Google™ News, Volatility and the Stock Market

HANS BYSTRÖM
Department of Economics
Lund University
PO Box 7082
220 07 Lund
Sweden
Phone: +46 46 2229478
Fax: +46 46 2224118
contact: hans.bystrom@nek.lu.se

In this paper we employ the news aggregator Google™ News, provided by Google Inc., to shed some light on the link between stock market volatility and the amount of news available to market participants. We believe this to be the first time a news aggregator such as Google News is used in financial research. As a proxy for the amount of news available to the market we collect stock market-related news by searching on particular word strings in Google News. Our findings indicate a strong correlation between market volatility and news volume; in months when large amounts of stock market-related news is available to the market, the volatility of the stock market is higher than when less news is available. Based on these findings, and as an example of how news aggregators potentially could be used, we suggest a truly global *investor fear gauge* constructed along the lines of the VIX index but based on (negative sentiment) news volumes in circulation (world-wide) instead of on market volatility (in the US). The approach taken in this paper is of course not limited to the study of stock market volatilities, and we believe that news aggregators such as Google News could be used by researchers, policy makers and investors in a range of different areas.

Keywords: Google News; news aggregator; news; volatility; stock market
JEL classification codes: G10; G14; D83

In this paper we use Google™ News, the news aggregator, to investigate how the amount of stock market-related news (world-wide) over a particular time period is related to the volatility of the global stock markets over the same time period. The effect of news on market volatility has been studied before, but we believe our study to be the first that uses news aggregators to specify the news volume dynamics.

The literature on news and the effect it has on financial markets is extensive. The share of this literature focusing particularly on the relationship between volatility and news, however, is somewhat smaller. In one of the earlier studies, French and Roll (1986) find significantly higher market volatility during trading hours than during non-trading hours. They accredit the difference to the variability in the information flow. Mitchell and Mulherin (1992), in turn, analyze the effect of public information on stock market activity (which is related to volatility), finding only a weak link between the two. In another study, Berry and Howe (1994) relate the number of news releases by Reuter's

News Service to stock price volatility and although the volatility is found to vary across time in a fashion similar to that of the news volumes, results from regressions show no significant relationship. Looking instead at interest rates and foreign exchange futures markets, Ederington and Lee (1994), find scheduled macroeconomic news announcements to explain a substantial share of the market volatility. Finally, in a more recent study, Johnson and Marietta-Westberg (2004) find that the increase in idiosyncratic stock return volatility observed in recent decades is significantly related to the increase in firm-specific news.

In this paper we rely on modern developments in web-based news aggregation in order to study the news volume and its effect on market volatility. In a way, our research is based on how certain modern market participants approach news. Today, many hedge funds, proprietary trading desks and other sophisticated investors use news content in a systematic way to make money. In doing so, an increasing share of these market participants are also using specially programmed computers to try and outsmart their rivals (van Duyn (2007)). In some cases, the news that is feed into the machines is not even interpretable by humans; news releases can be generated *by* computers *for* computers and this machine-machine interaction shortens the time delay between the release of the news and the trading on the news. Moreover, several companies are, allegedly, turning to the internet to capture market gossip. Measures of online activity, such as the number of times an issue is raised in blogs, news articles and scientific articles, could be extracted from news aggregators/filters and used in computer-based trading strategies (Scholtes (2006)).

An interesting example of an initiative along these lines is Google News. It is a news aggregator provided by Google Inc. that collects and presents web news as clusters on a particular topic chosen by the viewer. What makes Google News particularly interesting is that it collects the featuring news in a purely automatic fashion without intervention by human editors. In this way, political, ideological and other biases are efficiently eliminated from the news selection process. Currently, the English language version of Google News gathers news stories from more than 4500 news sources.[1]

The novel idea in this paper is to build on this modern development to answer traditional questions. In our case the question is whether there is a link between news volumes and market volatility. We use Google News to specify the volume of news on a certain topic (in our case the stock market) at a certain point in time. In this way, the entire dynamics of the news volume, and not only snapshots around certain news-generating events, can be traced. Moreover, the linkage between news volume dynamics and asset price dynamics can also be studied.

We limit ourselves to the link between news volume and stock market volatility and our main finding is that in months with many stock market-related news-releases the volatility of the stock market is higher than in months when less news is available. We would like to stress, however, that there are many other potential areas where news aggregators such as Google News could serve as useful tools for investors and/or academics.[2] Traders, for instance, could use Google News volumes to try and predict movements in various financial asset prices. Volatility sensitive asset classes such as options and, of course, the VIX volatility-index itself, are particularly likely candidates. Risk managers, in turn, could use the amount of news on a particular topic, such as stock market crashes, as an early warning signal; it is possible that the news volume leads the

market. Academics and policy makers, finally, could of course use proxies for the aggregated amount of news on a particular topic, say inflation, in models and policy decisions. And of course, the usage is by no means limited to finance and economics. Psychology, sociology and many other social sciences where news play a role could also profit from the use of news aggregators in research.

The organization of the paper is as follows. The next section gives a brief introduction to Google News, the news aggregator provided by Google Inc. The third section describes the data and the fourth section presents empirical evidence on the dynamics of news and the correlation between news volumes and stock market volatilities. The final section concludes the paper.

GOOGLE NEWS

Google News is an example of a so-called automated news aggregator that, without human intervention, collects, reorganizes and presents web news as clusters. Google News gathers news from a multitude of sources on the web, and currently, there are twelve different language versions of the aggregator. Each day (the search is updated every 15 minutes), Google News includes news articles that at some point during the past thirty days appeared in one or more of hundreds or thousands of news sources. The English-language version currently gathers news stories from more than 4500 news sources. The stories collected by Google News are selected by computer algorithms based on their frequency of appearing on selected sites. No human editors are used, which minimizes potential political or ideological bias in the selection process (Google (2007)).

Google News is very easy to use. You type the search word(s) that specifies the topic on which you want to gather news in a search box and the news stories appear sorted according to relevance. The total number of news entries is also presented, and this is the figure that is used in this paper as a proxy for news volume. For the more advanced users, special features such as personalized news, news-history tracing and country- or source-specific searching is also available. These features are not used in this paper.

DATA

The purpose of this study is to investigate the link between (stock market relevant) news volume and stock market volatility. In order to do this we collect neutral as well as negative news related to the stock market using the news aggregator Google News. More specifically, the volume data that we retrieve is the number of news stories appearing in the English language version of Google News when searching on the strings *stock market* and *stock market crash*, respectively. The data is collected on a daily basis over the time period August 8, 2006 to October 23, 2007.[3] The data is gathered in the morning (CET), approximately at the time of the opening of the major European stock exchanges. Consequently, the collected news on a particular day represents the situation *before* that particular day's opening of the US stock exchanges and *after* that day's opening of the Asian stock exchanges. We believe this way of collecting the data to be consistent with the "global investor" point of view taken in this study.

In addition to the Google News data, we have also collected stock market data.[4] Since the number of news stories listed in Google News on a particular day is related to the amount of news released over the last thirty days, one should preferably use stock market

data that can be readily represented as monthly averages. Since our focus is on volatilities, this causes no problems. In addition, contrary to stock prices for instance, there is no reason to believe our volatility and news volume data to be non-stationary.

As a proxy for the global stock market capitalization we use the MSCI ACWI index (an all country world index that includes stocks from 48 developed and emerging markets) and as a proxy for the global (conditional) volatility we use the historical standard deviation of the MSCI ACWI stock index returns over the last thirty days. As an additional volatility estimate we also use the CBOE VIX index of implied S&P500 volatilities.[5] This index is sometimes referred to as the *investor fear gauge* and since we want to link this fear indicator to the amount of news in circulation over the last thirty days we use an average of the last thirty days daily VIX quotes. This is an average of the last month's "investor fear" and we want to see whether there is any link between the fear level and the amount of available stock market news.

NEWS AND STOCK MARKET VOLATILITY

In this section we present some new evidence on the link between news volume and stock market volatility using time series samples of (time-varying) volatilities and news volumes. We calculate simple correlations between the number of news hits in Google News and the stock market volatility.

We use two search strings in Google News; *stock market* and *stock market crash*. The news gathered in this way is of course a very small part of the total news universe but we believe that these two strings capture the essence of the news volume relevant for the global stock market. One of the strings is clearly an example of negative news regarding the future stock market performance while the other is neutral in its stance. The volatility, in turn, is measured in two ways, as the historical 30-day sample standard deviation of the MSCI index returns and as the historical 30-day average of the VIX implied volatility.

One valid concern is whether the dynamics of the search string result for *stock market*, for instance, in fact captures the dynamics of the number of news articles on *stock market*, and not the number of articles in Google News more generally. The total amount of news available in Google News is likely to vary over time and in order to filter that variation out, we adjust the search results (the number of hits on a particular day) by dividing the number of hits on *stock market* by the sum of the number of hits on the search words *stock* and *market*, separately. This is of course not a perfect normalization but a scan of the entire news universe is impossible. Our adjustment at least tries to account for variations in the amount of news containing the words *stock* or *market*, separately. The same approach is applied to the search string *stock market crash* where all the three words are taken care of in the same fashion. Results for *adjusted* as well as *non-adjusted* volumes are presented below.

Figure 1 shows the movements of the MSCI ACWI stock index over the time period. The global stock market was struck by three small or medium-size corrections/crashes over this particular time period and, being a global stock index, the MSCI ACWI index of course reflects this. Other than that, the stock market of 2006 and 2007 must be considered a bull market with a clear positive trend. The most significant correction in the stock index level is associated with the sub-prime mortgage/structured credit crisis hitting the market in mid-2007.

Now, this study focuses on the volatility of the stock market. That is, we are mainly interested in the *magnitude* of the stock price movements, not their *directions*. The volatility, calculated on a daily basis, of the MSCI ACWI index, together with the VIX volatility, again calculated on a daily basis, is pictured in Figure 2. As expected, the volatility increases in times of market stress. During the credit crisis the level of volatility is about twice the level seen during more "normal" time periods.

The dynamics of the news volumes, in turn, is presented in Figures 3 and 4.[6] Again, the data is sampled on a daily basis. A comparison of Figure 2 and Figure 3 reveals that the number of news pieces containing the word string *stock market*, i.e. pieces of news containing negative, neutral or positive opinions on the stock market (on average neutral), increases at about the same time as the stock market volatility. What is more, this holds regardless of whether we adjust the volume series as described above or not. This is interesting, since it indicates that market volatility, in deed, does increase with the volume of market relevant news in circulation. The results are, qualitatively, the same when the number of news pieces containing the word string *stock market crash* in Figure 4, i.e. on average negative pieces of news, are compared to the volatilities in Figure 2. Again, whether the news series is adjusted or not has almost no effect, at least not on this purely qualitative discussion. In fact, the adjustment has an even smaller effect here than earlier. Finally, the high co-variation between news and volatility is also clearly visible in the scatter plots between volatilities and news volumes (non-adjusted) in Figures 5 and 6.

To quantify the findings in the previous paragraph, we calculate the ordinary Pearson correlation coefficient between the news volume time series and the stock market volatility time series. The results can be found in Table 1 and the numbers reveal that the correlations, overall, are fairly high. In some cases, most notably for the non-adjusted neutral *stock market* news volumes, the news-volatility correlations are well above 0.7, which must be considered very high. To summarize, in months when a lot of stock market-related news is available, the volatility of the stock market is higher than when less news is available.[7]

The two different volatility measures have fairly similar correlations with the news volume. As we would have expected, however, the investor fear gauge, i.e. the average VIX index level over the last thirty days, is more strongly correlated to the search string *stock market crash* (representing negative investor sentiment) than the historical MSCI AWCI volatility is. The opposite holds for the search string *stock market* (representing neither positive nor negative investor sentiment). Also, overall, *non-adjusted* volume series show a stronger link with stock market volatility than *adjusted* volumes. We have no good explanation for why that is the case.

Finally, the results in this paper raise the issue of whether there is a potential link between news volumes and the phenomenon of volatility clustering. We find some evidence of news clustering and it is possible that this clustering is related to the clustering of the stock return volatility. Basically, there could be a feed-back mechanism between news volumes and market volatility; when market volatility increases more market-related news is released, causing volatility to remain elevated, which in turn could result in more news releases that could help keeping volatility at higher levels, etc. This line of reasoning is not pursued further in this paper and any investigation of this potential link is left to future studies.

CONCLUSIONS

We believe this study to contribute to the literature in at least two ways. First, by using a new method of measuring the volume of relevant news available to market participants, based on so-called news-aggregators, we try to assess the strength of a possible link between stock market-related news volumes and stock market volatility. And second, more generally, in doing so, we hope to convince the reader about the potential benefits of using news aggregators in all sorts of research, policy, or trading activity (for free or at a cost, depending on the stance of the owner of the aggregator).

We have used a particular news aggregator called Google News, provided by Google Inc., to investigate the link between the stock market volatility over a certain time period and the amount of news available to market participants over the same time period. We believe this to be the first time news aggregators are used in financial research, and as a proxy for the amount of news available to the market participant we have collected neutral as well as negative market-related news by searching on particular word strings in Google News.

A simple correlation study reveals a strong link between volatility and news volume. We find that in months with an abundance of stock market-related news, the volatility of the stock market is higher than when less news is available. The causal direction between news and volatility is not studied in this paper but the link between the two seems to be quite strong. One related issue is whether news is the cause of the well known phenomenon called volatility clustering. The causes of volatility clustering are not fully understood and it is possible that there is a link between news (clustering) and volatility clustering.

There are of course many other areas where news aggregators potentially could serve useful. As an example, an alternative investor fear gauge could be constructed along the lines of the VIX index but based on news volumes instead of on market activity. Two advantages of such a fear indicator would be its truly global nature and its flexibility when it comes to the exact design. Moreover, traders could attempt to use the news volume dynamics to predict movements in various financial asset prices. And risk managers could use the (changing) amount of news on word strings such as *stock market crash* in their risk management models. Policy makers, finally, could potentially use the aggregated amount of news on a particular topic, say inflation, in their policy decisions. And, of course, the usage is not limited to finance and economics. Psychology, sociology and other social sciences where news play a role could also profit from the availability of news aggregators.

ACKNOWLEDGEMENTS

Financial assistance from *Handelsbankens Forskningsstiftelser* is gratefully acknowledged.

ENDNOTES

[1] The homepage of Google News is http://news.google.com.

[2] Upon permission by the owner of the news aggregator, of course. In the case of Google News, the Terms of Use can be found on http://news.google.com.

[3] A small number of missing data points were replaced by interpolated values.
[4] The stock market data was provided by EcoWin AB (Reuters).
[5] Of course, the VIX index refers almost exclusively to US stocks. Here, however, we interpret the VIX index rather as a fear indicator than as a proxy for the world stock volatility.
[6] The two adjusted news volume series are normalized to start at the same level as the other two series in the graphs in Figures 3-4. This is only done for presentation purposes.
[7] The daily data used in this paper is overlapping in the sense that, each day, the Google News search result refers to the last thirty days and, each day, a new 30-day historical volatility is calculated. When non-overlapping data series (one observation every thirty days) are used the results remain more or less unchanged. The results are left out of this paper but are available from the author upon request.

REFERENCES

Berry, D. B. and K. M. Howe, “Public Information Arrival”, *Journal of Finance*, Vol. 49 No. 4 (1994), pp. 1331–1346.

Ederington, L. H. and J. H. Lee, “How Markets Process Information: News Releases and Volatility”, *Journal of Finance*, Vol. 48 No. 4 (1993), pp. 1161–1191.

Google Inc., “About Google News”, http://news.google.com, (2007)

Johnson, W. C. and J. Marietta-Westberg, "The Effect of News on Volatility: A Study of IPOs", Working Paper, Michigan State University (2004).

Mitchell, M. L. and J. H. Mulherin, “The Impact of Public Information on the Stock Market”, *Journal of Finance*, Vol. 49 No. 3 (1994), pp. 923–950.

Scholtes, S., “Market Gossip Goes High-Tech”, *Financial Times*, September 20, (2006).

van Duyn, A., “City Trusts Computers to Keep up With the News”, *Financial Times*, April 16, (2007).

Table 1 Sample correlations (Pearson correlation coefficients) between daily calculated stock market-related news volumes (number of news pieces in Google News that contain the word strings *stock market* or *stock market crash*) and daily calculated stock market volatilities (historical 30-day MSCI ACWI volatilities or historical 30-day averages of implied VIX volatilities). The time period is August 8, 2006 to October 23, 2007, and both adjusted and non-adjusted news volumes are included in the study.

	30-day Historical volatility	30-day average VIX volatility
stock market, non-adjusted	0.77	0.73
stock market, adjusted	0.62	0.51
stock market crash, non-adjusted	0.61	0.72
stock market crash, adjusted	0.37	0.48

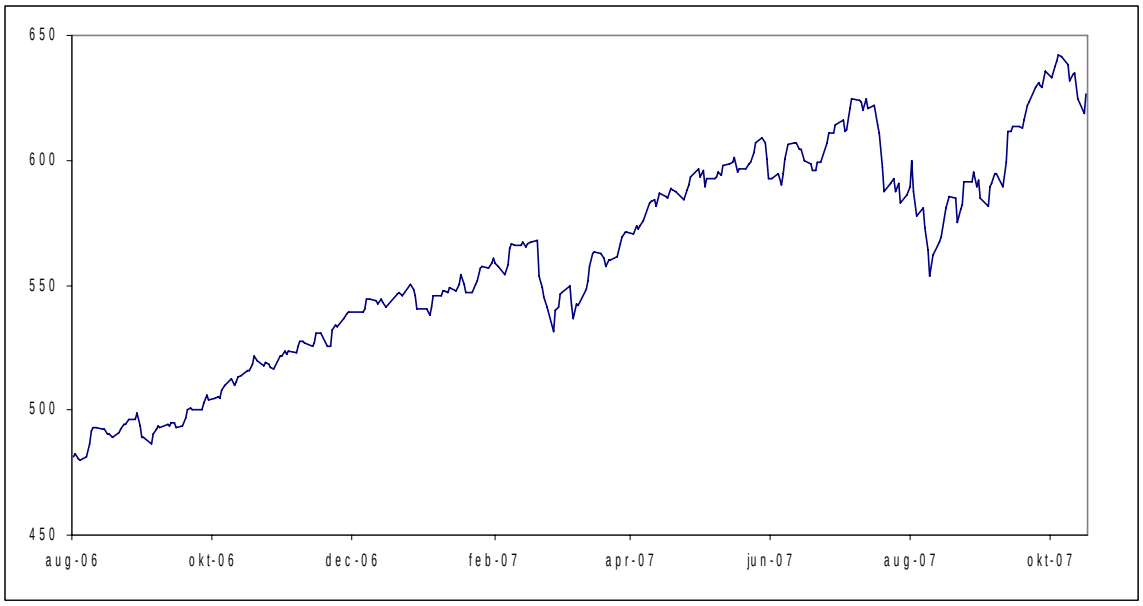

Figure 1
MSCI ACWI (All Country World Index). USD Index, Gross Total Return, Close, Daily Data.

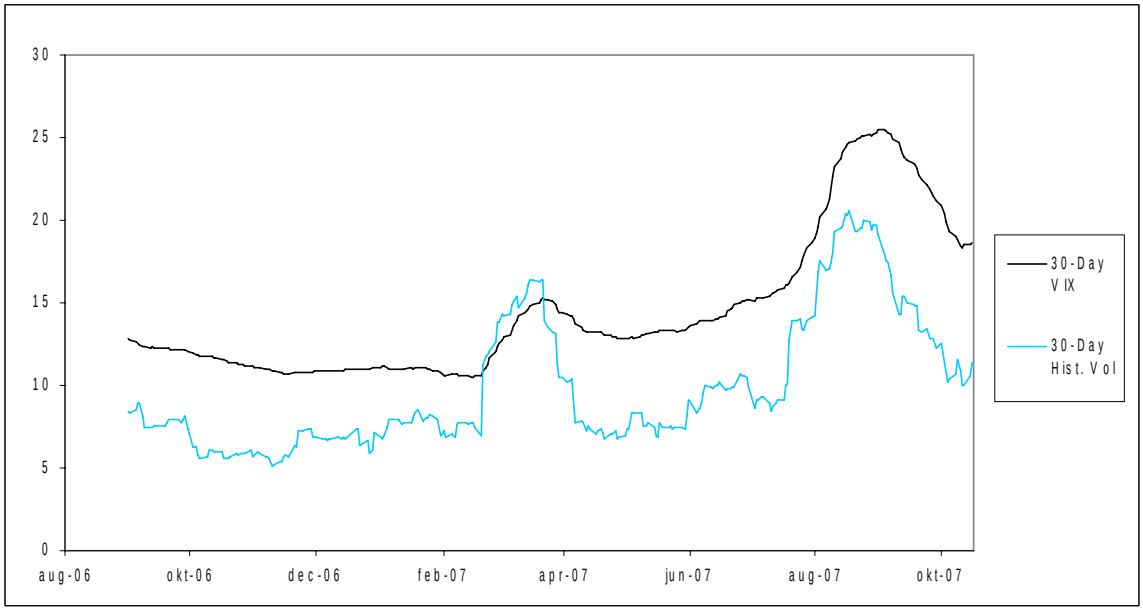

Figure 2
30-Day Historical Volatility of the MSCI ACWI Index and 30-Day Historical Average of the VIX Volatility-Index.

Figure 3
Google News Volume – Neutral News (*Stock Market,* adjusted /non-adjusted). The adjusted series is normalized to start at the same point as the non-adjusted series.

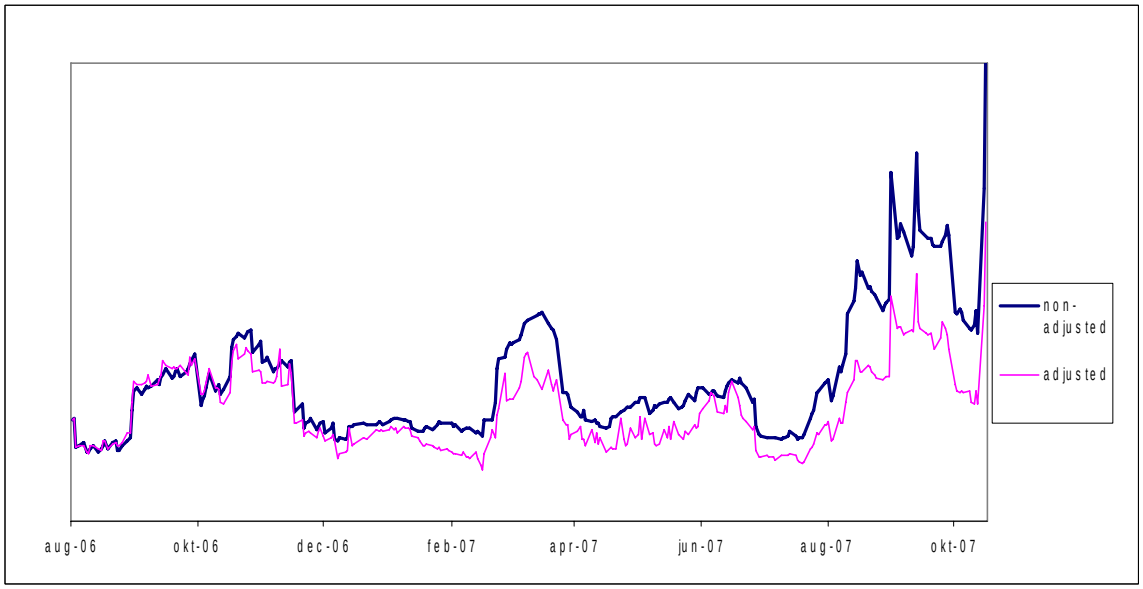

Figure 4
Google News Volume – Negative News (*Stock Market Crash,* adjusted /non-adjusted). The adjusted series is normalized to start at the same point as the non-adjusted series.

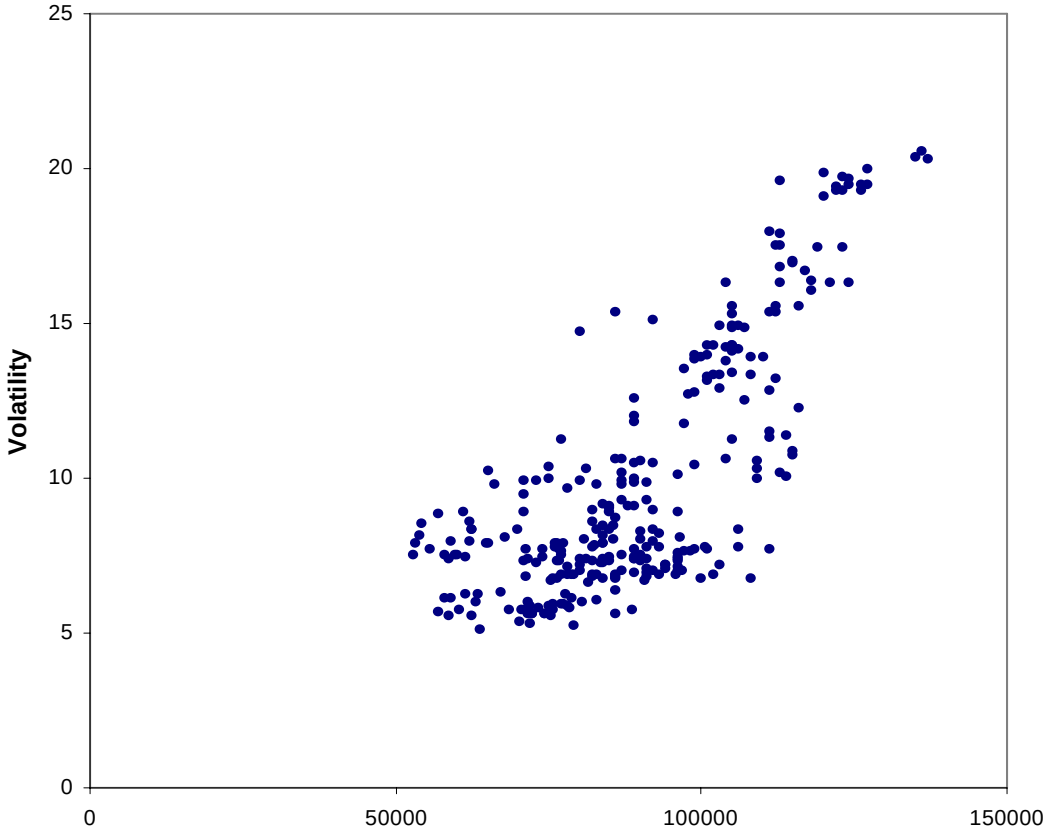

Figure 5
30-Day Historical Volatility of the MSCI ACWI Index versus Google News Volume – Neutral News (*Stock Market*, non-adjusted).

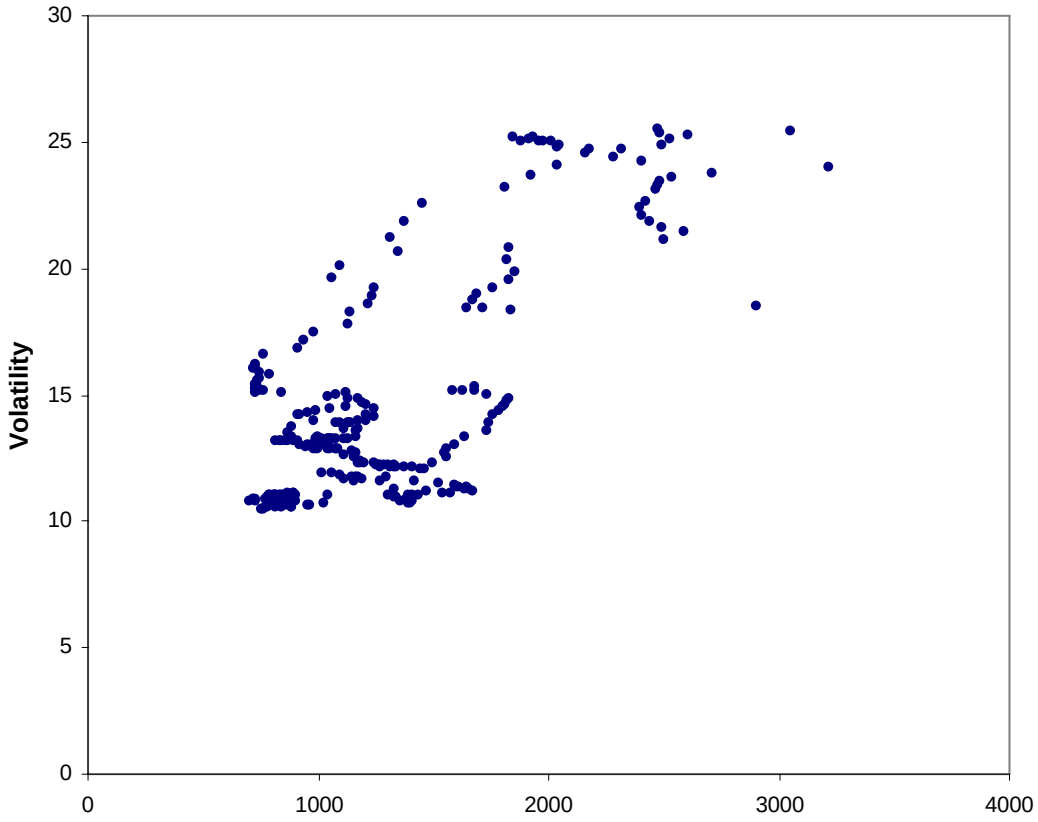

Figure 6
30-Day Historical Average of the VIX Volatility-Index versus Google News Volume – Negative News (*Stock Market Crash*, non-adjusted).