategy by Dataminr signals.

Using Dataminr signal data classified as either Alerts or Flashes and Notables and intra-day asset prices provided by DataMinr, along with publicly available asset price information, we find that assets had significantly higher absolute return values on days with either Dataminr signals. This effect was amplified for higher volatility stocks and is particularly strong in sectors like pharmaceuticals.

Our study of intra-day price dynamics concluded that stock prices movement co-occur with Dataminr signals most strongly in the first half an hour after the signal and that abnormal returns continue over the next three days after the signal. We find that abnormal returns are much higher for Flashes and Notables than for Alerts.

In addition we back-tested a reversal strategy that involves shorting past week winner stocks and rebalancing this portfolio on weekly basis. We have used event volume count by adding up the number of Dataminr signals over the past week and discovered that winner reversal strategy is significantly enhanced by filtering on high event volume stocks.

Finally, the analysis can be extended in many ways, e.g. to cover non-US stocks or take into account sentiment of the signals. The Damaminr signals could also be compared to other signals such as based on Bloomberg news or to study the effectiveness of the signals in identifying broader events not related with individual tickers.

INTRODUCTION

The following report provides details of the analysis carried out by FNA to analyze Dataminr signals data and its relationship with returns of the associated assets. We first describe the data used and then detail analysis carried out.

DATAMINR SIGNALS DATA

The results presented here are based on signal data provided by DataMinr in the files:

- 'alerts.3.1.2014-11.30.2014.dateString.tsv' which includes Alert signals
- 'notable_and_flash.3.1.201411.30.2014.dateString.tsv' which includes Flashes and Notables -signals

From these files we used only columns A and D, which provide the date/time of the signal and the associated asset ticker(s), respectively.

Since we are only counting the number of signals associated with each ticker within a given time span, the content of the tweets themselves was not used. We ignored any signals that had no ticker provided in column D.

As a coarse way to deal with time zone incompatibilities (e.g., European markets close much earlier than US markets, yet the timing of all signals is in EDT), we limited analysis to only signals associated with US tickers.

Finally, any signals made after 4pm EDT (closing time of markets) were considered to have their impact on the next day's prices. We also removed from the data set 7 rows that had a tweet in the column where the date and time should have been.

From here we created a file with the number of signals per ticker per day, a sample of which is shown below.

```
Date,duk,gm,orcl,cvx,nok,lmt,ibm,twc,cmcsa,ba,...
20141128,0,0,0,2,0,0,0,0,1,0,...
20141127,0,1,0,0,0,0,0,0,0,0,...
20141126,0,1,0,0,0,0,1,1,4,0,...
20141125,0,1,0,3,0,0,0,0,2,0,...
```

20141124,0,3,0,2,0,0,1,0,0,0,...
20141123,0,0,0,0,0,0,0,0,1,0,...
20141121,0,0,0,1,0,0,0,1,3,0,...
20141120,0,2,0,0,0,0,0,0,0,0,...
20141119,0,0,0,0,0,0,0,0,1,0,...
20141118,0,0,0,1,0,0,1,0,0,0,...
20141117,1,3,0,0,0,0,0,0,0,4,...
...

We first limited the price and signal files to those tickers and dates common to both. In particular, 55 of the 1993 US tickers in the signal data did not have prices available on Yahoo, so these tickers were excluded from the analysis[1]. There were also a few dates that were not common to both files (for example, there were some signals on Sundays), and those data were removed as well. This yielded 182 days of data on 2,065 assets, with 18,368 Alert signals in total.

Yahoo Price Data

Using Yahoo Finance, we downloaded price data for all US tickers between 13 March, 2014 and 28 November, 2014. From these data we calculated returns for the period between 14 March, 2014 and 28 November, 2014, which corresponds to the dates available in the DataMinr data. Any missing return values during this period were excluded from the analysis.

Intraday Price data

Our intra-day analysis is based on the minute by minute price data taken from the file 'price.1min.march.dec.tgz' provided by Dataminr. We divided the trading day into hour-long intervals into which we binned the signal counts.

[1] These tickers are brkb, wpo, nyx, pbra, gnk, peugy, pc, si, ubs, rdsa, bnjaf, jrcc, pcxcq, vrus, rnftf, barc,

ANALYSIS OF ALERT SIGNALS

End of day Analysis

We first consider whether there is any relationship between Dataminr signal activity and asset returns by comparing average absolute return values on “high" and “low" signal days. We consider two definitions for high and low signal days:

1. one or more signals vs. no signals
2. above vs below 90% percentile

In both cases we test whether the mean absolute return value is significantly different on high versus low activity days. We find that the mean absolute return on days with any signals is 3.0% and the mean absolute return on days with no signals is 1.6%. The p-value for testing equality of the two means is less than 10^{-16}.

For comparing many signals versus fewer signals, we calculate for each ticker the 90th percentile of number of signals per day. All days where the number of signals was higher than the 90th percentile for the given ticker were considered high signal days for that ticker. Defined this way, the mean absolute return value on high signal days is equal to 3.6% and the mean absolute return on days with low signal is equal to 1.6%, with p-value again less than 10^{-16}.

Table 1: Mean absolute returns for Alert -signals

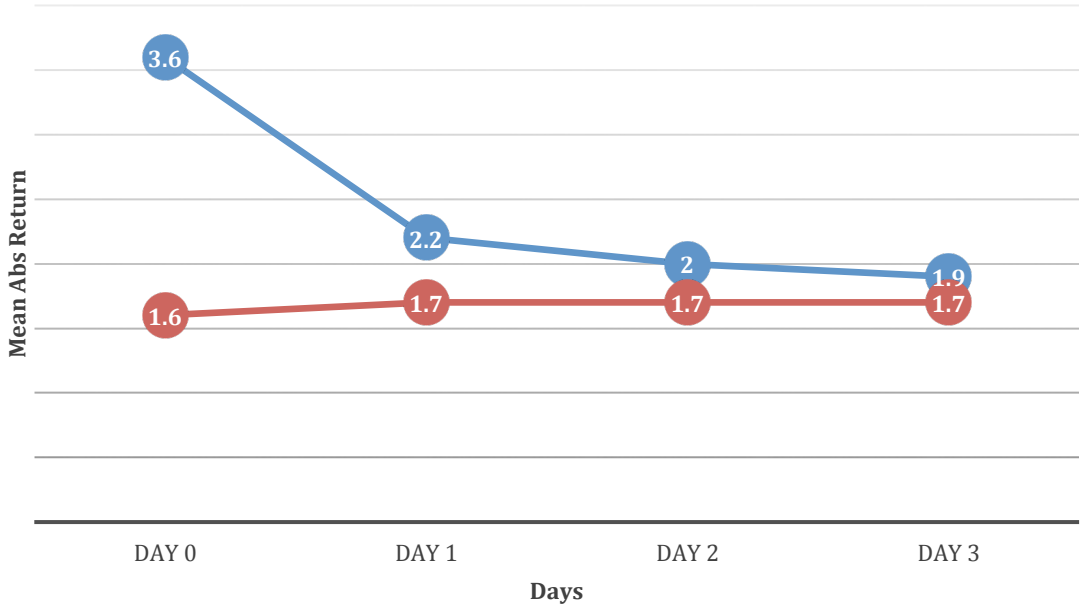
Definition for High vs Low activity	High Activity	Low Activity	p-value
Signals vs no Signal	3.0%	1.6%	10^{-16}
Above vs below 90th percentile	3.6%	1.6%	10^{-16}

Thus we see significantly higher returns on high signal days for both definitions of high and low signal. Because the difference in mean returns on high versus low signal days is greater for our second definition, we use the 90% percentile definition of a signal in the subsequent analyses, unless otherwise noted.

Lagged Effects

We have just shown that absolute return values are higher on signal days. Now we consider whether absolute returns on the days following the signal are also elevated. The table below summarizes results for comparing returns on the same day (as in the previous section) and one, two, or three days following the signal.

Figure 1: Lagged effects of Alert signals



Days Lagged	High Activity Mean abs(return)	Low Activity Mean abs(return)	p-value
0	3.60%	1.60%	< 10-16
1	2.20%	1.70%	< 10-16
2	2.00%	1.70%	< 10-10
3	1.90%	1.70%	< 10-10

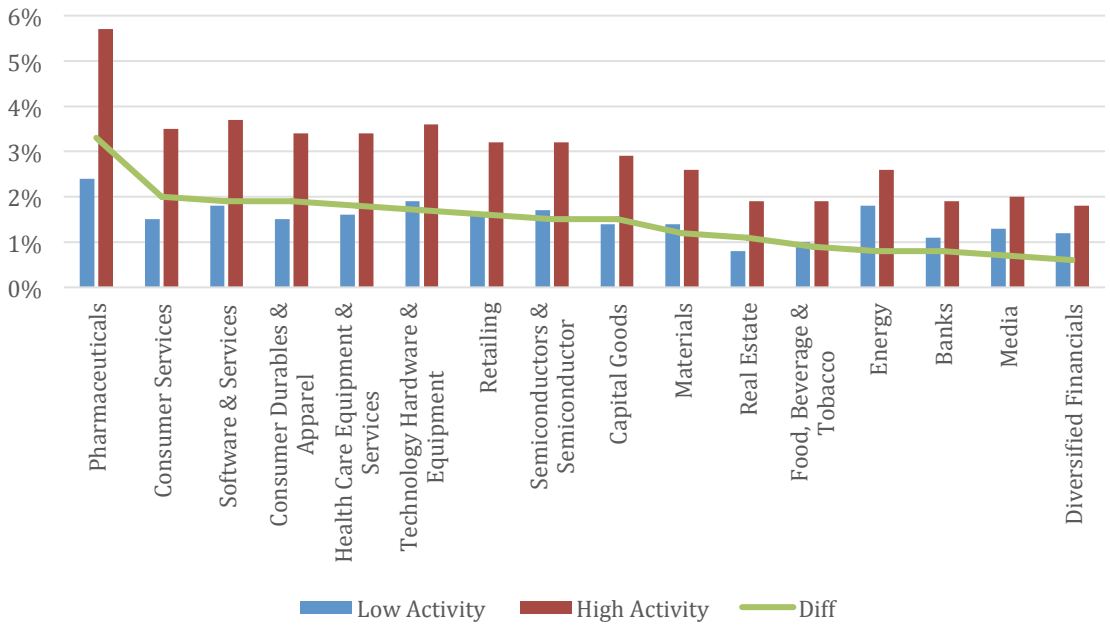
We see that while mean absolute returns remain significantly higher up to three days following a high activity day, the magnitude of the difference gets smaller as the lag increases. Since we see the most impact on returns on the same day as the high activity,

subsequent analyses for sectors, volatility and returns prior to the signal will consider returns on the same day, unless otherwise noted.

Sector Analysis

Now we consider how signals are associated with asset returns across different sectors. The table below shows absolute returns on signal and no-signal days across GICS industry groups. We show results for all groups with at least 50 tickers, ordered by the magnitude of the impact. We find that mean absolute returns are significantly higher ($p < 0.05$) on high activity days in all sectors and particularly in Pharmaceuticals.

Figure 2: Mean absolute returns by sector



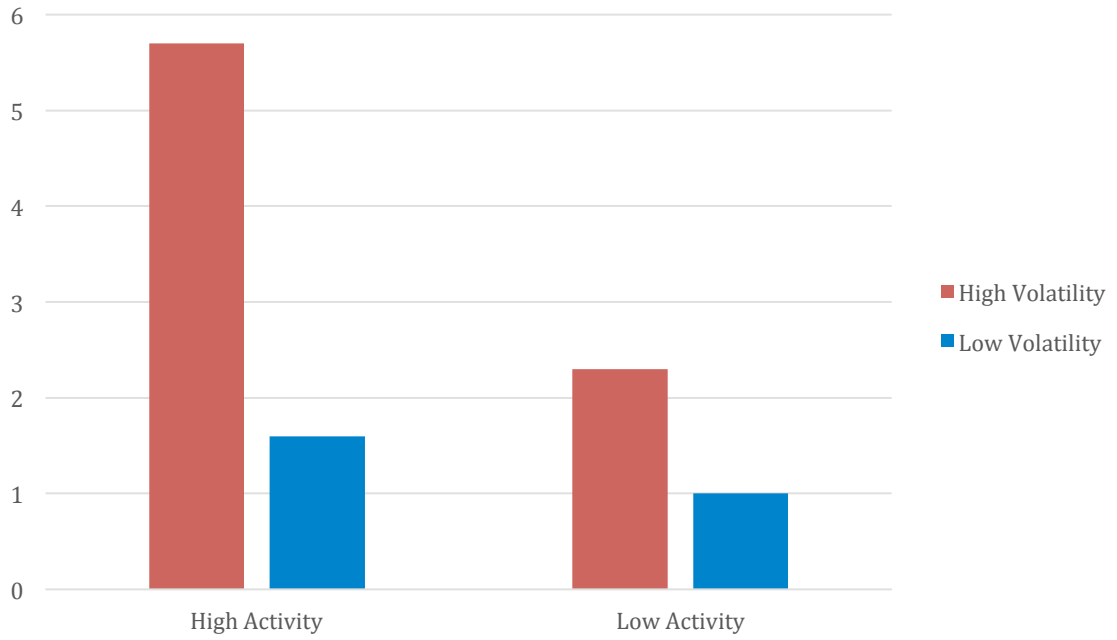
Sector	# of Tickers	Signal day	No-signal day	p-value
Pharmaceuticals, Biotechnology & Life Sciences	257	5.70%	2.40%	$< 10^{-16}$
Consumer Services	79	3.50%	1.50%	0.0001
Software & Services	183	3.70%	1.80%	$< 10^{-11}$

Consumer Durables & Apparel	73	3.40%	1.50%	$< 10^{-7}$
Health Care Equipment & Services	105	3.40%	1.60%	$< 10^{-5}$
Technology Hardware & Equipment	87	3.60%	1.90%	0.0001
Retailing	107	3.20%	1.60%	$< 10^{-13}$
Semiconductors & Semiconductor Equipment	70	3.20%	1.70%	$< 10^{-6}$
Capital Goods	122	2.90%	1.40%	$< 10^{-9}$
Materials	89	2.60%	1.40%	$< 10^{-7}$
Real Estate	54	1.90%	0.80%	0.0037
Food, Beverage & Tobacco	51	1.90%	1.00%	$< 10^{-4}$
Energy	154	2.60%	1.80%	$< 10^{-6}$
Banks	62	1.90%	1.10%	0.0042
Media	56	2.00%	1.30%	$< 10^{-4}$
Diversified Financials	93	1.80%	1.20%	0.0009

Volatility Analysis

To study how the relationship between Dataminr signals and asset returns is impacted by asset volatility, we grouped the tickers into two categories - high and low realized volatility - based on a year of returns between 28 November 2013 and 28 November 2014. We split the groups at the median volatility value so to have the same number of low volatility and high volatility tickers.

Figure 3: Mean absolute returns (1-day) by volatility



Volatility	High Activity Mean abs(return)	Low Activity Mean abs(return)	p-value
Low	1.6%	1.0%	$< 10^{-15}$
High	5.7%	2.3%	$< 10^{-15}$

While the difference in mean absolute returns is significantly different from zero for both high and low volatility assets, the impact is much greater for the high volatility assets.

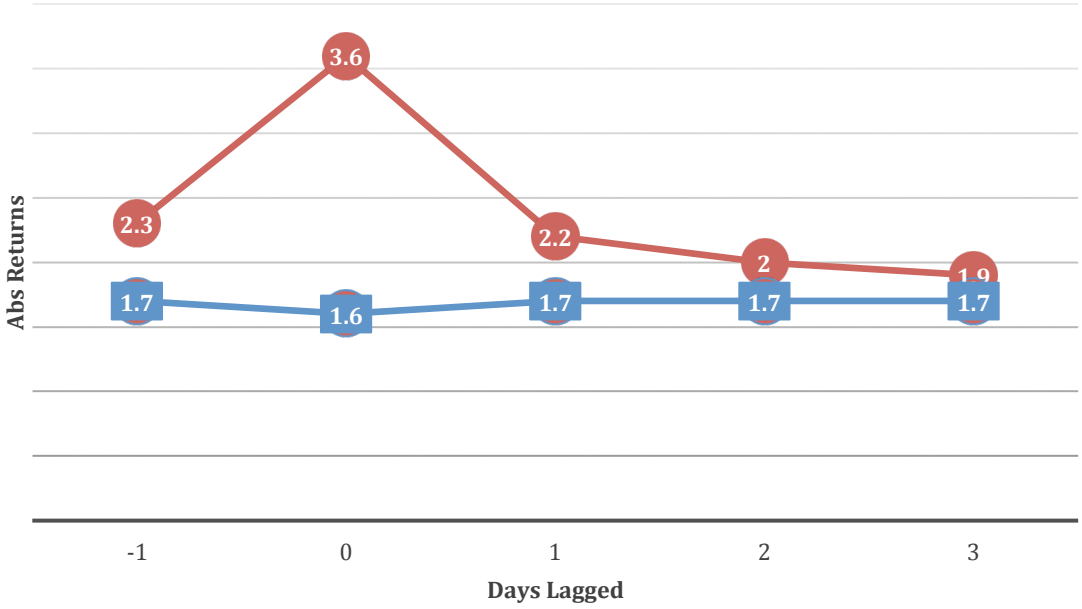
Returns Prior to Dataminr Signals

Finally, we test whether absolute returns are higher on the day before a high activity day. We find results very similar to those for the day after a high activity day: The mean absolute return prior to a high activity day is equal to 2.2% and the mean prior to a low activity day is equal to 1.7%, with p-value less than 10^{-15}.

Combining this result with those from the lagged analysis, we see a trend where returns are elevated the day before the signal, are elevated even more on the signal day, and

remain elevated, to a decreasing extent, on the days following the signal. The figure below shows this trend. We interpret the trend as new information first being reflected in somewhat elevated absolute returns, then being early identified by Dataminr as a signal after which returns are elevated even further, and the elevation diminishes on subsequent days.

Figure 4: Lagged returns around Alert signals



Days Lagged	High Volatility Mean abs(return)	Low Volatility Mean abs(return)	p-value
-1	2.30%	1.70%	$< 10^{-15}$
0	3.60%	1.60%	$< 10^{-16}$
1	2.20%	1.70%	$< 10^{-16}$
2	2.00%	1.70%	$< 10^{-10}$
3	1.90%	1.70%	$< 10^{-10}$

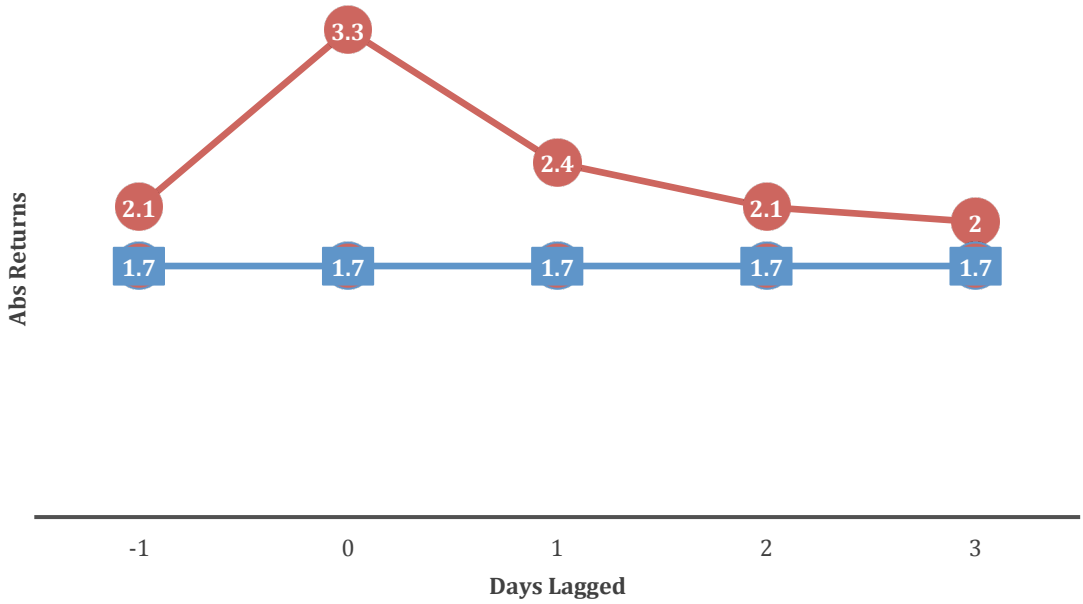
Flashes and Notable Signals

Overall, the results for flashes and notables data are very similar to those for Alerts Signals.

Lagged Effects

Lagged effects are higher on high signal days, peaking on the signal day with absolute returns having very similar values as for the alert signals. The results are statistically significant, but the p-values are somewhat higher than those with alert signals because of the smaller sample size for flashes and notables.

Figure 5: Returns around Flashes and Notables signals



Days Lagged	High Signal Mean abs(return)	Low Signal Mean abs(return)	p-value
-1	2.10%	1.70%	$<10^{-8}$
0	3.30%	1.70%	$<10^{-25}$
1	2.40%	1.70%	$<10^{-7}$
2	2.10%	1.70%	$<10^{-5}$
3	2%	1.70%	$<10^{-6}$

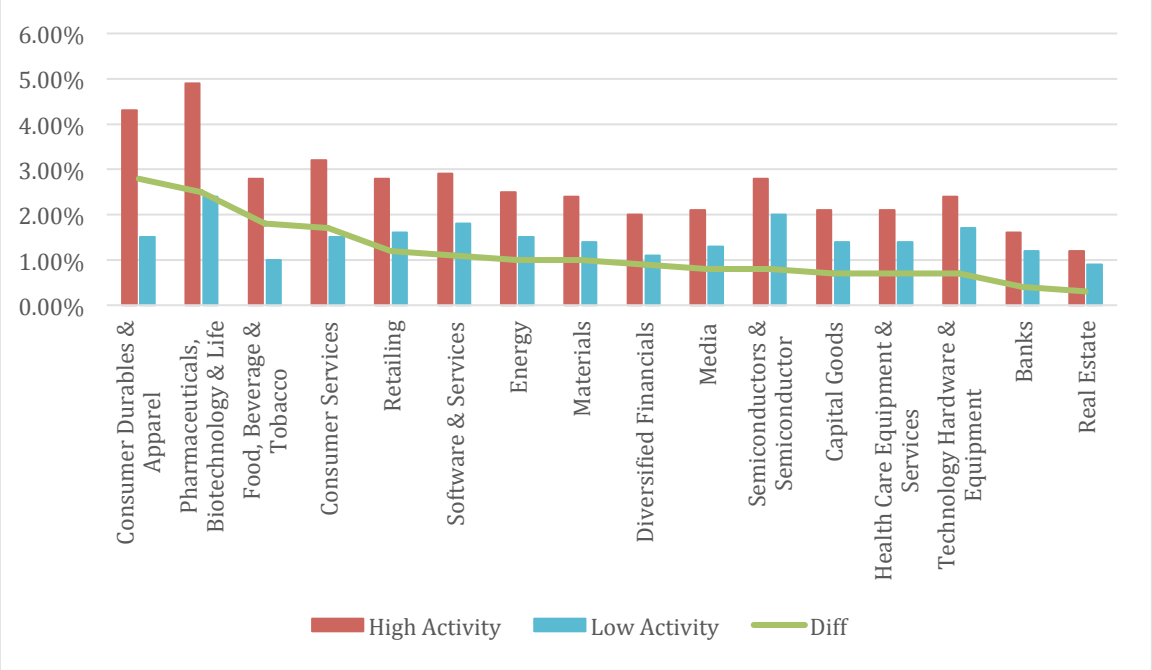
Sector Analysis

Pharmaceuticals again show the highest difference in absolute returns between the high and low signal days. The results show consistently higher returns on high signal days in all sectors.

Sectors with higher volatility highlight the effect much more with absolute returns almost doubling on high signal days. Lower volatility sectors show higher absolute returns on high signal days, however the effect is not nearly as strong.

Results are mostly statistically significant, but again with higher p-values due to the smaller sample size.

Figure 6: Mean absolute returns by sector



		Signal	No Signal	
Sector	# of tickers	Mean abs(return)	Mean abs(return)	p-value
Pharmaceuticals, Biotechnology & Life Sciences	257	4.9%	2.4%	$4.5*10^{-8}$
Software & Services	183	2.9%	1.8%	0.0002
Energy	154	2.5%	1.5%	0.0032
Capital Goods	122	2.1%	1.4%	0.0333
Retailing	107	2.8%	1.6%	$1.4*10^{-8}$
Health Care Equipment & Services	105	2.1%	1.4%	0.1186
Diversified Financials	93	2%	1.1%	0.0750
Materials	89	2.4%	1.4%	0.1343
Technology Hardware & Equipment	87	2.4%	1.7%	0.0113
Consumer Services	79	3.2%	1.5%	0.0418
Consumer Durables & Apparel	73	4.3%	1.5%	0.0002
Semiconductors & Semiconductor Equipment	70	2.8%	2%	0.1044
Banks	62	1.6%	1.2%	0.2686
Media	56	2.1%	1.3%	0.0058
Real Estate	54	1.2%	0.9%	0.3121
Food, Beverage & Tobacco	51	2.8%	1%	0.0019

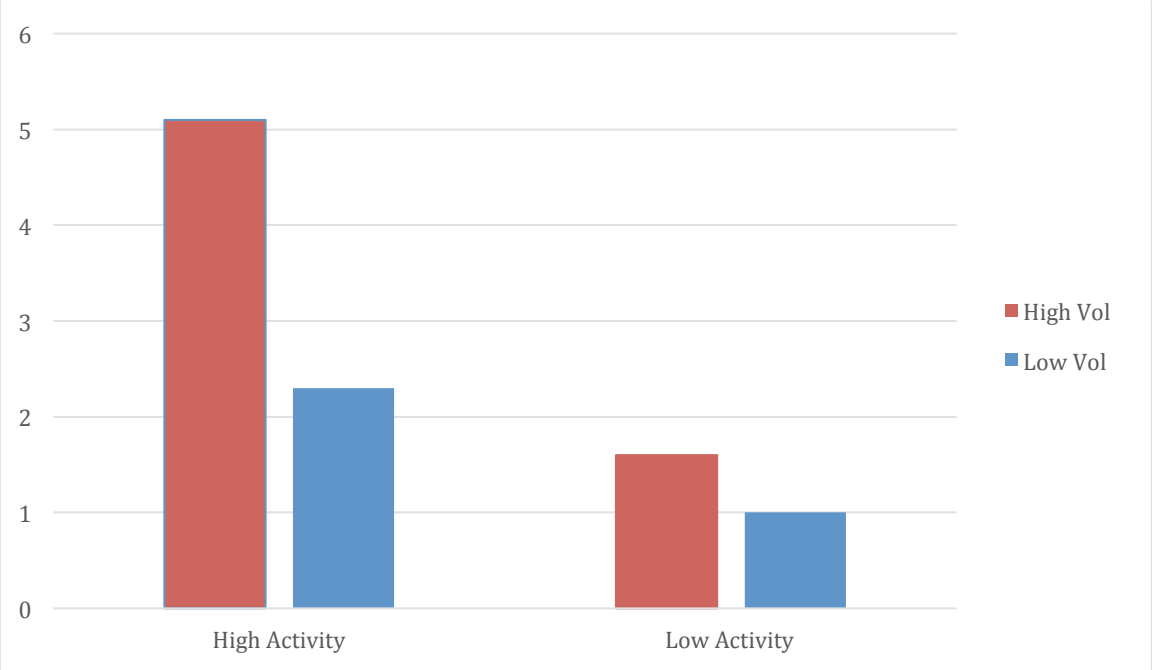
Results for the sectors listed below are not statistically significant at level 0.05.

- Health Care Equipment & Services
- Diversified Financials
- Materials
- Semiconductors & Semiconductor Equipment
- Banks
- Real Estate

Volatility Analysis

Analysis based on stock volatility percentiles for flashes and notables looks almost identical to alert signals.

Figure 7: Mean absolute returns by volatility



Volatility	High Activity Mean abs(return)	Low Activity Mean abs(return)	p-value
Low	1.6%	1.0%	$< 10^{-15}$
High	5.1%	2.3%	$< 10^{-15}$

INTRADAY ANALYSIS

The objective of this analysis is to see whether Dataminr signals are related to intraday returns. Both signal timing and asset prices are available down to the minute. However, since the signal data is fairly sparse, we bin the data to hourly buckets. To assess the impact of signals on intraday returns, we compare the mean absolute intraday return from 10am to 4pm on days where the asset had one or more signals between 9:30am and 10am versus days where the asset had no signals during that time. Because it was rare for tickers to have more than one signal in such a short time period, high activity here is defined as one or more signals.

Table 2: Alerts signals from 9:30 am - 10 am

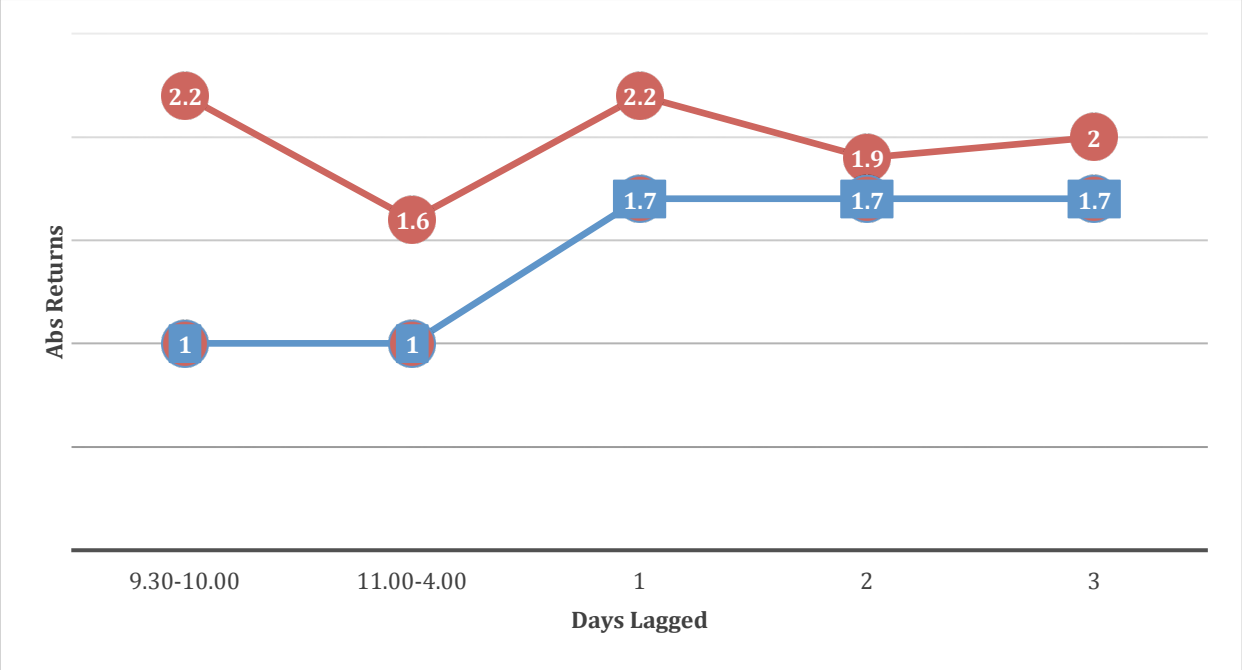
Return Period	Signal present, Mean abs(return)	No Signal, Mean abs(return)	p-value
10:00 am – 4:00 pm	1.9%	1.3%	0.0019

Table 3: Notable and Flash signals from 9:30 am - 10 am

Return Period	Signal present, Mean abs(return)	No Signal, Mean abs(return)	p-value
10:00 am – 4:00 pm	1.9%	1.2%	0.00014

Now let us look at the second hour that had the most signals, the hour from 10 am and 11 am. We can compare returns during the pre-signal hour, i.e. from the market open until 10 am, and returns from 11 am until the market closing at 4 pm. In addition we also consider returns from after the signal at 11 am until 4pm the following day (holding the stock for 2 days) and until 4pm two days later (holding stock for 3 days). It looks like it makes sense to hold stocks for more than 1 day since the excess daily returns compared with the day when there are no signals remain elevated.

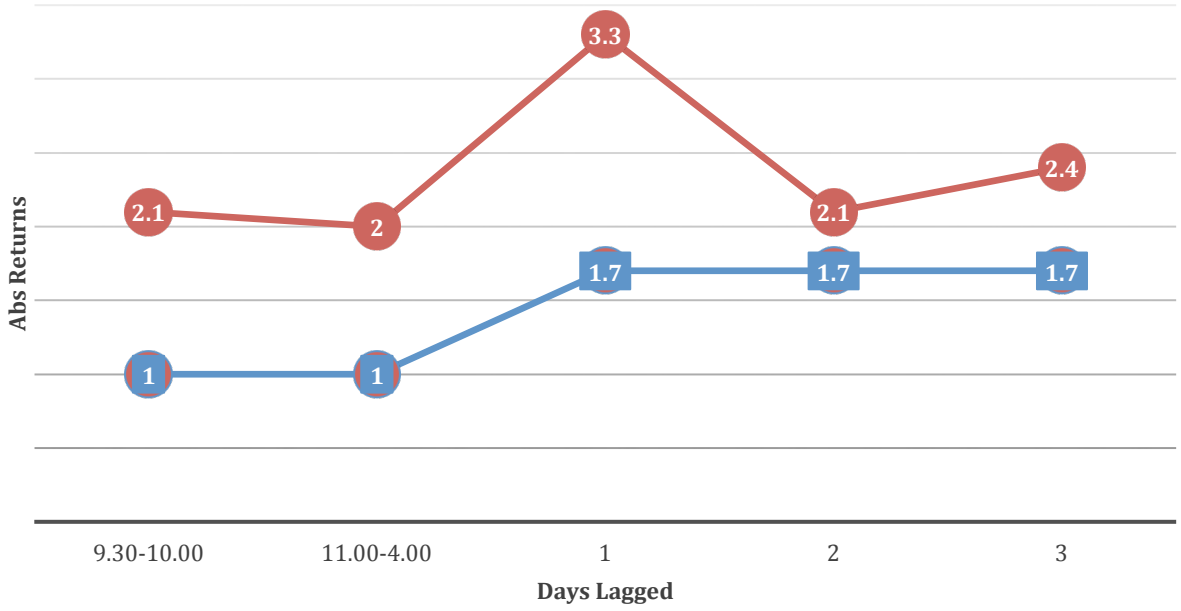
Figure 8: Intraday returns around Alert signals



Return period	Signal, Mean abs(return)	No Signal, Mean abs(return)	p-value
9:30 am – 9:59 am	2.2%	1.0%	$2.341*10^{-15}$
11:00 am – 4:00 pm	1.6%	1.0%	$1.714*10^{-8}$
1 day lag	2.2%	1.7%	$6.121*10^{-6}$
2 days lag	1.9%	1.7%	0.0071
3 days lag	2.0%	1.7%	0.0005

This especially pronounced for Flashes and Notables where cumulative excess return reaches 1.3% on the 3rd day.

Figure 9: Returns around Flashes and Notable signals

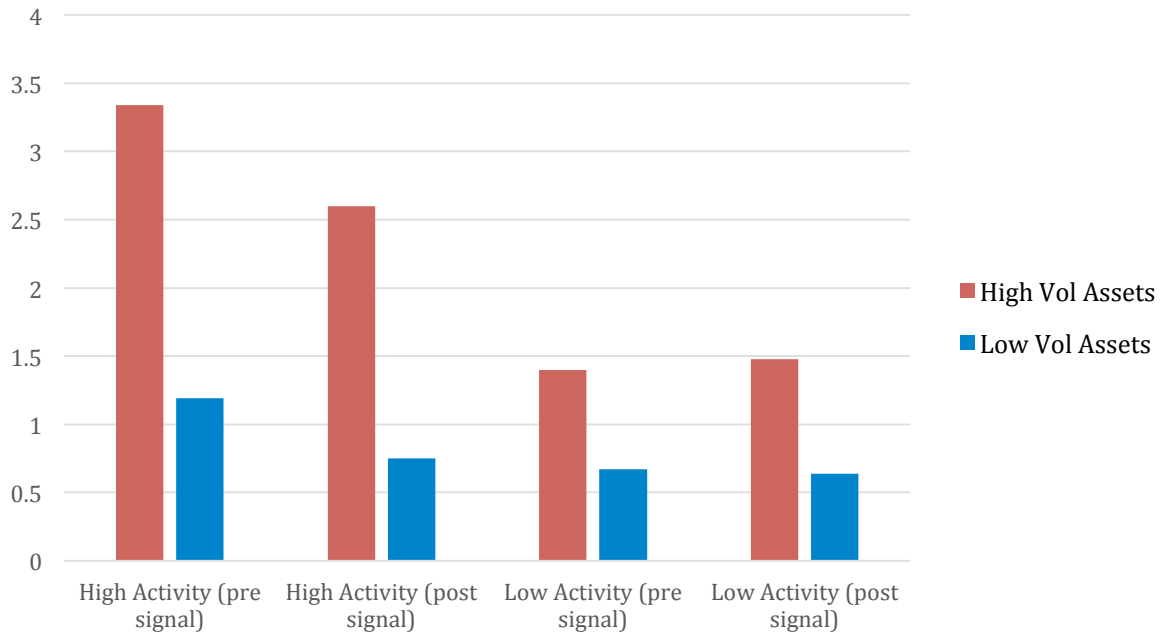


Return period	Signal Mean abs(return)	No Signal Mean abs(return)	p-value
9:30 am – 9:59 am	2.1%	1%	0.0009
11:00 am – 4:00 pm	2.0%	1%	$1.321*10^{-5}$
1 day lag	3.3%	1.7%	0.0124
2 days lag	2.1%	1.7%	0.0046
3 days lag	2.4%	1.7%	0.0001

Volatility Analysis

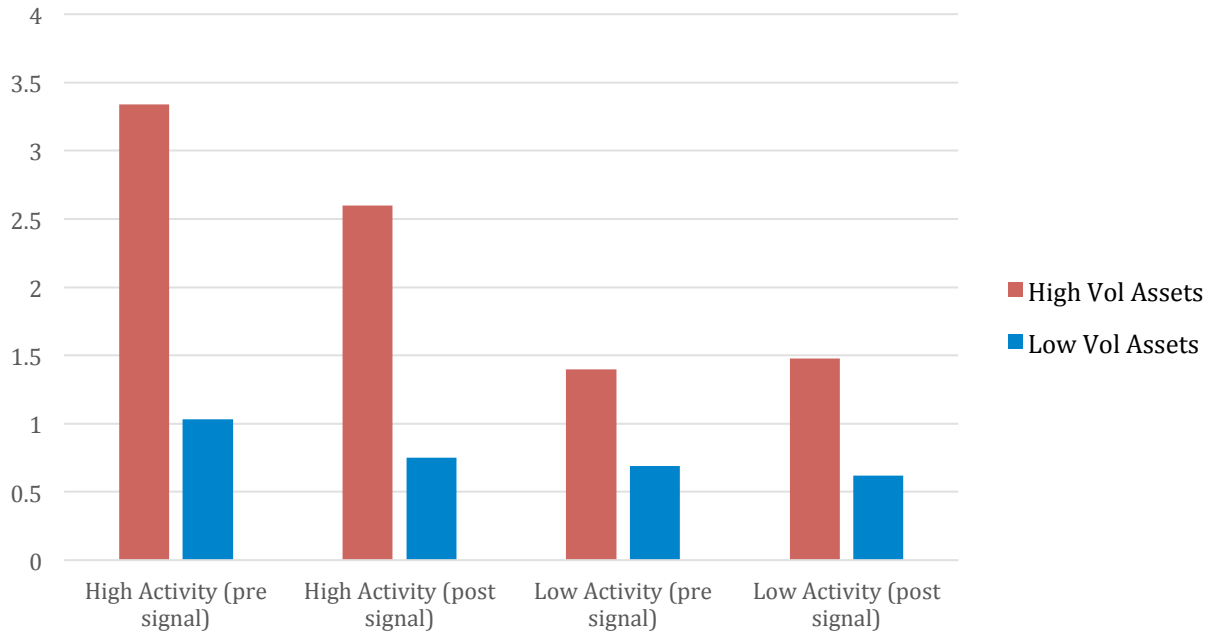
In this analysis we separate the stocks by volatility and analyze the signals for tickers with high and low volatility. We find that on high signal days returns are much higher both pre- and post- signal for high volatility tickers whereas for low volatility tickers the difference in returns on low vs high signal days is not very pronounced. Nearly all results are statistically significant, with p-values smaller (more significant) for alert signals.

Figure 10: Mean absolute returns by volatility for Alert signals



Return period	High Signal, Mean abs(return)	Low Signal, Mean abs(return)	p-value
Low volatility assets			
9:30 am – 9:59 am	1.19%	0.67%	$3.86*10^{-9}$
11:00 am – 4:00 pm	0.75%	0.64%	0.0003
High volatility assets			
9:30 am – 9:59 am	3.34%	1.4%	$3.814*10^{-11}$
11:00 am – 4:00	2.6%	1.48%	$4.635*10^{-8}$

Figure 11: Mean absolute returns by volatility for Flashes and Notables signals



Return period	High Signal, Mean abs(return)	Low Signal, Mean abs(return)	p-value
Low volatility assets			
9:30 am – 9:59 am	1.03%	0.69%	0.0042
11:00 am – 4:00 pm	0.75%	0.62%	0.0200
High volatility assets			
9:30 am – 9:59 am	3.34%	1.4%	0.0047
11:00 am – 4:00	2.6%	1.48%	$4.591*10^{-5}$

Sector Analysis

In this section we analyze intra-day results within industry sectors. For the second hour that had the most signals (10 am - 11 am), we compare absolute returns during the pre-signal hour, i.e. from the market open until 10 am and returns from 11 am until the market closing at 4 pm.

In nearly all cases the mean absolute return is higher when there was at least one signal vs. when there were no signals, both before and after the 10am-11am

window. However, we note that for many sectors the results are not statistically significant.

Figure 12: Mean absolute returns (9:30am – 10am) by sector

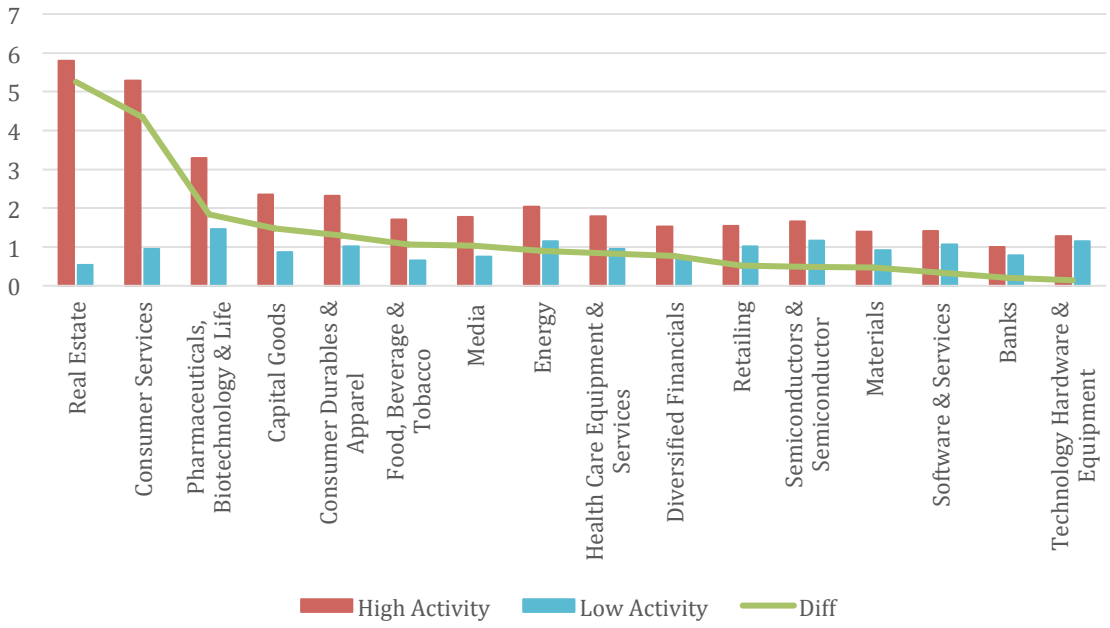


Figure 13: Mean absolute returns (11am– 4pm) by sector

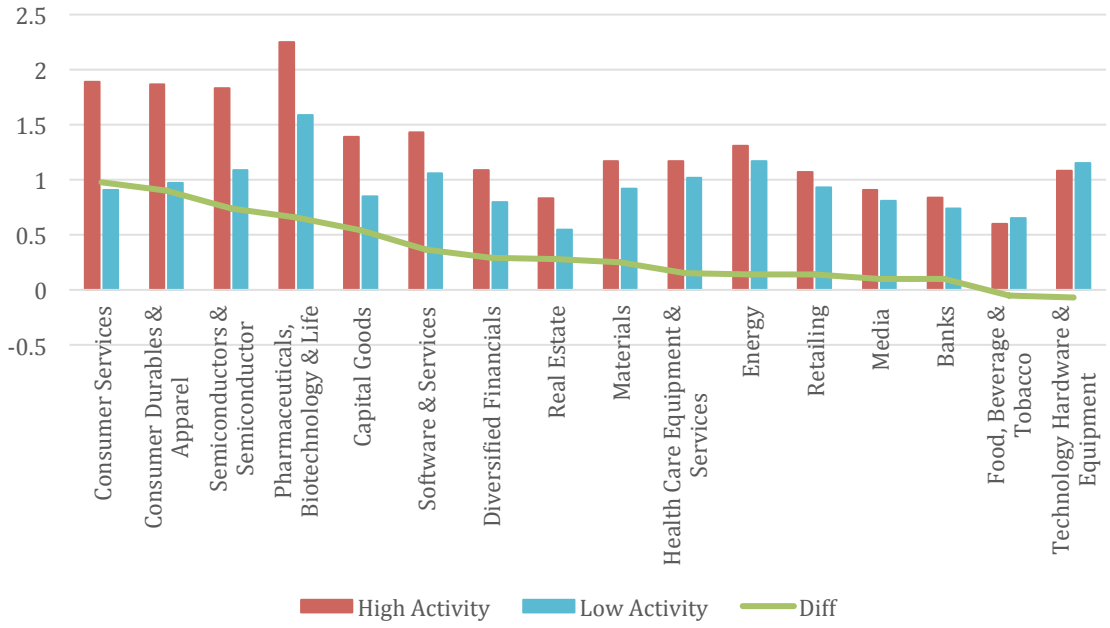


Table 4: Alerts Signals from 10 am - 11 am

		Signal	No Signal	
9:30 am – 9:59 am	# of tickers	Mean abs(return)	Mean abs(return)	p-value
Consumer Durables & Apparel	73	2.32	1.02	0.0219**
Consumer Services	79	5.29	0.95	0.0806*
Pharmaceuticals, Biotechnology & Life Sciences	257	3.30	1.46	1.76*10-5***
Semiconductors & Semiconductor Equipment	70	1.66	1.17	0.1108
Health Care Equipment & Services	105	1.79	0.95	0.4701
Capital Goods	122	2.35	0.87	0.0124**
Energy	154	2.04	1.14	0.1321
Materials	89	1.39	0.92	0.2211
Software & Services	183	1.41	1.07	0.0550*
Retailing	107	1.54	1.02	0.1102
Real Estate	54	5.08	0.54	0.1266

Diversified Financials	93	1.53	0.76	0.2725
Media	56	1.78	0.75	0.0451**
Technology Hardware & Equipment	87	1.28	1.14	0.6957
Banks	62	1.00	0.79	0.1820
Food, Beverage & Tobacco	51	1.71	0.65	0.0110**

Table 5: Alerts Signals from 11 am – 4pm

11:00 am – 4:00 pm				
Consumer Durables & Apparel	73	1.87	0.91	0.0685*
Consumer Services	79	1.89	0.97	0.1993
Pharmaceuticals, Biotechnology & Life Sciences	257	2.25	1.59	0.0040***
Semiconductors & Semiconductor Equipment	70	1.76	1.14	0.0072***
Health Care Equipment & Services	105	1.45	0.95	0.2015
Capital Goods	122	1.39	0.85	0.0831
Energy	154	1.31	1.17	0.6813
Materials	89	1.17	0.92	0.3910
Software & Services	183	1.43	1.06	0.0211**
Retailing	107	1.07	0.93	0.1821
Real Estate	54	0.83	0.55	0.4712
Diversified Financials	93	1.09	0.8	0.3614
Media	56	0.91	0.81	0.357
Technology Hardware & Equipment	87	1.08	1.15	0.8096
Banks	62	0.84	0.74	0.5095
Food, Beverage & Tobacco	51	0.60	0.65	0.6940

The only sector whose results are statistically significant for both pre- and post-signal analysis at 5% level was Pharmaceuticals, Biotechnology & Life Sciences.

For Notable and Flash -signals, most of the sectors don't have enough data for meaningful analysis

REVERSAL STRATEGY

Finally, we consider a price reversal strategy on stocks[2] and test whether solely returns based strategy can be improved by taking Dataminr signals into consideration.

In order to have reliable and liquid asset price information and a somewhat uniform signal distribution among the assets we consider stocks in S&P 100 – a broad market index with the most liquid constituents.

We will look into weekly stock returns since the above analysis indicates that the signals have an approximately weekly information diffusion cycle. We first calculate weekly returns from 2014/03/14 till 2014/11/28 and then compute excess returns with respect to the average return for S&P 100 constituents for the corresponding week. We then consider the stocks that are in the highest quintile based on their excess return ("winners") and expect their prices to revert back on the following week.

Running such winner reversal portfolio for the time period from 2014/03/14 to 2014/11/28 we find that for that period of time its cumulative return is **0.85%.** Since the strategy consists of shorting the winner stocks, running this strategy will actually sustain small losses over the above-mentioned period of time.

Let us now introduce a statistic based on the Dataminr signals called event volume. Event volume is simply the number of the Alert signals for the given stock in the past week.

If we add event volume and narrow down our selection of stocks to be shorted to those that are in the highest quintile of returns and in the highest quintile based on their event volume for the past week, we get the cumulative return of **-12.42%** over the same time period. This is a significant improvement and in fact presents a good opportunity for enhancing the strategy.

Repeating the above procedure for winner stocks portfolio but changing the filter to the lowest event volume quintile stocks get a cumulative return of **-1.91%** for the strategy. This is still an improvement over solely return based strategy but the result is not nearly as good as for the stocks in the highest event volume quintile.

[2] See http://www.investopedia.com/terms/r/reversal.asp

Looking into a loser stocks portfolio, i.e. the stocks that showed lowest excess returns for the past week and going long with those stocks for the coming week produces a return of **9.54%** for the same period of time. This is also a good result.

If in addition we choose the loser stocks to be in the lowest event volume quintile (EV=0), we get a **9.13%** overall return. This is actually slightly worse that for all stocks without taking the signals into consideration. We think it may be because there are not enough signals that can produce a meaningful low quintile for the event count.

Looking into losing stocks portfolio with highest event volume we get an even worse result – overall return is only **2.53%.**

These results are summarized in the table below. The best strategy is to short past winners that had a high Event Volume based on Dataminr signals.

Table 6: Reversal strategy returns with and without Dataminr signals

	Returns Only	Low EV	High EV
Winners	0.85 %	-1.91%	**-12.42%**
Losers	9.54 %	9.13%	2.53%

FURTHER EXTENSIONS

Several extensions to the above analysis and for investigating other aspects of the effects of Dataminr signals on stock returns can be thought of.

1) Extension of the above analysis to non-US stocks.
2) Taking into account the sentiment of the signals, and identifying whether the price impact is likely positive or negative.
3) Comparison of Dataminr signals with mainstream news releases such as Bloomberg – the objective here would be to understand if acting on Dataminr signals provides an edge to trading strategies using mainstream news as signals ;
4) To study signals that are not associated to any tickers – such as signals indicating political unrest or a start of epidemics. We would like to research if one can induce regional or industry specific market dynamic from early warnings derived

from Twitter signals related to geopolitical events. We expect a lag from the signals in such cases to be measured in months ;

5) Explore consumer sentiment effect associated to new products on company stock prices (such as Apple iPad or iPhone success with consumers).