

Event-Based Historical Value-at-Risk

Frederik Hogenboom, Michael de Winter, Milan Jansen, Alexander Hogenboom,
Flavius Frasincaar, and Uzay Kaymak

Abstract—Value-at-Risk (VaR) is an important tool to assess portfolio risk. When calculating VaR based on historical stock return data, we hypothesize that this historical data is sensitive to outliers caused by news events in the sampled period. In this paper, we research whether the VaR accuracy can be improved by considering news events as additional input in the calculation. This involves processing the historical data in order to reflect the impact of news on the stock returns. Our experiments show that when an event occurs, removing the noise (that is caused by an event) from the measured stock prices for a small time window can improve VaR predictions.

I. INTRODUCTION

IN today's financial markets, Value-at-Risk (VaR) is a widely used risk measure quantifying the risk of loss on a portfolio of financial equities. For a given portfolio, confidence, and time horizon, VaR is defined as a threshold value such that the probability that the loss on the portfolio over the given time horizon does not exceed this value is at the given confidence level. Although VaR assumes normal market conditions, meaning that there are no sudden (unexpected) trend breaks, in the real world we do see derivations from trends, mainly caused by emerging news. For instance, when Google announced its plans to buy Motorola Mobility, Google shares went down around 1.5%, while Motorola Mobility's stock jumped 57% right after the news was published.

According to the weak form of the efficient market hypothesis, news that contains information on an equity is not perfectly incorporated in the price the moment it goes public. There are many studies that indicate such a delay exists in the incorporation of information in the price [1]. This delay is caused by an initial over- or under-reaction to the news. There have also been studies that indicate that news events have an effect on the volatility of equities [2]. Since VaR is based on predicting the distribution of returns, it is likely that taking into account news events for VaR calculations is beneficial, as the volatility is the standard deviation of this distribution.

Frederik Hogenboom, Alexander Hogenboom, and Flavius Frasincaar are with the Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands (email: {fhogenboom, hogenboom, frasincaar}@ese.eur.nl). Michael de Winter and Milan Jansen are with Semlab, Alphen a/d Rijn, The Netherlands (email: {dewinter, jansen}@sem lab.nl). Uzay Kaymak is with the Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands (email: u.kaymak@tue.nl).

The authors are partially sponsored by the NWO Physical Sciences Free Competition project 612.001.009: Financial Events Recognition in News for Algorithmic Trading (FERNAT) and the FES COMMIT Infiti project Information Retrieval for Information Services. We would like to thank Mark Vreijling and Wijand Nuij from Semlab for fruitful discussions and technical support during this research.

There are multiple methods to compute VaR, but ultimately it comes down to predicting the distribution of future returns. The three most widely used implementations are the parametric method (where a normal or log-normal distribution of equity returns is usually assumed), a Monte Carlo simulation-based method that predicts future returns by fitting a distribution based on historical data, and the historical method, where the assumption is that historical changes in the price accurately predict changes in the future.

The parametric method requires an assumption about the distribution of the returns of an equity. Popular assumptions are that the distribution of returns can be captured by a normal or a log-normal distribution, as these distributions offer simplicity and robustness. However, in practice, equity returns are almost never normally distributed. Assuming a specific distribution could therefore lead to a bias in the risk measure, and hence we do not use this method as the basis of our research.

Alternatively, the Monte Carlo simulation randomly samples the historical data multiple times to approximate its distribution. This method is complex to implement and its operations are computationally intense. As we aim for an application that is able to run real-time, Monte Carlo simulation-based methods are not suitable.

In contrast to the previous method, the historical method analyzes a set of historical returns instead of an assumed distribution. It therefore assumes that changes in the past are predictive for the future. This is only true if the portfolio consists of independent equities because using heteroscedastic data¹ will result in a biased standard error and as such, any conclusion we would reach based on these biased standard errors is not reliable. In order to make sure our portfolio has no heteroscedasticity we evaluate only single equity portfolios. An advantage of the historical method is its simplicity, which fosters real-time computation, and therefore, in this paper we utilize the historical method.

VaR has a few known limitations [3]–[5], as it is only meaningful on losses that are normally distributed, which in practice is often not the case. Additionally, optimization could be difficult due to the existence of local optima, and also VaR outcomes are usually biased towards optimism instead of conservatism [6]. Despite its shortcomings, VaR is still widely adopted by practitioners and has become a standard in the finance industry, while other approaches with better properties such as the expected shortfall or conditional value at risk (CVaR) [7] – which measures the market risk

¹A sequence of random variables is heteroscedastic if the random variables have interdependent variances, as is often the case with different financial equities in a portfolio.

of a portfolio and is more sensitive to the tail of the loss distribution than regular VaR – have not become a standard.

Therefore, in this paper we focus on the improvement of VaR due to its applicability in practice. We hypothesize that we can improve VaR computations (using the historical method) by introducing financial news events [8], [9] as an additional input. Using information extracted from text in a financial context has proven to be a vital strategy in many financial applications [10]–[16]. In our approach, we clean the historical data set for data that is subject to a relatively high news influence, and aim to obtain a data set which is a better, more stable representation of the expected returns distribution. With this improved distribution we try to calculate a more accurate VaR.

In our research, we employ ViewerPro [17] for the extraction of ticker data and news events. For ticker data – to be used for VaR calculation – the platform extracts data from financial feeds, e.g., Bloomberg or Google Stocks. For news events, the ViewerPro application selects, structures, and presents news stories for trading purposes by scanning news wire sources – e.g., Reuters or AP Dow Jones – for predefined key words or phrases related to financial events. Identified events are subsequently linked to companies. For example, a message such as “*Unilever announced a 14% rise in profit*”, would cause ViewerPro to recognize an event in the format of “*Company Profit Rise*”, and link it to the company “*Unilever*”. Such identified events could potentially be useful in trading applications, such as the prediction of risk.

We assume news events are unique, i.e., they do not repeat in the future. For this reason, news events and their effects on the market are a distorting factor in the historical data set used in traditional methods of calculating VaR. This influence of (i.e., the overreaction to) an arbitrary news event is maximized immediately after the news is released and diminishes in the following hours as the market interprets and reacts to the new information. Eventually, the market returns to its normal state when the effect of the news message goes down to zero.

This paper is organized as follows. First we describe related approaches to this research in Sect. II. Then we introduce our framework in Sect. III. Section IV presents our implementation, our data set, and an evaluation of the framework. Last, in Sect. V we draw our conclusions and provide directions for future work.

II. RELATED WORK

In the existing body of literature, various researchers have underlined the relationship between news events and the stock market [18]–[21]. Furthermore, a correlation has been found between the number of news events and trading activity [2]. Both the efficient market hypothesis and the random walk theory support that news information is fully and immediately processed into the value of the share. In practice though, evidence indicates this is not always the case [22], [23]. Hence, for traders, timely and accurately reacting on news before the competition does, yields profitable trades.

A viable strategy could be the use of algorithmic trading, where machines trade automatically at high speeds based on an array of inputs.

Hull and White [24] improve the VaR calculation by updating the volatility in the historical method by means of GARCH/EWMA models in order to reflect the difference between the volatility at the time of the observation and the current volatility. The difference between their research and ours is that the authors of [24] analyze multiple equity portfolios, whereas we only use a single equity. The authors propose a method to update the volatility in the appropriate time interval so that the volatility becomes a more dynamic factor in VaR calculation, leading to a more accurate VaR prediction. The method is compared to another method, which involves assigning weights to observations that are more recent, so that they get sampled more frequently [25]. The evaluation of these two methods and the traditional historical method is performed based on mean absolute percentage error (MAPE). The authors find that the first method outperforms both the traditional historical method and second method for exchange rates. However, this method’s results for stock indices are mixed.

The authors of [26] aim to improve technical indicators with news as well. They use only a simple text classification algorithm with a supervising learning method. Instead of only using company specific news, the authors are also integrating general market news in combination with technical indicators. The authors arrive at the interesting conclusion that technical indicators and news events alone are inaccurate as estimators, but that the combination of both could lead to better results. Based on a real life market simulation the authors show that by using their approach it is possible to make profit.

III. FRAMEWORK

In order to be able to assess whether the incorporation of news into the calculation of the VaR of a specific equity improves the overall quality of the outcomes, we propose a framework that is based on two inputs: a list of stock prices and a list of financial events, which in our case stem from the ViewerPro application. Figure 1 depicts the processing steps of our framework, which are discussed in more detail in the following subsections.

A. Data Processing

It should be noted that in order to enable optimal processing, we first need a proper data set containing financial data and news messages with event annotations. Hence, the first step in our algorithm is to clean collected equity prices stemming from the ViewerPro application. As stock markets are only open on specific dates and times, we filter the prices and keep those recorded within market opening times. Also, the time intervals between individual prices are increased so that the hourly prices are kept, because a more fine-grained data set would tremendously increase the computational complexity.

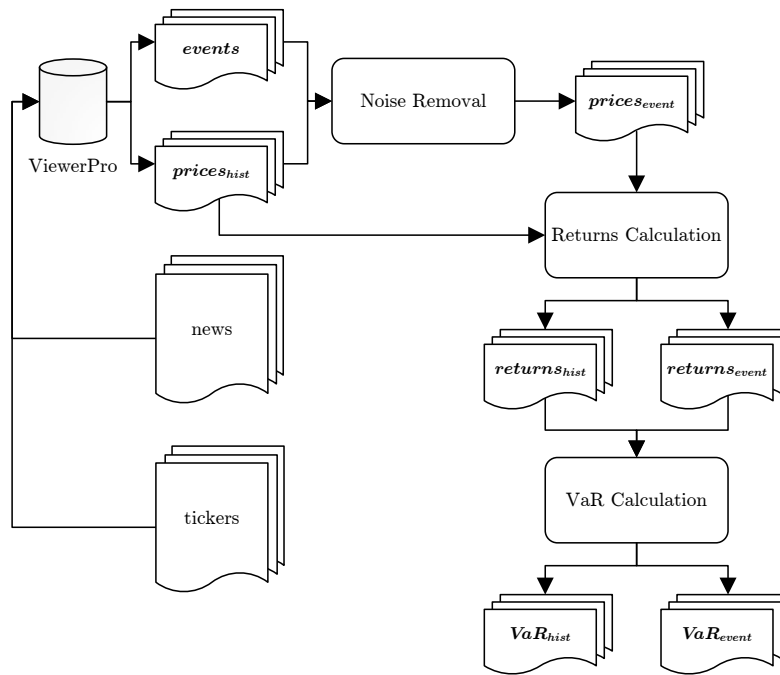


Fig. 1. Overview of data flows and processing steps within the event-based VaR framework

Subsequently, our second step is to read in the annotated news events. ViewerPro receives (financial) news feeds from disparate sources, e.g., RSS feeds and any other textual news sources. These streams are processed by means of computational linguistics, semantic analysis, and formal logic. This way, ViewerPro determines the positive and negative impacts of the information described in the news on the equities that are relevant to the user. Once the information is fed into the ViewerPro system, it undergoes several steps in order to filter out unwanted information and select the information that is relevant for our research. Large amounts of news messages are filtered for equity-specific news, and the semantic analysis system of ViewerPro analyzes each individual news message for economic impact. This yields a list of relevant annotated news events. Some general types of news events that are covered by the ViewerPro annotations are described in Table I.

B. Removing Noise

The identified events are associated with times in which they occurred. These event times are matched to the times of the recorded equity prices. Our assumption is that whenever

a news event occurs, the prices of a subsequent time window are directly influenced by this event. As the influenced prices are not reliable as predictors for future prices (due to the temporal imbalance they cause), we adjust the collected prices for a fixed time window to account for the generated noise by updating their values to the previously measured value.

Algorithm 1 illustrates our update mechanism, which processes a list of chronologically ordered (hourly) recorded prices. For each stock price $price$ in price list $prices$, we check whether there has been an occurrence of an event on the time the price was measured by comparing the stock price time with the time of each event $event$ (rounded down

Algorithm 1 News event processing (per equity)

Require: $prices$ = array of stock prices and associated times

Require: $events$ = array of events and associated times

Require: $window$ = integer representing time window

```

1:  $previousprice.value = prices.first.value$ 
2: for all  $price$  in  $prices$  do
3:   for all  $event$  in  $events$  do
4:     if  $impact > 0$  then
5:        $impact = impact - 1$ 
6:        $price.value = previousprice.value$ 
7:     end if
8:   if  $price.time = event.time$  then
9:      $impact = window$ 
10:  end if
11: end for
12:  $previousprice.value = price.value$ 
13: end for
  
```

TABLE I

SOME GENERAL NEWS EVENTS COVERED BY VIEWERPRO ANNOTATIONS

Type	Sub-types
CEO	hiring, resignation
Acquisition	consideration, start, completion, stop
Bid	receiving, consideration, acceptance, drop, raise
Profit	down, up
Legal conflict	loss, resolution, win
Bankruptcy	–

on an hourly basis) stored in event list *events*. If an event occurrence is identified, *impact* is set to the window size *window* (which is optimized to 8 after a series of initial experiments using a hill-climbing procedure with values ranging from 1 to 100), causing the value of the subsequent *price* items to be set to the current value. The value of *impact* is decreased with 1 every next *price* in price list *prices*, so that subsequent price values are updated up until the window size has been reached. In case of overlapping events, naturally the *impact* counter is reset to the window size *window*.

After all prices have been processed, the denoisation process returns a new list of prices that have been adapted in case an event occurred in their vicinity: *prices_{event}*. Together with the original prices *prices_{hist}*, these prices are fed into the next processing step of the framework.

C. From Prices to Returns

Both sets of original (“*hist*”) and denoised (“*event*”) prices are converted to sets containing hourly returns. We compute the return set *returns* of a price set *prices* as

$$returns = \frac{prices_{t+1} - prices_t}{prices_t} \quad \forall t = 1, \dots, N - 1, \quad (1)$$

where N represents the number of items in the list. Hence, the return is calculated as the relative change between the price at time $t+1$ and the previous price at time t . A specific return $returns_t$ is to be interpreted as the profit that can be obtained if a share is bought at time t and sold at time $t+1$.

D. Calculating Value-at-Risk

As we implement the historical method of VaR calculation, we do not assume a distribution, but use the historical returns (both original and adapted) to estimate the future returns. The time horizon here is considered to be 1 day. After sorting the return list *returns*, we calculate the Value-at-Risk, *VaR*, as

$$VaR = returns' [\lceil \alpha \cdot \text{length}(returns) \rceil], \quad (2)$$

where *returns'* represents the ordered (sorted) list of returns and where the confidence level is denoted by α . For example, if the data set contains 100 historical returns – with the first element being located on position 1, and the last on position 100 – we select the fifth worst return (i.e., position 95) when our confidence level is 95%.

A more detailed example is depicted in Table II. Here, the results of a VaR calculation are presented, based on 21 prices – with and without noise removal – with an *event* occurring at $t = 6$ while using a *window* size of 8. With a confidence interval of 95%, this would result in a VaR of -0.39 or -0.10 (printed in bold font) for daily returns for the historical method or the updated historical method proposed here, respectively. A VaR of -0.39 or -0.10 with confidence of 95% means that it is not expected to make losses lower than 35% or 10%, respectively. The observed difference stems from the proposed removal of noise inherently associated with events, i.e., the noise in prices generated at time $t = 6 + 1$ up until time $t = 6 + window$. These differences can

then subsequently be evaluated by assessing the quality of both predicted values using common measures such as the mean squared error (MSE), where the error is defined as the amount by which the calculated VaR differs from the actual worst return (at the same confidence level) that is measured in the future.

IV. EVALUATION

For evaluation of the performance of the proposed VaR calculation that involves an additional denoisation process for historical data in case of event occurrences, the framework elaborated on in Sect. III is implemented as a Java-based application that calculates the VaR of a single equity based on a data set containing news events and stock prices. Note that we are not analyzing a portfolio of equities in order to rule out any bias caused by heteroscedasticity.

A. Data Set

The data set that is used in our conducted experiments covers news events and stock data collected on an hourly basis for 363 equities – differing in volatility and amount of news coverage received – during the year 2010. Prices that were collected during weekends and after market closing hours are filtered out, since no price changes occur during these hours, leaving us with approximately 2,000 stock data points and 50 up to 75 associated events per equity. In order to evaluate the performance of the calculation, we predict the VaR with both our adjusted method (referred to as *VaR_{event}*) and the traditional method (referred to as *VaR_{hist}*) for 75% of our data set. The remaining 25% is used as a test set for comparing the predicted VaR with the actual VaR. Traditionally, in VaR research, models are benchmarked against GARCH models. However, the main focus of the paper is on the improvement of a specific VaR calculation method, i.e., the historical method. Therefore, we do not take into consideration GARCH models, as the characteristics and underlying assumptions are too different.

B. Metrics

Even though many VaR analyses are currently performed using the Kupiec test [27], we employ a different set of measures. As explained by Kupiec in his original work, the test is hampered by the fact that it is statistically weak with sample sizes consistent with the current regulatory framework (one year). As our data comprises the year 2010, it would not be a good idea to perform Kupiec’s test. Therefore, we employ different measures that provide insights into the effectiveness of our proposed event-based approach.

In order to analyze for how many equities our adjusted event-based historical method provides better quality predictions in comparison to the traditional historical method, we measure each method’s squared error: the smaller the value (i.e., the closer to zero), the better. The squared error *SE* for equity e is defined as the squared difference between the equity’s actual VaR (*VaR_{e,actual}*) measured in our test

TABLE II

EXAMPLE CALCULATION FOR RETURNS, BASED ON 21 PRICES WITH AND WITHOUT NOISE REMOVAL (*event* AND *hist* COLUMNS, RESPECTIVELY) WITH AN *event* OCCURRING AT $t = 6$ WITH A *window* SIZE OF 8; TOGETHER WITH A CONFIDENCE INTERVAL OF 95%, THIS WOULD RESULT IN A PREDICTED VALUE-AT-RISK OF EITHER -0.39 (*hist*) OR -0.10 (*event*), PRINTED IN BOLD FONT, WHICH ARE DISTINCT PREDICTED VALUES CAUSED BY THE PROPOSED REMOVAL OF NOISE INHERENTLY ASSOCIATED WITH EVENTS

Prices			Returns			Returns (sorted)	
t	<i>hist</i>	<i>event</i>	t	<i>hist</i>	<i>event</i>	<i>hist</i>	<i>event</i>
1	0.35	0.35	1	0.14	0.14	2.50	0.44
2	0.40	0.40	2	0.07	0.07	0.48	0.14
3	0.43	0.43	3	-0.37	-0.37	0.44	0.08
4	0.27	0.27	4	0.44	0.44	0.36	0.07
5	0.39	0.39	5	-0.03	-0.03	0.14	0.05
6	0.38	0.38	6	-0.68	0.00	0.14	0.05
7	0.12	0.38	7	2.50	0.00	0.07	0.00
8	0.42	0.38	8	0.36	0.00	0.05	0.00
9	0.57	0.38	9	-0.39	0.00	0.05	0.00
10	0.35	0.38	10	-0.23	0.00	-0.02	0.00
11	0.27	0.38	11	0.48	0.00	-0.03	0.00
12	0.40	0.38	12	-0.08	0.00	-0.03	0.00
13	0.37	0.38	13	-0.03	0.00	-0.03	0.00
14	0.36	0.38	14	0.14	0.08	-0.08	0.00
15	0.41	0.41	15	0.05	0.05	-0.08	-0.02
16	0.43	0.43	16	-0.02	-0.02	-0.10	-0.03
17	0.42	0.42	17	-0.10	-0.10	-0.23	-0.03
18	0.38	0.38	18	0.05	0.05	-0.37	-0.08
19	0.40	0.40	19	-0.08	-0.08	-0.39	-0.10
20	0.37	0.37	20	-0.03	-0.03	-0.68	-0.37
21	0.36	0.36					

set and the predicted VaR ($VaR_{e,predicted}$) that has been predicted based on our training set, i.e.,

$$SE_e = (VaR_{e,actual} - VaR_{e,predicted})^2. \quad (3)$$

Please note that $VaR_{e,predicted}$ is one of VaR_{event} or VaR_{hist} .

For both methods, the squared errors for all equities of a set are subsequently combined into the mean squared error (MSE), yielding an MSE_{hist} and MSE_{event} . In general, the MSE is calculated as the summation of the squared errors (SE) of all equities $e \in E$ divided by the number of equities, i.e.,

$$MSE = \frac{\sum_{e \in E} SE_e}{|E|}, \quad (4)$$

with $|E|$ being the total number of equities in set E , i.e., 363.

Additionally, we perform a two-sample one-tailed t -test on the sets of individual squared errors SE_{hist} and SE_{event} (containing $SE_{e,hist}$ and $SE_{e,event} \forall e \in E$, respectively) in order to assess the significance of the measured difference between MSE_{hist} and MSE_{event} . For this, we use a significance level of 0.05 to reject the null hypothesis that there is no difference between the measured MSE values.

Finally, we evaluate the number of times the both methods outperform one another, OPT (OutPerformed Total). This

is done by comparing the squared errors $SE_{e,hist}$ and $SE_{e,event}$ for each equity $e \in E$, yielding

$$OPT_{hist,event} = \sum_{e \in E} OP(SE_{e,hist}, SE_{e,event}), \quad (5)$$

$$OPT_{event,hist} = \sum_{e \in E} OP(SE_{e,event}, SE_{e,hist}), \quad (6)$$

$$OP(X, Y) = \begin{cases} 1 & \text{if } X < Y \\ 0 & \text{else} \end{cases}. \quad (7)$$

It should be noted that $OPT_{event,hist}$ cannot be equal to $|E| - OPT_{hist,event}$, nor can $OPT_{hist,event}$ be equal to $|E| - OPT_{event,hist}$, as both methods could perform equally on some events, and hence the condition $OPT_{event,hist} + OPT_{event,hist} = |E|$ does not always hold.

C. Experimental Results

When comparing the results from both VaR calculation methods, we obtain the results depicted in Table III, which shows the traditional and event-based historical VaR calculation (columns *hist* and *event*, respectively). The last column, $\Delta\%$, contains the percentage of improvement for both the mean squared error (MSE) and the number of times an outperformance is measured (OPT).

Based on these results, we observe that on our data set, our adjusted (event-based) historical method made predictions that were of higher quality than the ones of the traditional

TABLE III

EXPERIMENTAL RESULTS OF THE PERFORMANCE OF TRADITIONAL AND
EVENT-BASED HISTORICAL VaR CALCULATION

Measure	<i>hist</i>	<i>event</i>	$\Delta\%$
<i>MSE</i>	1.0590E-05	8.2965E-06	21.66%
<i>OPT</i>	89	234	162.92%

historical method. The mean difference measured between the predicted and actual VaR was merely 8.2965E-06, which is an improvement of 21.66% compared to 1.0590E-05. Also, our event-based method outperformed the traditional method in 234 out of the 363 cases (i.e., 64.46%), which is an improvement of 162.92% compared to the score of the traditional method, which only outperformed 89 times (i.e., 24.52%). Note that in our data set, when comparing the traditional and event-based historical methods, there are cases of equal performance of both methods, and hence $OPT_{hist,event} + OPT_{event,hist} \neq |E|$.

In order to assess significance of the measured *MSE* improvement of 21.66% when comparing the traditional historical method with our event-based method, we perform a paired two-sample one-tailed *t*-test based on SE_{hist} and SE_{event} , containing squared errors for all equities, using the hypotheses

$$H_0 : MSE_{event} = MSE_{hist} ,$$

$$H_1 : MSE_{event} > MSE_{hist} ,$$

and ($|E| - 1 = 363 - 1 =$) 362 degrees of freedom. From our two sample *t*-test we obtain a *p*-value of 9.3529E-06. When applying a significance level of 0.05, we can reject the null hypothesis that there is no difference between the measured *MSE* values, and hence the measured improvement is significant, i.e., the proposed event-based historical VaR calculation method produces more reliable VaR predictions when compared to the traditional method that does not take into account events.

V. CONCLUSIONS

In this paper we have proposed a way to enhance the calculation and prediction of Value-at-Risk (VaR) based on historical data, by introducing a denoisation process linked to news events causing noise in stock prices through over- and underreactions. We have presented a framework that takes event and stock price data stemming from the proprietary ViewerPro software for predicting VaR using historical data, both in the traditional way and in the event-based way. Our experiments on a substantial data set with an implementation of the proposed framework underline that our enhanced VaR calculation outperforms the traditional approach (in 64.46% of the cases), causing the mean squared errors to drop significantly with 21.66%, making thus our enhanced VaR predictions more reliable. Therefore, the calculation and prediction of VaR can be improved with news (or more specifically, extracted events and stock rates) as an additional input.

For future work, we suggest to investigate accounting for the type of news events. This could be related to the influence a news event has on the price of an equity. For example, a CEO resignation might have a bigger impact on prices than an acquisition offer. Another direction for future research is related to additionally accounting for general stock market events, instead of only just the company specific news. Moreover, we would also like to build a real life market simulation for our improved historical VaR method. Finally, it would be interesting to expand the research presented in this paper to other financial risk measurements, such as the conditional Value-at-Risk (CVaR).

REFERENCES

- [1] E. F. Fama, "The Behavior of Stock-Market Prices," *Journal of Business*, vol. 38, no. 1, pp. 34-105, 1965.
- [2] M. L. Mitchell and J. H. Mulherin, "The Impact of Public Information on the Stock Market," *Journal of Finance*, vol. 49, no. 3, pp. 923-950, 1994.
- [3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Thinking Coherently," *Journal of Risk*, vol. 10, no. 11, pp. 68-71, 1997.
- [4] —, "Coherent Measures of Risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203-228, 1999.
- [5] R. McKay and T. E. Keefer, "VaR is a Dangerous Technique," *Corporate Finance, Searching for Systems Integration Supplement*, p. 30, 1996.
- [6] R. T. Rockafellar and S. Uryasev, "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1443-1471, 2002.
- [7] —, "Optimization of Conditional Value-at-Risk," *Journal of Risk*, vol. 2, no. 3, pp. 21-41, 2000.
- [8] J. Borsje, F. Hogenboom, and F. Frasinca, "Semi-Automatic Financial Events Discovery Based on Lexico-Semantic Patterns," *International Journal of Web Engineering and Technology*, vol. 6, no. 2, pp. 115-140, 2010.
- [9] A. Hogenboom, F. Hogenboom, F. Frasinca, U. Kaymak, O. van der Meer, and K. Schouten, "Detecting Economic Events Using a Semantics-Based Pipeline," in *Twenty-Second International Conference on Database and Expert Systems Applications (DEXA 2011)*, ser. Lecture Notes in Computer Science, A. Hameurlain, S. W. Liddle, K.-D. Schewe, and X. Zhou, Eds., vol. 6860. Springer, 2011, pp. 440-447.
- [10] D. L. Ikenberry and S. Ramnath, "Underreaction to Self-Selected News Events: The Case of Stock Splits," *Review of Financial Studies*, vol. 15, no. 2, pp. 489-526, 2002.
- [11] J. van Bommel, "Rumors," *Journal of Finance*, vol. 58, no. 4, pp. 1499-1520, 2003.
- [12] W. S. Chan, "Stock Price Reaction to News and No-News: Drift and Reversal After Headlines," *Journal of Financial Economics*, vol. 70, no. 2, pp. 223-260, 2003.
- [13] M.-A. Mittermayer and G. F. Knolmayer, "Text Mining Systems for Market Response to News: A Survey," Institute of Information Systems, University of Bern, Working Paper 184, 2006, <http://www2.ie.iwi.unibe.ch/publikationen/berichte/resource/WP-184.pdf>.
- [14] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science*, vol. 53, no. 9, pp. 1375-1388, 2007.
- [15] P. C. Tetlock, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, vol. 62, no. 3, pp. 1139-1168, 2007.
- [16] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More Than Words: Quantifying Language to Measure Firms' Fundamentals," *Journal of Finance*, vol. 63, no. 3, pp. 1437-1467, 2008.
- [17] Semlab, "ViewerPro," <http://viewerpro.semlab.nl/>, 2011.
- [18] H. Byström, "News Aggregators, Volatility and the Stock Market," *Economics Bulletin*, vol. 29, no. 4, pp. 2673-2682, 2009.
- [19] J. E. Engelberg and C. A. Parsons, "The Causal Impact of Media in Financial Markets," *Journal of Finance*, vol. 66, no. 1, pp. 67-97, 2009.

- [20] R. Goonatilake and S. Herath, "The Volatility of the Stock Market and News," *International Research Journal of Finance and Economics*, vol. 3, no. 11, pp. 53–65, 2007.
- [21] P. S. Kalev, W.-M. Liu, P. K. Pham, and E. Jarnecic, "Public Information Arrival and Volatility of Intraday Stock Returns," *Journal of Banking & Finance*, vol. 28, no. 6, pp. 1441–1467, 2004.
- [22] B. Rosenberg, K. Reid, and R. Lanstein, "Persuasive Evidence of Market Inefficiency," *Journal of Portfolio Management*, vol. 11, no. 3, pp. 9–16, 1985.
- [23] S. Basu, "Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis," *Journal of Finance*, vol. 32, no. 3, pp. 663–682, 1977.
- [24] J. Hull and A. White, "Incorporating Volatility Updating into the Historical Simulation Method for Value-at-Risk," *Journal of Risk*, vol. 1, no. 1, pp. 5–19, 1998.
- [25] J. Boudoukh, M. Richardson, and R. F. Whitelaw, "The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk," *Risk*, vol. 11, no. 5, pp. 64–67, 1998.
- [26] Y. Zhai, A. Hsu, and S. K. Halgamuge, "Combining News and Technical Indicators in Daily Stock Price Trends Prediction," in *4th International Symposium on Neural Networks (ISSN 2007)*, D. Liu, S. Fei, Z. Hou, H. Zhang, and C. Sun, Eds. Springer-Verlag Berlin, Heidelberg, 2007, pp. 1087–1096.
- [27] P. H. Kupiec, "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, vol. 3, no. 2, pp. 73–84, 1995.