Research Report

Forecasting News-Related Liquidity Shocks: Extracting Signals from Unstructured Data

TO DERIVE OPTIMAL ORDER EXECUTION STRATEGIES THAT STRIVE TO MINIMIZE TRANS-ACTION COSTS, INVESTORS AS WELL AS AUTOMATED TRADING ENGINES MUST BE ABLE TO ANTICIPATE CHANGES IN THE AVAILABLE MARKET LIQUIDITY. BASED ON AN EVENT STUDY ON THE LIQUIDITY IMPACT OF AD-HOC DISCLOSURES, WE PROPOSE A NOVEL IT ARTIFACT THAT ALLOWS AUTOMATED TRADING ENGINES TO APPROPRIATELY REACT TO NEWS-RELATED LIQUIDITY SHOCKS. FURTHERMORE, WE PROVIDE A SIMULATION-BASED EVALUATION THAT SHOWS ITS ECONOMIC RELEVANCE.

Sven S. Groth   Michael Siering

Peter Gomber

Introduction

Financial markets are characterized by high levels of complexity and non-linearity, whereas information systems play a crucial role in supporting human and computer-based decision-makers alike. Due to the far-reaching electronification of the securities trading value chain, computer-based automated traders (often referred to as "algortihmic trading engines") generate a large amount of the trading activity on major European markets.

They "emulate a broker's core competence of slicing a large order into a multiplicity of smaller orders and of timing these to minimize the market impact" (Gomber and Gsell, 2006). The decision on the investment or portfolio allocation itself is performed by the respective portfolio manager at a fund management company, and the primary task of automated traders is to execute the orders that are received from these fund management companies or institutional investors at the best available conditions.

To determine an optimal execution strategy for a pre-defined execution time period, i.e., to achieve the best available conditions, automated traders must handle the trade-off between the different transaction cost components in the order's execution. First, there are explicit costs, such as commissions, fees, or taxes. Second, there are implicit costs, such as market impact, timing costs, or opportunity costs (Bikker et al., 2006). Especially for large trades, implicit transaction costs mostly are much larger than explicit transaction costs. Liquidity constitutes the main determinant of implicit transaction costs: if the number of shares that other market participants are willing to trade at a given limit is reduced, then the market impact of an order is increased. The higher the liquidity, the lower the implicit transaction costs, and vice versa (Schwartz and Francioni, 2004). Therefore, to derive an optimal execution strategy and to minimize the associated transaction costs, the ability to forecast future liquidity levels is very important.

Bikker et al. (2006), however, conclude that "forecasting market impact costs appears notoriously difficult and traditional methods fail". Domowitz and Yegerman (2005) find that the execution quality of automated traders is inferior to the executions that are handled by brokers. One reason for this observation might be the fact that the employed models are solely based on purely quantitative input data and largely neglect one of the most important sources of information, which is unstructured qualitative data. If, for example, a listed com-

pany issues an unanticipated regulatory ad-hoc disclosure, then automated traders cannot react sufficiently fast because they cannot analyze the content of that disclosure. Because unanticipated news, by definition, is very unlikely to be reflected in quantitative time series data prior to its publication date, automated traders can respond only to other (human) market participants' reactions.

Against this background, we investigate whether and how unstructured qualitative data can be used as input for automated trading engines. We are especially interested in whether useful information can be extracted automatically from qualitative data to predict future levels of liquidity after the publication of corporate disclosures. For extracting such information, data mining techniques are utilized.

Research Approach

Data mining aims at discovering useful patterns in data that can serve for predictions. However, for a successful application of machine learning techniques, additional steps such as domain and data understanding as well as data reduction and pre-processing are crucial, too (Han and Kamber, 2006).

In our study (Groth et al., 2014), we acquire a dataset that is composed of ad-hoc disclosures, stock prices, and order book data. We perform an event study on the liquidity impact of ad-hoc disclosures. Support Vector Machine as an appropriate data mining algorithm for classification is selected. Finally, the results of

the data mining phase are analyzed by means of a simulation-based evaluation.

Dataset Analyzed

The news dataset analyzed is composed of ad-hoc disclosures that are published by Deutsche Gesellschaft für Ad-hoc-Publizität (DGAP) on behalf of the companies admitted to trading on an organized market in Germany. To fulfil the legal requirements, these companies must publish immediately any insider information or other information that is highly relevant to investors. We concentrate on this news type because the disclosures are expected to primarily contain new information and event studies have shown that these are often followed by abnormal stock returns (Muntermann and Guettler, 2007). The final dataset comprises 415 ad-hoc disclosures. Furthermore, in order to be able to determine the liquidity impact, the dataset at hand contains high-frequency (level-2) order book data, i.e., it allows insights into the order book breadth and depth, including the best ten bid and offer limits and the respective order quantities at those limits. The order book data were extracted from Thomson Reuters Tick History.

Liquidity Impact of Ad-hoc Disclosures

By means of an intraday event study, we determine the impact of ad-hoc disclosures on stock liquidity. First, it can be observed that there are no significant abnormal liquidity levels prior to the publication of corporate disclosures. This finding provides evidence that the chosen event type actually contains new and previously unknown information that has not been widely anticipated by market participants. It is therefore expected that forecasting of such liquidity shocks is especially challenging for models that are based solely on historical quantitative data.

Second, we find strong empirical evidence that transaction costs increase subsequent to the publication of corporate disclosures. This finding is most likely due to the fact that the disclosures' contents persuade traders to adjust their valuations of the respective company and adjust their existing limit orders in the order book accordingly. During the adjustment process, fewer limit orders (or limit orders with a lower volume) remain in the market, and therefore, the cost of execution increases (i.e., the liquidity decreases).

Given these results, one might typically assume lower liquidity levels (i.e., higher implicit transaction costs) subsequent to the publication of corporate disclosures. Therefore, the simplest strategy to avoid high implicit transaction costs would be to either execute orders immediately at the disclosure (and thereby prior to the negative liquidity impact) or wait for execution until the liquidity reverts to normal levels. The latter case would, however, incur timing costs.

If we, however, take a closer look at the distribution of the abnormal levels of liquidity, we note that certain corporate disclosures are followed by better levels of liquidity (about 25% of all cases). We are especially interested in identifying those corporate disclosures that are associated with implicit transaction costs that are below historical levels and, consequently, liquidity levels that are above historical levels.

As a result, forecasting this liquidity impact by means of data mining methodologies appears to be promising because investors as well as automated trading engines must adjust their strategies according to the expected liquidity impact to reduce the implicit transaction costs. Therefore, we use the ad-hoc disclosures for training a Support Vector Machine classifier that is able to distinguish between ad-hoc disclosures followed by higher or lower levels of liquidity.

Simulation-Based Evaluation

To evaluate the Support Vector Machine classifier, we introduce a novel domain-specific simulation-based evaluation approach that aims at acting on the discovered knowledge. In general, a simulation-based model evaluation allows for additional statistical analysis and provides insights into the results' robustness. Previous applications of text mining techniques in the financial industry have highlighted the need for domain-specific evaluation metrics (Groth and Muntermann, 2011).

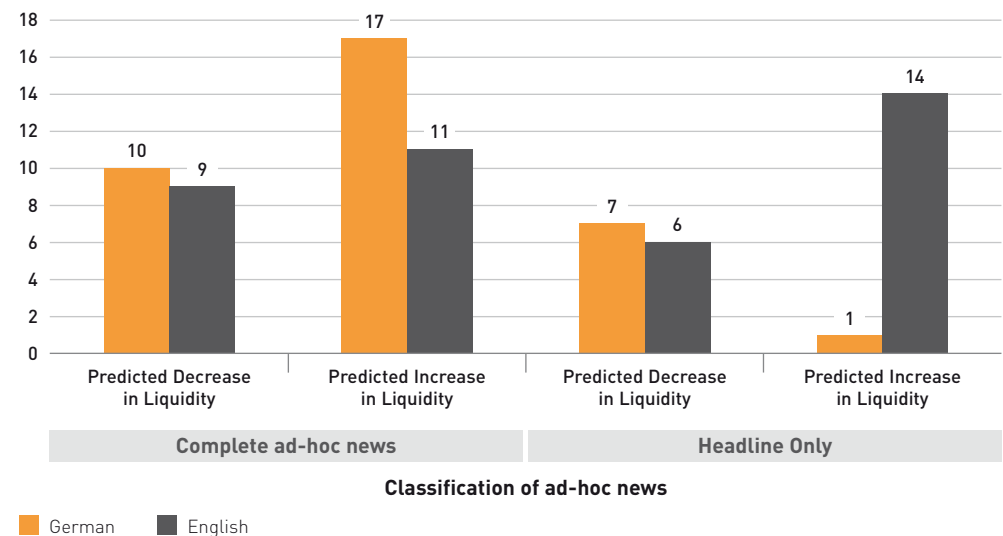**Savings by proposed strategy**
in percentage points

Figure 1: Simulation-based evaluation results of the proposed IT artifact

The proposed simulation setup constitutes an automated trading engine use case. This simulation allows to quantify the economic value of the text mining system. Within the simulation, an automated trader receives the task to execute an order during a specified time interval. The goal is to minimize the implicit transaction costs. To achieve this goal, an optimal execution strategy must be determined. Therefore, each corporate disclosure is classified by the above proposed text mining approach.

If it is predicted that the liquidity of the underlying security decreases during the time interval, the automated trader should prefer to execute at the very beginning of the interval, i.e., immediately after the release of the news but prior to the liquidity impact or at the end of the interval (naïve strategy). In this way, the automated trader can make use of the available (high) level of liquidity before other traders (human traders) change their bids and offers in response to the news disclosure.

In contrast, if the liquidity of the underlying security is expected to increase – or at least not to decrease – during the time interval, a liquid strategy is followed. Therefore, the volume is split into different orders that are executed step-by-step. For each of these cases, we calculate the implicit transaction costs as well as the relation to the opposite strategy which was not followed. The cost savings in percentage points by following the recommended strategy are presented in Figure 1 for different input languages (language of the ad-hoc news) as well as different disclosure components (complete ad-hoc news versus headline only).

Overall, it is shown that the proposed approach is of value since the cost savings are positive in all cases. If the trading signal that is produced by the proposed text mining approach is followed, then the liquidity levels can be forecasted correctly in order to decrease the implicit transaction costs. In the case of, e.g., complete German ad-hoc disclosures, a trader who follows the proposed liquid strategy (i.e. the recommendation of our text mining system) would have to bear implicit transaction costs that are 17% lower compared to the naïve approach, i.e., the execution of the whole volume at the time of publication. These findings provide evidence that our proposed text mining system works well and that the proposed approach can be seen as economically relevant.

Conclusion

Text mining techniques are already applied in various research projects and practical applications to electronically classify financial news. However, most research is focused on the prediction of future price changes of a security. Given that liquidity constitutes one of the most important determinants of implicit transaction costs, we investigate whether text mining allows to predict future levels of liquidity.

Following a structured domain-specific knowledge discovery process, it is possible to extract useful information from unstructured qualitative data to predict future levels of liquidity.

This study contributes in terms of both methodology and practical relevance. Being a highly relevant group of traders and despite their technical capabilities, automated traders require an appropriate decision support system as well.

In this research project, we proposed and successfully tested different ways of enhancing automated trading engines to address news-related liquidity shocks in a timely manner. Future work will concentrate mainly on solving the current limitations of this research, i.e., the proposed forecasting approach shall be compared to existing quantitative forecasting approaches. Because this paper's proposed IT artifact is not intended to replace existing systems (and is instead intended to complement them), future work will concentrate on the integration of our trading signal into existing execution models.

References

Bikker, J.; Spierdijk, L.; Hoevenaars, R.; van der Sluis, P. J.:
Forecasting Market Impact Costs and Identifying Expensive Trades.
In: Journal of Forecasting, 27 (2006) 1, pp. 21–39.

Domowitz, I.; Yegerman, H.:
The Cost of Algorithmic Trading – A First Look at Comparative Performance.
In: Bruce, B. (ed.): Algorithmic Trading: Precision, Control, Execution, Institutional Investor Inc., New York, 2005.

Gomber, P.; Gsell, M.:
Catching Up with Technology – The Impact of Regulatory Changes on ECNs/MTFs.
In: Competition and Regulation in Network Industries, 1 (2006) 4, pp. 535–557.

Groth, S.; Muntermann, J.:
An Intraday Risk Management Approach Based on Textual Analysis.
In: Decision Support Systems 50 (2011), pp. 680-691.

Groth, S.; Siering, M.; Gomber, P.:
How to Enable Automated Trading Engines to Cope with News-Related Liquidity Shocks? Extracting Signals from Unstructured Data.
In: Decision Support Systems, 62 (2014), pp. 32-42.

Han, J.; Kamber, M.:
Data mining: Concepts and techniques.
2nd ed. Elsevier, Morgan Kaufmann, San Francisco, 2006.

Muntermann, J.; Guettler, A.:
Intraday stock price effects of ad-hoc disclosures: the German case.
In: Journal of International Financial Markets, Institutions and Money, 17 (2007) 1, pp. 1–24.

Schwartz, R.; Francioni, R.:
Equity Markets in Action: The Fundamentals of Liquidity, Market Structure & Trading.
John Wiley & Sons, Hoboken, N.J., 2004.