

Big Data Analytics Capstone

Background Analysis

Damian Etchevest

St. Thomas University

CIS-627-170

Table of Contents

Background Analysis.....	3
Introduction	3
Section 1 - Information Driven Bars & Entropy.....	3
Section 2 – Sentiment Analysis & Momentum.....	3
Section 3 – Network Analysis & Instrument Profiling.....	4
References.....	5
Highlighted Pages	6
Section 1	6
High Price Impact Trades in Different Information Environments: Implication for Volatility and Price Efficiency	7
Tracking Retail Investor Activity	11
Discerning Information from Trade Data.....	15
Section 2	20
Stock Market Forecasting Based on Text Mining Technology: A Support Vector Machine Method	20
Predicting Market Reactions to Bad News	24
The Democratization of Investment Research and the Informativeness of Retail Investor Trading ..	26
Kinetic Component Analysis.....	30
Section 3	34
Clustering Financial Return Distributions Using the Fisher Information Metric	34
Partial Mutual Information Analysis of Financial Networks	38
Causal Non-Linear Financial Networks.....	41

Background Analysis

Introduction

These article serves as an extension of the work performed by Maximilian Frankie and Sean Mondesire [1]. On it, they explore the relationship between published news and the stock market reaction. Their research results achieve corresponding *Mean Square Error (MSE)* of 0.039 and an accuracy of 0.652. In contrast to their research, we explore the development of an information driven wave pre-news release and test the confirmation of the tendency on post-news release.

Section 1 - Information Driven Bars & Entropy

Section 1 deep dives into the development of an automated signal generator which performs with live feed stock data to determine when the presence of informed traders exists. This last term, informed traders, refer to the presence of traders that possess fundamental or alternative company's information and generate trades upon understanding its respective future market impact.

Different methods have been tested that can be useful to disclose the information contained in stocks prices. Georges Dionne and Xiaozhou Zhou [1] use *Limit Order Book (LOB)*, a type of order to buy or sell an instrument at a determined price or better, to identify *High Price Impact Trades (HPTIs)*, trades that relate to informed traders which make transactions associated with large price changes relative to their volume proportion. Passive limit orders, from retail investment, are contrarian, consistent with the liquidity provision hypothesis. In contrast, the implementation of Fama-MacBeth regressions on order imbalance measures proposes that aggressive market orders (another form of retail investments) can predict future news, suggesting that retail investors are mainly informed about firm-level news, and that they are likely to have valuable private information [2].

On the other hand, Lopez de Prado [3], claims that with the advances of algorithmic trading, which takes parent orders and chops them into child orders, trading become more complex, since orders and traders are not congruent, and 87% of the executed child orders are passive, therefore blurring the underlying intentions of the parent order. In his study, Lopez de Prado, analyses market microstructure and examines the accuracy of three methods for classifying information driven trades, *the tick rule, the aggregated tick rule, and the bulk volume classification methodology*.

Section 2 – Sentiment Analysis & Momentum

Lots of research has been done exploring the sentiment of news and their market reactions. This section utilizes time series predictions to measure the inference of alternative data sources (offers the opportunity to work with truly unique, hard-to-process datasets) on the stock prices as a means to determine the characteristics of the information-driven price waves.

An interesting approach aims to predict Chinese stock market movements, tendency [4]. To do that, they implement *support vector machines (SVM)*, to classify the tendency, using *text Mining* on online news, their respective comments, and stock market data. Also, their findings show that source's news quality and audience number can serve to calculate their influence by predicting results' difference from the normal prediction results.

A similar study examines the predictability of market reactions to bad news [5]. Their approach is innovative as they first use *time series clustering*, to generate news clusters based on subsequent stock returns, and then apply *SVM* to classify the features extracted from each cluster, by *natural language processing (NLP)* which is a method that interprets human language. Their research identifies four types of market reactions after the news becomes public: downward drift, short-term reversal, medium-term reversal, upward drift.

In their article [6], the authors explore the impact of *Seeking Alpha* (crowd-sourced content service for financial markets) on the stock market's movements after SA research articles are published. They use classification and regression methods and find incremental in order imbalance, significantly related to the sentiment of research articles and comments, which begins within half-hour after SA publications.

Lopez de Prado and Rebonato [7] examines the use of *Kinetic Component Analysis (KCA)* - a state-space application that extracts the signal from a series of noisy measurements by applying a Kalman Filter on a Taylor expansion of a stochastic process - on tick data to determine what they refer as “financial inertia”. They compare the use of KCA to other signal processing tools as *Fast Fourier Transform (FFT)* or *Locally Weighted Scatterplot Smoothing (LOWESS)*, and find that their method: provides confidence intervals estimates of the signal’s position, define the wave’s characteristics (velocity and acceleration of the series), does not exhibit Gibbs phenomenon, and it can be updated online to be forward-looking and resilient to structural changes.

Section 3 – Network Analysis & Instrument Profiling

Finally, we develop clusters of instruments and their respective networks to identify significant information transfers in the different stock networks. The aim is to profile the different stocks to develop behavioral classification models to predict the price wave.

Stephen Taylor [8] fundaments the use of Fisher Information Metric, which is relevant in the fields of information geometry and computing geodesics, to generate clusters of stocks based on their distribution. They contribute valuable theory examples: nearest neighbor comparison using the generalized Pareto distribution, and the maximum daily loss over and annual period distribution hierarchical clustering techniques using the generalized extreme value distribution. Even though his findings do not consider instruments networks, it provides behavioral features that contribute to the stocks profiling.

Pawel Fiedor deeply explores the development of financial networks on the New York Stock Exchange. He analysis the causal relationships within the markets by using *partial mutual information* (a generalization of *partial correlations*), which is sensitive to non-linear dependencies. Fiedor compares partial mutual information methods against those networks methods based on correlation, partial correlation, and mutual information; he proves that the use of his network implementation provides different and more accurate networks [9]. Then, the author extends the analysis using transfer entropy, which is a measure that uses time lags to quantify causal information transfer between systems evolving in time, based on conditional transition probabilities. Pawel Fiedor confirms that asynchronous links are rarely based on the sector of economic activity of the connected stocks [10]. We will extend his studies by identifying the behavior of the stocks in each network when information from one of the stocks in the cluster has news release.

References

- [1] G. Dionne and X. Zhou and H. Montréal, “High Price Impact Trades in Different Information Environments: Implication for Volatility and Price Efficiency”. [Online] Available: <https://www.researchgate.net/publication/340606392> [April 13, 2020]
- [2] E. Boehmer and C. Jones and X. Zhang and X. Zhang, “Tracking Retail Investor Activity”. [Online] Available: <https://ssrn.com/abstract=2822105> [April 8, 2020]
- [3] D. Easley and M. López de Prado and M. O'Hara, “Discerning Information from Trade Data”. *Journal of Financial Economics*, 120(2), pp. 269-286. May 2016; Johnson School Research Paper Series No. 8-2012. [Online] Available: <https://ssrn.com/abstract=1989555> [March 1, 2015]
- [4] Y. Xie and H. Jiang, “Stock Market Forecasting Based on Text Mining Technology: A Support Vector Machine Method.”. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1909/1909.12789.pdf> [July 28, 2016]
- [5] Y. Xiaowen and X. Xin and C. Liangliang and K. Hang Sun, “Predicting Market Reactions to Bad News”. [Online] Available: <https://ssrn.com/abstract=3144041> [March 19, 2018]
- [6] M. Farrell and T. Green and J. Russell and S. Marcov, “The Democratization of Investment Research and the Informativeness of Retail Investor Trading”. [Online] Available: <https://ssrn.com/abstract=3222841> [October 15, 2019]
- [7] M. López de Prado and R. Rebonato, “Kinetic Component Analysis”. *Journal of Investing*, Vol. 25, No. 3, 2016. [Online] Available: <https://ssrn.com/abstract=2422183> [June 5, 2016]
- [8] S. Taylor, “Clustering Financial Return Distributions Using the Fisher Information Metric”. [Online] Available: <https://ssrn.com/abstract=3182914> [May 22, 2018]
- [9] P. Fiedor, “Partial Mutual Information Analysis of Financial Networks” *Acta Physical Polonica A*, vol. 127, no. 3, pp. 863–867. [Online] Available: ResearchGate, https://www.researchgate.net/publication/260678996_Partial_Mutual_Information_Analysis_of_Financial_Networks [March 11, 2014]
- [10] P. Fiedor, “Causal Non-Linear Financial Networks”. [Online] Available: ResearchGate, https://www.researchgate.net/publication/264084873_Causal_Non-Linear_Financial_Networks [August 8, 2018]

Highlighted Pages

Section 1

High Price Impact Trades in Different Information Environments: Implication for Volatility and Price Efficiency

5.2 Trade Types and Price Contribution

Using all transactions from all stocks, we first report the weighted price contribution associated with trade size as in Barclay and Warner (1993). Then we extend our analysis to the weighted

price contribution of roundedness and matchedness²⁸ separately. Finally, we provide a more detailed analysis by jointly considering trade size, roundedness and matchedness. To statistically identify HPITs, we estimate equation (6) with dummy variables of all 12 trade types (3 sizes \times 2 roundedness \times 2 matchedness).

We take size as the first dimension for our trade classification and analysis of weighted price contribution. Specifically, for each stock, trades are first classified as small-, medium- or large-size according to the 30th and 70th percentiles of its own trade size distribution. Then, the corresponding proportion of trades and volumes, and price contribution are computed. Table 7 summarizes the results aggregated by size dimension for the DAX, MDAX and SDAX stocks.

It follows that in our dataset it is the small-size trades that are associated with disproportionately large price changes relative to their proportion of volume. However, using a data sample between 1981 and 1984, Barclay and Warner (1993) find that medium-size trades are the trades associated with disproportionately large price changes relative to their proportion of volume.

The difference between their findings and ours suggests a migration of informed trades from medium-size to small-size trades. One explanation is that trading cost decreases in financial markets. Informed traders always have to trade off between the gains related to their private information and the costs associated with the trading implementation. In previous quote-driven

²⁸At this step of analysis, we keep size (e.g., small, medium, or large) as our first dimension.

to size and matchedness. The results confirm our first hypothesis that in a high public disclosure environment, as liquidity is already high, uninformed traders are likely to place more unmatched trades to meet their given objectives, and informed traders, who are sensitive to both liquidity and price, are likely to submit LOB-matched trades when correcting mispricing. However, in a low public disclosure environment, accompanied by a liquidity shortage, uninformed traders care more about liquidity and are likely to submit matched trades, and informed traders are likely to submit unmatched trades (e.g., marketable trades). Specifically, for DAX stocks (Panel A2), the information quality of matched trades (1.17) is higher than that of unmatched ones (0.82). The opposite is true for MDAX (0.66 for matched trades vs. 1.37 for unmatched trades) and SDAX (0.68 for matched trades vs. 1.23 for unmatched trades) stocks. However, for all three indexes, we observe a migration of informed trades from matched trades to unmatched ones when trade sizes increase. Take DAX stocks (Panel A2) as an example, with only about 2.68% of the total trade volumes, small-matched trades produce about 17.14% of the cumulative price change, which amounts to an information quality of 6.39. However, small unmatched trades, representing 2.73% of the total trade volumes, result in a WPC of -4.61%.³⁰ For medium-size trades, matched and unmatched trades have almost the same information quality. While for large-size trades, the information quality of matched trades is dominated by that of unmatched ones. Panels B2 and C2 report the price contribution of different trade types according to size and matchedness for MDAX and SDAX stocks. We observe slightly different but consistent results: 1) the information quality decreases when trade size increases. 2) Even for the small-size group, information quality of matched trades does not dominate any more than that of unmatched ones. Overall, our results suggest that matchedness, as a complement to roundedness, is an important dimension in informed trading identification.

In sum, the two-dimensional results of size-roundedness and size-matchedness provide evidence that both are important dimensions in informed trader identification. Further, informed trades are mainly associated with small-size trades, which is consistent with the stealth trading hypothesis.

³⁰The negative WPC indicates that on average the associated trades run in the opposite direction to price.

to get rewarded when faced with a large transaction cost (a large bid-ask spread). The same arguments also hold for the least liquid SDAX stocks. That is, SDAX stocks are characterized by the highest bid-ask spread, the lowest proportion of uninformed traders, the least likely short-term price deviations. Overall, in line with previous theoretical models, the negative relationship between HPITs and intraday volatility suggests that HPITs are rewarded for acting as price stabilizers. However, this negative effect decreases when information asymmetry and bid-ask spread increase, which are consistent with the idea that informed traders have to limit their involvement as price stabilizer in presence of unfavorable trade conditions.

7.2 Information Conveyed by HPITs: Autocorrelation Test

Thus far, we have empirically shown that a higher proportion of HPITs leads to a decline in volatility. We now provide more detailed empirical evidence on the channel through which informed trading could reduce volatility and how the magnitude of this decline in volatility varies with information environments. To do so, we define contrarian (herding) HPITs as trades that are against (after) the current price trend. Specifically, buy (positive) HPITs in the presence of decreasing price and sell (negative) HPITs during a price increase are designated as contrarian HPITs. Similarly, buy (positive) HPITs during a price increase and sell (negative) HPITs in the presence of a price decrease are defined as herding HPITs. By definition, contrarian (herding) HPITs involve price correction (price convergence) informed trading. If our conjecture that HPITs are sent by informed traders is correct, we should have negative coefficients for contrarian

³⁹Trading ahead of one's own customers is illegal, but the mentioned strategy entails trading after or against one's customers.

greater public disclosure. The two-dimensional results of size-roundedness and size-matchedness provide evidence that both are important dimensions in informed trading identification. In addition, we show that HPITs not only have a higher cross-sectional price impact, but also a higher price impact over time

With the above identified informed trading, we validate the second hypothesis that a stronger presence of HPITs leads to a decline in volatility for all three groups of stocks. However, this negative effect increases with the level of stocks' public disclosure (i.e., it is highest for DAX stocks and lowest for SDAX stocks). To further explore the trading behavior conveyed by HPITs, we decompose HPITs into contrarian and herding HPITs, and show that contrarian HPITs are responsible for this decline in volatility and lead to price reversals. Again, this negative effect increases with the level of stocks' public disclosure. Our empirical findings support our third hypothesis that 1) contrarian HPITs are more present in the greater public disclosure market and trade against uninformed traders; and 2) herding HPITs are mainly related to fundamental-based information and have an insignificant effect on return autocorrelation.

Finally, we use variance-ratio and autocorrelation-based price efficiency measures to test our fourth hypothesis that HPITs increase price efficiency for stocks with greater public disclosure and high levels of liquidity. Our results show that for DAX stocks, the increase in HPITs improves price efficiency significantly. Similar results are found for MDAX stocks, but are not significant. For the least liquid SDAX stocks, the presence of HPITs slightly reduces market efficiency. We provide two possible explanations for this phenomenon: First, for stocks with more private-information traders and a wide bid-ask spread, informed trades might be followed by uninformed trades that lead a price to deviate from its fundamental level. Further, a high bid-ask spread impedes the price correction made by informed traders. Therefore, price efficiency deteriorates. Second, the midquote price we use to compute the efficiency measure might not be a reliable proxy for the expected true price, especially for the least liquid stocks with a large bid-ask spread. More research is needed to find a better proxy.

Tracking Retail Investor Activity

Can retail investors' activity provide useful information for future stock returns? In this section, we examine the predictive power of our order imbalance measures using Fama-MacBeth regressions as follows:

$$\begin{aligned} \text{ret}(i, w) = & c0(w) + c1(w)\text{oib}(i, w - 1) + c2(w)'\text{controls}(i, w - 1) \\ & + u2(i, w), \end{aligned} \quad (6)$$

where we use the retail order imbalance measure from the previous week, $\text{oib}(i, w - 1)$, and various control variables to predict the next week's stock return, $\text{ret}(i, w)$, for firm i during week w . As in the previous section, because we use overlapping daily frequency data for weekly order imbalances and return measures, the standard errors of the time-series are adjusted using Newey-West (1987) with five lags. If past retail order imbalance can predict future returns in the right direction, we expect coefficient $c1$ to be significantly positive. If coefficient $c1$ turns out to be significantly negative, then retail investors might be making systematic trading mistakes, and if $c1$ is close to zero, we cannot reject the null that retail trading is uninformative on average about the cross-section of future stock returns.

We again include past returns as control variables, using three different horizons: the previous week, the previous month, and the previous six months (from month $m-7$ to month $m-2$). In addition, we include log market cap, log book-to-market ratio, turnover, and daily return volatility, all from the previous month. We report the estimation results in Table III. In regression I, we use the order imbalance based on share volume, oibvol , to predict the next week's return based on bid-ask midpoints. The coefficient on oibvol is 0.0009, with a t -statistic of 15.60. The positive and significant coefficient shows that, if retail investors buy more than they sell in a given week, the return on that stock in the next week is significantly higher. In terms of magnitude, we

B.2. Subgroups in the Cross Section

Our sample includes on average more than 3,000 firms each day. Is the predictive power of retail investor order imbalances restricted to a particular type of firm? Do informed retail investors have preferences for particular types of firms? We investigate these questions by analyzing various firm subgroups in this section. We first sort all firms into three groups based on a firm or stock characteristic observed at the end of the previous month. Then, we estimate equation (6) within each characteristic group. That is, we allow all coefficients in equation (6) to be different within each group, which allows substantial flexibility in the possible predictive relationship across these different groups.

To save space, we include only the results on weekly returns that are computed using the end-of-day bid-ask average price. We first sort all stocks into three different size groups based on market capitalization: small, medium, and large. The results are reported in Panel A of Table IV. In the left panel, we report coefficients on *oibvol*, the order imbalance computed from share volume. When we move from the smallest one-third of firms by market cap to the largest tercile, the coefficient on *oibvol* decreases from 0.0013 to 0.0003, and the *t*-statistic decreases from 13.90 to 3.68. Clearly, the predictive power of retail order imbalance is much stronger for smaller firms than for larger-cap firms, but the predictability remains reliably present in all three groups. Economically, the interquartile difference in weekly returns is 21.9 basis points for the smallest firms (11.39% per year), and 2.6 basis points for the largest firms (1.35% per year). The results in the right panel using order imbalance based on the number of trades (*oibtrd*) are quite similar.

In Panel B of Table IV, we sort all firms into three groups based on the previous month-end share price. In the left panel, moving from the lowest share-price firms to the highest, the coefficient on *oibvol* decreases from 0.0014 to 0.0002, and the *t*-statistics go from 13.34 to 3.23.

three hypotheses. First, as in Chordia and Subrahmanyam (2004), order flows are persistent, and, as the buying/selling pressure is persistent, this could lead directly to the predictability of future returns. Second, as in Kaniel, Saar, and Titman (2008), the retail traders are contrarian at weekly horizons, and their contrarian trading provides liquidity to the market, and thus their trades might positively predict future returns. Third, as in Kelley and Tetlock (2013), retail investors, especially the aggressive investors using market orders, may have valuable information about the firm, and thus their trading could correctly predict the direction of future returns. The above three hypotheses are not exclusive. In Section II.C.1, we conduct a simple decomposition to separate alternative hypotheses. In Section II.C.2, we provide more evidence regarding the liquidity provision hypothesis.

C.1. Two-Stage Decomposition

To distinguish among these alternative hypotheses, we adopt a two-step decomposition procedure. In the first step, we decompose the weekly retail order imbalance into three components for each week w , with the following cross-sectional regressions:

$$oib(i, w) = d0(w) + d1(w)oib(i, w - 1) + d2(w)'ret(i, w - 1) + u4(i, w). \quad (8)$$

After we obtain the time-series of coefficients, $\{\widehat{d0}(w), \widehat{d1}(w), \widehat{d2}(w)'\}$, we define the following terms:

$$\begin{aligned} \widehat{oib}_{i,w}^{persistence} &= \widehat{d1}(w)oib(i, w - 1), \\ \widehat{oib}_{i,w}^{contrarian} &= \widehat{d2}(w)'ret(i, w - 1), \\ \widehat{oib}_{i,w}^{other} &= \widehat{u4}(i, w). \end{aligned} \quad (9)$$

That is, we denote the part related to the past order imbalance as the “persistence,” which is related to the price pressure hypothesis. The part related to past returns over different horizons is denoted as “contrarian,” which relates to the liquidity provision hypothesis. After we take out predictability

due to “persistence” and “contrarian” trading, we denote the residual part as “other,” which we attribute to retail investors’ information about future returns.

At the second stage, we estimate the following regression using the Fama-MaBeth methodology, which is parallel to equation (6):

$$\begin{aligned} ret(i, w) = & e0(w) + e1(w)\widehat{oib}_{i,w-1}^{persistence} + e2(w)\widehat{oib}_{i,w-1}^{contrarian} \\ & + e3(w)\widehat{oib}_{i,w-1}^{other} + e4(w)'controls(i, w - 1) + u5(i, w). \end{aligned} \quad (10)$$

Since we decompose the original order imbalance measure into three parts, related to order flow persistence, a contrarian trading pattern, and the residual, the coefficient estimates in equation (10) reveal how each component helps to predict future stock returns.

We report the decomposition results in Table VII. Panel A presents the first-stage estimation as in equation (8), which is quite similar to those reported in Table II. Take the first regression as an example. The order imbalance measure, *oibvol*, has a highly significant and positive coefficient on its own lag at 0.22, which indicates order persistence. In terms of past returns, the coefficients for the past week, past month, and past six-month returns are -0.9286, -0.2029 and -0.0267, respectively, all implying contrarian trading patterns.

After we decompose the previous week’s order imbalance into “persistence,” “contrarian,” and “other,” we use them separately and together to predict future stock returns, as in equation (10). In the first regression, we use the past week’s *oibvol* to predict future bid-ask return. The coefficient estimate on *oib (persistence)* is 0.0027, with a t-statistic of 8.75, which implies that price pressure significantly and positively contributes to the predictive power of the retail flow. The coefficient estimate on *oib (contrarian)* is -0.0044, and it is insignificantly different from zero, implying that we cannot reject the null hypothesis that the contrarian component does not contribute to the predictive power of retail order imbalances. Finally, for the *oib (other)*

Discerning Information from Trade Data

These difficulties suggest that any practical solution to this problem will involve an approximation. The simplest approximation is the tick rule, which assigns a trade to be a buy if the trade price was an uptick relative to the previous trade and to be a sell if it was a downtick (in the case of a zero-tick the signing relies on the movement relative to the last price change). This approach eschews any distributional assumptions and relies instead on the basic notion that buys raise prices and sells lower them. But how well this approximation works to infer trades, or underlying information, is debatable, particularly in light of the trading practice and market structure issues raised earlier.

If there is noise in the data, the Bayesian approach does not provide a point prediction (of a buy or sell, for example) but rather a posterior probability. This is the intuition that underlies the bulk volume classification algorithm. Our approximation aggregates trades over short time or volume intervals and then uses the standardized price change between the beginning and end of the interval to approximate the percentage of buy and sell order flow. Thus, this approach can be interpreted as assigning probabilities to buys and sells given the observable data. Intuitively, we say that the underlying trade was more likely to have been buyer-initiated the larger, and more positive, is the price change and more likely to have been seller-initiated the smaller, and more negative, is the price change, relative to the distribution of past price changes.⁷

A (time or volume) bar τ is assigned the price change $P_\tau - P_{\tau-1}$, where P_τ is the last price included in bar τ , and $P_{\tau-1}$ the last price included in bar $\tau - 1$.⁸ To define the *bulk volume* procedure, let

⁷ See also Easley, et al. (2012a) where we apply this technique in estimating VPIN measures, and Gollapulli and Bose (2013) who use this approach to estimate order imbalances in swap markets.

⁸ We start the first bar with the second transaction in our sample, so that the algorithm has a P_0 for initialization.

$$\hat{V}_\tau^B = V_\tau \cdot t \left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}, df \right) \quad (1)$$

$$\hat{V}_\tau^S = V_\tau \cdot \left[1 - t \left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}, df \right) \right]$$

where V_τ is the volume traded during (time or volume) bar τ which we wish to classify in terms of buy and sell volume \hat{V}_τ^B and \hat{V}_τ^S , and t is the CDF of Student's t distribution, with df degrees of freedom.⁹ $P_\tau - P_{\tau-1}$ is the price change between two consecutive bars and $\sigma_{\Delta P}$ is our estimate of the standard derivation of price changes between bars. Our procedure splits the volume in a bar equally between buy and sell volume if there is no price change from the beginning to the end of the bar. Alternatively, if the price increases, volume is weighted more toward buys than sells depending on how large the price change in absolute terms is relative to the distribution of price changes.

2.1 Statistics for Bulk Volume Classification and the Tick rule

Comparing the performance of the tick rule and the BVC approaches is not straightforward as they do not produce the same type of output. The tick rule produces a list of buy and sell classifications, one for each trade, whereas BVC produces a list of fractions of buys and sells, one for each bar (time, volume or trade) to which it is applied. Even on a single bar (with multiple trades in the bar) they produce different output: the tick rule provides a list of buys and sells and BVC provides fractions of buys and sells. To compare the two approaches we consider two transformations. The first, and most obvious, is that we compare the actual tick rule with an application of BVC to a bar containing a single trade. Applying BVC on a single trade makes sense if we interpret BVC as assigning to any observation a probability that the

⁹ We use the t-distribution because the parameters of the true distribution are unknown. Other distributions, such as the Normal or the actual empirical distribution of the data, can be used, but in empirical testing we found no improvement over results from the t-distribution. Based on calibration, we used $df = 0.25$ to account for the fat tails present in the data.

underlying trade was a buy. In the second transformation we create an aggregate version of the tick rule and compare it to the BVC procedure which is already in a bulk form.

We first show that, even in the single trade case, whether BVC or the tick rule does a better job of trade classification depends on how informative the trade is about the underlying data we want to infer. BVC does better if the observation is very noisy, which seems more likely if we interpret the underlying data as information (good or bad news) and less likely if we interpret the underlying data as a trade (Buy or Sell).

2.2. Observations Classified One-By-One

Suppose that we observe a price change ϵ where the distribution of ϵ differs if the (unobservable) trade type was a Buy or Sell.¹⁰ Assume that $\epsilon \sim d(\bar{\epsilon}, \sigma^2)$ if the trade was a Buy, probability that the news (unobservable) was buy or sell: The price change must be different than the predicted price change (the probability that the price change ϵ was buy or sell differ from the probability of the news being buy or sell)

and $\epsilon \sim d(-\bar{\epsilon}, \sigma^2)$ if the trade was a Sell, where $\bar{\epsilon} > 0$. We denote the prior probability that the unobservable trade was a Buy by $PR(Buy) = p$, where $0 < p < 1$. We consider three methodologies to assign a probability that the underlying trade type was a buy or a sell given the observation of a single draw of ϵ : Bayes rule, the tick rule and BVC specialized to a single observation. The tick rule assigns probability one or zero to the trade having been a Buy. BVC when applied to one observation can be interpreted as assigning the probability of a Buy. Bayes rule, of course, actually assigns a probability of the trade having been a Buy. For each methodology, the formula for the conditional probability of a Buy is:

1. Bayes: $B(\epsilon) = \frac{pPR(\epsilon | Buy)}{PR(\epsilon)}$ where $PR(\epsilon) = pPR(\epsilon | Buy) + (1-p)PR(\epsilon | Sell)$
2. Tick: $T(\epsilon) = 1$ if $\epsilon > 0$ and $T(\epsilon) = 0$ if $\epsilon < 0$

¹⁰ We will refer to trade types and Buy or Sell, but the analysis also applies if we interpret the unobservable event as information which can be good or bad news.

Table 6 also reports the result of a regression including lagged high-low spread and both order imbalance measures. Again the coefficients on tick rule and BVC order imbalances have opposite signs. This confirms our hypothesis that, once we take into account the combined market impact from aggressive and passive trades, we have a much better (and consistent) explanatory model of changes in liquidity. To the extent that the high-low spread is a good proxy for the portion of trade arising from informed traders, these results suggest that the tick rule order flow imbalance fails to detect the presence of informed traders, while the BVC order flow imbalance succeeds in doing so.

3.3 Order imbalance and daily price changes

Over short intervals, underlying new information should affect the willingness of market makers to provide liquidity. Over a longer interval, however, market efficiency dictates that new information should affect market prices. This linkage between trading and market efficiency of prices is a fundamental insight of market microstructure research, and it sets the stage for our third accuracy test. If order imbalances are signals of informed trade, then these imbalances should be correlated with price movements.

Testing this proposition using futures data requires careful consideration of some particular features of futures market microstructure. Unlike equity markets which have defined closing times, futures trade on a 24 hour cycle. Trading volume, however, can vary wildly over this interval. In the case of the e-mini S&P future, for example, trading can be frenetic around the open and close of the U.S. equity markets, and somnolent during U.S. overnight hours. Such time patterns might be expected to influence tests relating order imbalances measured using time intervals, but should be of much less importance for imbalances measured over volume increments.

has its limitations and in this section we discuss these in more detail. We also consider the broader issues related to implementing this approach over asset classes more generally.

One issue has to do with price changes. Both tick rule approaches and bulk volume classification techniques rely on price changes to classify trades. For tick rule approaches, these price ticks are computed for each trade; for BVC, it is the net price change between the beginning and end of a time bar or volume bar that matters. As discussed in Section 2, under certain conditions these approaches will be identical – the cumulative sum of the individual ticks will equal the net change in prices over the interval (see Result 4). In this case, BVC would be expected to provide no added benefit over using a tick rule approach. But, generally, these two approaches will not be the same and it is useful to understand why this is the case. If trades are of different sizes, or have different price impacts over the day, or if sequences of trades elicit different price responses, then BVC and tick rule estimates will differ. They will also differ if changes in the book can signal new information as this can result in large price changes arising from little trading volume. All of these are features characteristic of high frequency markets. In such settings, adding up individual up ticks and down ticks will not provide the same information as is captured in the price change used in BVC.

Why does this matter? We are interested in this trade data because it tells us about the underlying information motivating trade, and that information is unobservable. So we must look for its reflection in markets – and that generally involves some aspect of price behavior. This is a standard approach in microstructure where tick-signed order imbalance have been shown to relate well to bid and ask price movements, or tick-signed trade price movements decomposed into temporary and permanent price effects are related to inventory and information respectively.

The non-linear transformation we apply in BVC is another example of how price-based proxies

Section 2

Stock Market Forecasting Based on Text Mining Technology: A Support Vector Machine Method

types of communications or machine learning algorithms [5]. Besides, data sources also experiences a big change. Researchers are no more satisfied with historic stock data. They try to make use of different sources' data and make assumptions based on that [4].

Among these additional sources, text sources are starting to become heated because of its maneuverability [5]-[7]. Actually, text sources are also divided into two parts. One is focused on the direct feedback of customer and market reception for a new product to estimate the popularity and price of a product, even a stock. It is mainly based on identifying positive and negative words and processing text with the purpose of classifying its emotional stance as positive or negative, widely used in predict turning point or tendency of a specific stock [8]-[10]. The other one tries to deal with passive influence from external text sources. Such sources are usually from mass-media, like main websites, newspapers and magazines [9], [11]. These texts tell objective information and investors react to this information differently in their investment decision-making process.

Chinese stock market is special around the world. Unlike western countries' stock market, most stock investors are ordinary people, not investment firms and corporations. Besides, the majority of Chinese investors do not have enough financial or economic knowledge. Most of them make money by short-term speculative trading and can be easily manipulated by external information. Thus, news should have significant effect on Chinese stock market. However, till now there are not many researches trying to deal with this phenomenon. Facing so many difficulties, this research will focus on support vector algorithm using text mining method to deal with online news and predict Chinese stock market. We hope to demonstrate a new approach to describe stock market movement and uncover unique mechanism of Chinese stock investor behavior as well.

2. Research Method

Like many other similar researches, this research is mainly divided into 3 stages: data collection; data pre-processing; machine learning and forecasting, as shown in Fig. 1.

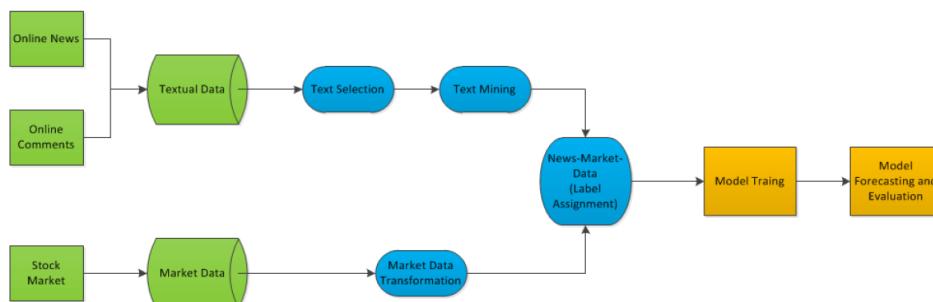


Fig. 1. General research process.

2.1. Data Preparation

This research merges textual data and market data together to form input vector. When collecting textual data, it is essential to find out all possible news for any stocks. There are two benefits doing so: 1) text mining dictionaries can be more precise; 2) more latent influence from text data can be captured. By contrast, stock type does not need to be overall. However, the type of testing stocks should be representative. Thus,

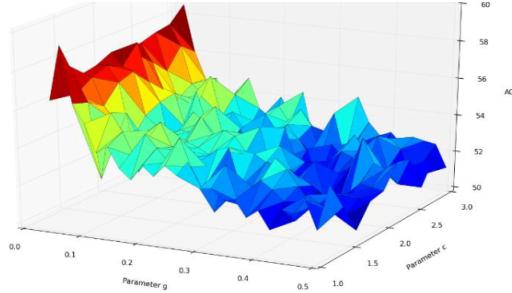


Fig. 9. Traverse algorithm best parameter plane (SVC).

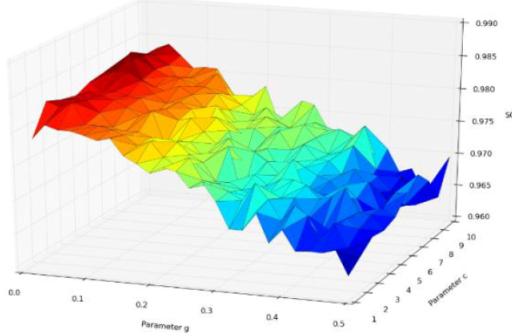


Fig. 10. Traverse algorithm best parameter plane (SVR).

The figures also show that g is the main factor in the model which proves that searching g first is correct in approximate algorithm discussed the latter. Actually, when g is near 0, the model will perform much better. For all 20 stocks, the traverse algorithm result is shown below.

Table 3. Traverse Algorithm's result of All Stocks for SVC

Com_code	Best g	Best c	Best ACC (%)	Com_code	Best g	Best c	Best ACC(%)
601288	0.02	1.4	60.16462	601398	0.02	1.6	54.83435
600016	0.3	1.6	53.10473	601899	0.28	2.4	54.74664
601668	0.48	2	53.65812	600837	0.16	2.8	51.43552
601988	0.04	1.2	59.1734	601989	0.06	2.2	55.49496
601818	0.02	1.4	59.12176	601328	0.02	1.6	57.57212
600030	0.44	1.2	53.36816	601166	0.08	1.2	53.47609
600050	0.26	2.6	53.95256	600028	0.02	2.4	52.82336
600000	0.28	2.4	51.70183	600036	0.02	1.2	52.86296
600010	0.16	2.4	53.23738	601901	0.34	1.8	52.84631
600795	0.02	3	56.99535	601390	0.02	1.8	59.3652

Table 4. Traverse Algorithm's Result of All Stocks for SVR

Com_code	Best g	Best c	Best MSE	Best SCC	Com_code	Best g	Best c	Best MSE	Best SCC
601288	0.02	4	0.002898	0.982576	601398	0.02	7	0.003794	0.975532
600016	0.02	9	0.02135	0.992012	601899	0.04	10	0.012369	0.990894
601668	0.06	4	0.009668	0.98962	600837	0.06	10	0.143558	0.987999
601988	0.04	7	0.00328	0.985429	601989	0.02	9	0.027772	0.986388
601818	0.02	10	0.006065	0.984559	601328	0.04	5	0.01065	0.990556
600030	0.04	9	0.226013	0.991033	601166	0.04	8	0.064489	0.987322
600050	0.08	8	0.009361	0.99278	600028	0.04	10	0.020217	0.992863
600000	0.04	6	0.061748	0.991011	600036	0.04	9	0.053716	0.981747
600010	0.04	8	0.009559	0.98801	601901	0.04	7	0.074697	0.989365
600795	0.02	9	0.003774	0.985014	601390	0.04	8	0.023091	0.993408

As we can see from the tables, best parameter set (c, g) of 20 stocks for SVC are differential, while for SVR

$$Z_j = \sum_{i=1}^{20} M_{ij} \cdot V_i$$

where V_i represents the trading volume of No. i stock.

After scaling, the final source weight is shown in Table 5.

Table 5. Source Weight

No.	Source	Weight	No.	Source	Weight
1	Wallstreetcn	19.4154	11	ccstock	18.78984
2	Stockstar	19.37817	12	Cnstock-SH	18.78599
3	Chinanews	19.08397	13	Cnstock	18.73956
4	Cnstock & CS	19.03179	14	Cnstocknet	18.73758
5	21st CBH	19.01998	15	STCNnet	18.73686
6	NBD	18.95937	16	Eastmoney	18.71921
7	Yicai	18.95242	17	Xinhuanet	18.7151
8	Tencent Finance	18.90835	18	NNBD	18.65674
9	Xinlang Finance	18.88499	19	Aastocks	18.65194
10	STCN	18.87678	20	Hexun	18.64479

The result is reasonable relating to Chinese stock market case. It seems there are two major factors determining the source's weight. One is the news quality. Like the *wallstreetcn*, *chinanews* and 21st *CENTURY BUSINESS HERALD*, they are all famous for its high news quality. The other is the audience number. Actually *Tencent Finance* and *Xinlang Finance* are not the original birthplace of news documents. Just because they have huge amount of audience, they can stand so front in this rating. In addition, the main high weight sources come from the newspapers and magazines, showing that the majority of Chinese investors are still depend on paper news as their financial information source.

5. Conclusions and Discussions

In general, the SVM method based on text mining technology shows an outstanding result in predicting stock market especially when predicting the specific stock price. In comparison, when predicting the stock tendency, SVC is not good enough, though it also proves the importance of news documents.

In addition, this paper proposes two simple and basic parameter-find algorithms and discusses the relationship between them. The result shows that traverse algorithm is better. Furthermore, based on the test best parameter-searching plane, in SVM parameter g is the main factor and one can just choose search g first in order to increase searching speed.

This paper also proposes a simple approach to evaluate the online sources' weight in affecting investors' trading decision-making. The weight result is based on the real effect and is surely much better than methods based on the audience number or specialists' proposal.

However, the SVM is not perfect yet. When news amount is low, the predicting result is not good enough. In such cases, even some basic and simple linear regression algorithm can do a better job.

There are also several aspects needing to make sure and improve in order to get a more robust SVM method. 1) Adding more text sources; 2) Designing a standard sentiment evaluation system and finding more specialists to score the sentiment dictionary; 3) Making sure the stock needing to be predicted has enough text documents and trading volume; 4) Trying to expand the whole dataset.

References

Predicting Market Reactions to Bad News

classify the news in the test set into one of the clusters based on their features.

This is a procedure of predicting subsequent market reaction patterns for

updated news. We conduct SVM classification model for both short-term and

medium-term horizons. For short-term horizon, SVM is used to classify the

news in the test set to be one of the two labels, i.e., upward and downward.

While for medium-term horizon, SVM is used to classify the news to be one of

the four labels we obtained through time series clustering.

5.3.1 Accuracy for short-term prediction

For short-term prediction, the accuracy scores on the training set and the

test set when features are generated by LDA, LSA and NMF models are shown

in Table 5.3.1. The accuracy score is the ratio of correctly classified observation

by SVM, compared with the upward/downward labels we made based on

next-day stock price movement, to the total observations. It shows that LDA

gives the best in-sample classification accuracy, while NMF gives best

out-of-sample classification accuracy. The overall accuracy scores are not high,

most of which are just slightly over 0.5. We think this is mainly caused by the

size of the final dataset for SVM model. However, as we will prove in the

trading strategy part, even though the prediction accuracy is not very high, we

still can obtain relatively good returns using our short-term trading strategy.

Furthermore, we have developed a machine learning based framework for predicting subsequent market reactions to negative news update for both short-term and medium-term horizons. We have implemented different NLP models to extract the features for each cluster and applied the SVM classifier on the news in the test set to label each news based on its features. The accuracy scores for the short-term prediction are not as good as we have anticipated, but the accuracy scores for the medium-term prediction are relatively good.

Lastly, we have moved our analysis further by developing trading strategies for both short-term and medium-term horizons that are viable from the framework we developed. The short-term strategy we have designed to bet on the next-day stock price movement gives us good returns, for both in-sample and out-of-sample implementation, even though the daily returns are not very stable over time, which is common when the signal is generated from the news. The medium-term strategy we have developed to bet on the medium-term reversal also provides acceptable performance, with its cumulative return beating the market. The back-test results over ten years demonstrate the effectiveness of our trading strategies.

We would like to conclude by sharing our plans in the future. It is always important to remember that having a good source of data is essential in any academic research. We believe that our research can have more depth in the

informative if a Seeking Alpha article will be published in the next one to five days (days $t+1$ through $t+5$), with none of the estimates being statistically significant at the 10% level. Furthermore, for the subset of more informative articles, the estimates are always significantly smaller than the estimate on $SA_Event_{i,t-1,t}$. Collectively, the evidence mitigates the concern that unobserved public information spuriously explains the relation between SA research and informative retail trading.

6.2 Decomposing Retail Trading into Price Pressure, Liquidity Provision, and Informed Trading

In Table 4 we find that retail investor order imbalances are highly persistent (i.e., the coefficient on $Retail\ OIB_{w-1}$ is significantly positive). This finding raises the concern that buying or selling pressure could explain the predictability of returns, particularly if there is greater persistence in order imbalances after Seeking Alpha research. In other words, Seeking Alpha research may amplify noise trading among retail investors, generating price pressure and resulting in short-term return predictability.²² Table 4 also documents that retail investor order imbalances are contrarian over short horizons (e.g., the coefficient on Ret_{w-1} is significantly negative). Short-term contrarian trading is a common proxy for liquidity provision (e.g., Nagel, 2012; Jame, 2018), raising the possibility that the positive association between retail order imbalances and future returns is attributable to liquidity provision rather than informed trading.

²¹ This finding is also consistent with the evidence in Table 5, which shows that roughly 41% of the incremental 10-day returns is captured in the first day subsequent to the trade (0.050%/0.123%).

²² This view is bolstered by the evidence that “fake” articles, i.e., those that are financially sponsored (but undisclosed) by the covered firm, are at least occasionally associated with significant price impact (Kogan, Moskowitz, Niessner, 2018, Clarke, Chen, Du, and Hu, 2019). We note that articles that have been identified as fake by the SEC have been removed from Seeking Alpha and are excluded from our sample. In separate analysis, we collect a list of fake articles from the *Financial Times* (<https://ftalphaville-cdn.ft.com/wp-content/uploads/2017/04/10231526/Stock-promoters.pdf>). Examining the incremental informativeness of retail order imbalance around these fake articles, we find negative but statistically insignificant estimates. The fake news evidence is inconsistent with retail investors benefiting from the price pressure associated with fake articles.

We explore the potential role of price pressure and liquidity provision following the approach in BJZZ. In particular, we decompose retail order imbalances into three components: *OIB Persistence* (a proxy for price pressure), *OIB Contrarian* (a proxy for liquidity provision), and *OIB Other* (a proxy for informed trading). The three components are estimated as the fitted value from the following panel regression: $Retail\ OIB_{it} = \alpha + \beta_1 Retail\ Oib_{iw-1} + \beta_2 Ret_{i,w-1} + \varepsilon_{it}$, where $OIB\ Persistence = \hat{\beta}_1 OIB_{w-1}$; $OIB\ Contrarian = \hat{\beta}_2 Ret_{w-1}$; and $OIB\ Other = \hat{\varepsilon}_{it}$. We then estimate Equation (4) after replacing total retail order imbalance (*Retail OIB*) with *OIB Persistence*, *OIB Contrarian*, or *OIB Informed*.

Specifications 1 through 3 of Table 8 report the results for the full sample, and Specifications 4 through 6 report results for the subsample of *Informative* articles, as defined in Table 6. In both samples, the coefficients on *OIB Persistence* \times *SA* and *OIB Contrarian* \times *SA* are statistically insignificant, whereas the coefficient on *OIB Other* \times *SA* is highly significant. The evidence is inconsistent with either the liquidity provision or price pressure explanations and points towards *SA* articles contributing to more informative retail trading.

A second approach to disentangle price pressure from informed trading is to explore the return patterns over longer horizons. In particular, the price pressure explanation predicts reversals over longer holding periods as prices revert to fundamentals. In contrast, the informed trading explanation predicts, at a minimum, that returns should not revert. To the extent that retail investors trade on information that is slowly impounded in prices, we may observe a drift over longer horizons.

We re-estimate Specification 3 of Table 5 using returns measured over the subsequent 10, 20, 40, or 60 trading days. For brevity, in Figure 2 we only plot the coefficients on *Retail OIB* (non-*SA* days) and the coefficients on *Retail OIB* + *Retail OIB* \times *SA* (*SA* days) in the full sample

investment research, with the popular provider of informative crowdsourced research Seeking Alpha playing a central role (Chen et al. 2014). In this article, we explore whether this phenomenon enhances the informativeness of retail investor trading.

Our initial findings confirm anecdotal evidence that Seeking Alpha research is geared towards retail investors, with SA coverage being significantly negatively related to institutional ownership and positively related to number of shareholders. We also find strong evidence that retail investors react to Seeking Alpha research, with significant increases in retail investor trading activity on days with Seeking Alpha articles. Moreover, retail order imbalances are significantly related to the sentiment of research articles and comments, and the relation begins within a half-hour of publication.

More importantly, we document that Seeking Alpha research enhances the informativeness of retail investor trades. In particular, the relation between retail order flow and future stock returns is roughly three times as strong on days with Seeking Alpha research articles, and the SA findings are stronger than the analogous evidence for media articles or traditional brokerage research. We find that the informativeness of retail trading after SA research continues to hold after controlling for SA tone, suggesting that retail investors extract value-relevant information from SA articles that extends beyond article tone.

Consistent with this view, we find that retail investor trading is particularly informed after more informative SA research. For example, articles written by contributors with longer bios, and by those whose past research has been more impactful, result in a stronger post-article connection between retail order flow and future returns. The predictive ability of retail order flow after SA research is also stronger for negative sentiment research, consistent with these articles being more novel, and when the SA article elicits more comments, consistent with active investor engagement

Kinetic Component Analysis

One important advantage of KCA over FFT is that the former provides estimates of the means as well as the standard deviations of the hidden states. Figure 3 plots the states estimates as well as the 2 standard deviation confidence intervals.

[FIGURE 3 HERE]

In summary, KCA presents three advantages over FFT: i) KCA provides point as well as confidence interval estimates of the signal's position, while FFT only provides a point estimate. ii) beyond the position state, KCA also reveals information regarding the velocity and acceleration of the series (with confidence bands for the three of them). iii) KCA's extracted signal is closer to the true signal at the extremes of the series, because KCA does not exhibit the Gibbs phenomenon. This third advantage is critical, because a researcher is typically interested in extrapolating or forecasting a signal, which requires that the most recent estimates are the most accurate.

5. KCA vs. LOWESS

Locally Weighted Scatterplot Smoothing (LOWESS) is another popular method used to deal with noisy measurements. LOWESS fits weighted linear regressions to localized subsets of the data in order to build a function that filters noise point by point (Cleveland [1979]). Figure 4 plots the result of fitting several LOWESS functions to the same observations used in Section 4. When LOWESS is fit on local regressions that employ 50% of the data, the fit barely resembles the signal. A LOWESS function that uses 25% of the data gives a result similar to FFT's. A LOWESS function of 10% is very close to KCA's estimate. As the fraction of data is reduced, the LOWESS function fits the signal more closely, but unfortunately it also becomes more unstable. We can appreciate the onset of that phenomenon on the LOWESS(0.1) function, which exhibits several irregularities or bumps.

[FIGURE 4 HERE]

Figure 5 provides an example of a 20 step forward forecast performed by the KCA algorithm. KCA presents several advantages compared to LOWESS: i) LOWESS is not equipped to generate forecasts, because each estimate is highly dependent on the neighboring sample (before and after the observation). KCA incorporates a Kalman filter, thus inheriting the features of state-space signal processing methods. In particular, KCA is a forward-looking method that is robust to structural changes. ii) LOWESS requires a subsample to generate every single estimate, which makes it computationally intensive. Not only KCA can produce accurate forecasts several steps forward, but it can be updated online. This means that the last known state of the system is all KCA needs to forecast the next. iii) Like in the case of FFT, LOWESS does not decompose the signal into the three kinetic components, nor provides confidence intervals for those estimates.

[FIGURE 5 HERE]

6. SOME APPLICATIONS

6.1. FINANCIAL INERTIA

As we mentioned in Section 2, the financial literature has debated for decades whether financial momentum exists. We have called this section “Financial Inertia” rather than “Financial Momentum” because we believe that the former is what the academic literature actually meant. In plain English, *inertia* is the tendency of an object to keep moving in a straight line at a constant speed. The principle of inertia is postulated in Newton’s First Law of Motion. In contrast, *momentum* is the product of mass and velocity. Newton’s Second Law tells us that it takes the same force to deviate an object at double speed with half mass or at half speed with double mass. Momentum requires the definition of mass and velocity, while inertia is merely the observation that an object’s velocity remains unaltered unless a force acts upon it.

It is not the goal of this section to settle a long-standing controversy, but to demonstrate the use of KCA. We have studied the price dynamics of some of the most liquid investments across all asset classes. By applying KCA on price series, we can extract estimates of price acceleration. An instrument exhibits financial inertia when its price acceleration is not significantly greater than zero for long periods of time.

Table 1 summarizes our data. Our source is level 1 tick data recorded by TickWrite. For each instrument we used the front contract, rolled forward by volume. Tick series were grouped in volume buckets, at an average of 1 bucket per day. KCA was then applied on the series of volume weighted average prices (VWAP) computed on each bucket.

[TABLE 1 HERE]

Table 2 lists the key inertia statistics per contract. *Mean_Accel* is the average value of the acceleration. *Std_Accel* is the standard deviation of the estimated acceleration values. *Inertia* is the proportion of acceleration estimates that did not exceed a 95% confidence bound centered around zero. This value can be interpreted as the proportion of market activity associated with insignificant acceleration. The greater the inertia, the greater was the amount of activity (measured as transacted volume) that occurred under a relatively unchanged price speed.

[TABLE 2 HERE]

Inertia results are generally high, with an average value of 0.8942 and a standard deviation of 0.067. Corn (CN) registers the lowest reading, at 0.7615, and Euro FX (EC) the highest, at 1. Out of the 19 instruments studied, 9 exhibit an inertia greater than 0.9: Dollar Index (DX), Euro FX (EC), Natural Gas (NG), E-mini S&P500 (ES), XX (Eurostoxx 50), E-mini Dow-Jones (YM), Eurodollar (ED), T-Note 5 years (FV) and T-Note 2 years (TU).

An intuitive result is given by futures on treasury notes, where the inertia is lower as the duration increases. As short term interest rates are anchored by the Federal Reserve’s policy, inertia gradually increases as we move away from the yield curve’s front end: From 0.9986 in the case of Eurodollar (ED), all the way to 0.8588 in the case of T-Bond 30 years (US).

In terms of average inertia per asset class, the highest average values seem to be associated with currencies (0.9540, although we only count with two examples) and rates/fixed income (0.9298). These are followed by equity indices (0.8963) and commodities (0.8425).

Much has been published in recent financial outlets regarding the poor performance experienced by momentum or trend following funds. This contrasts with the above results, as they evidence strong inertia (the basis for the profitability of momentum funds). One possible explanation is the application of flawed trading rules to the monetization of existing momentum opportunities. A popular trading rule is called “crossing moving averages momentum”, and it consists in taking a long position on product X whenever the moving averages of sample sizes m and n satisfy the condition $\bar{x}_m > \bar{x}_n$, where $m < n$. A short position would be triggered by $\bar{x}_m < \bar{x}_n$. Since the sample size determines a limited number of parameter combinations that m and n can adopt, it is relatively easy to determine the pair (m, n) that maximizes the backtest’s performance. This sort of backtest overfitting has been shown to lead to negative performance in the presence of memory effects (Bailey et al. [2014a, 2014b]). Furthermore, let us not forget that “crossing moving averages” is an adaptive trading rule, in the sense that its estimates are purely historical and do not attempt to anticipate future behavior. That is not the case of KCA, which relies on Bayesian learning to update forward-looking priors.

For these two reasons (overfitted and adaptive trading rules), it is not unreasonable to think that momentum funds can generate substantial losses even in the presence of strong momentum. A better approach may have been to invest in momentum opportunities by applying forward-looking trading rules that are not so easily overfitted.

6.2. MICROSTRUCTURAL NOISE

Beyond the traditional study of financial momentum, KCA can be used in a variety of contexts. One such application is the modeling of price dynamics under microstructural noise. With the advent of High Frequency Trading, a large percentage of quotes are generated with no intention of actual trading. Also, trades often lead to positions that last barely a few seconds or even less. This microstructural noise makes it difficult to determine towards what levels prices are trending, and at what speed.

Suppose that a market maker targets to hold inventory for a maximum period h , in chronological or volume time (see Easley et al. [2012b] for the difference). In Section 4 we demonstrated how KCA provides estimated means and confidence intervals for our three components. We can use these estimates to determine the levels at which market makers can provide liquidity. In particular, let $\underline{p}_s, \underline{v}_s, \underline{a}_s$ be the lower bound estimates for position, velocity and acceleration, and $\bar{p}_s, \bar{v}_s, \bar{a}_s$ the respective upper bound estimates associated with a significance level α . For example, $\underline{p}_s = p_s + Z_\alpha \sigma_{p,s}$ and $\bar{p}_s = p_s - Z_\alpha \sigma_{p,s}$, with analogous expressions for the other components. Then, we can apply the Taylor expansion in Section 3.1 to determine a confidence interval for p_{s+1} , at which liquidity can be provided

$$\begin{aligned}\underline{p}_{s+1} &\approx p_s + Z_\alpha \sigma_{p,s} + (v_s + Z_\alpha \sigma_{v,s})h + \frac{1}{2}(a_s + Z_\alpha \sigma_{a,s})h^2 \\ \bar{p}_{s+1} &\approx p_s - Z_\alpha \sigma_{p,s} + (v_s - Z_\alpha \sigma_{v,s})h + \frac{1}{2}(a_s - Z_\alpha \sigma_{a,s})h^2\end{aligned}\tag{7}$$

Section 3

Clustering Financial Return Distributions Using the Fisher Information Metric

4.4. Computing Geodesic Distances. Computing geodesic distances in two and higher dimensional models in a robust manner is challenging due to complex forms that the geodesics equations may take. Recent advances in [38, 39] have overcome some of these issues by focusing on solving for the distance function first from which geodesic paths are extracted. Specifically, the author develops numerical schemes to solve Eikonal equations for the distance function of a Riemannian manifold numerically on uniform grids. He focuses on a type of Hamiltonian formulation of the length functional minimization definition of geodesics. Then the fast marching algorithm, which is a generalization of Dijkstra's minimal path method, is used to solve these equations. These techniques work particularly well for complicated geometries which exhibit a high degree of anisotropy. We utilize them below to compute distances between GEV and generalized Pareto models.

5. APPLICATIONS

We now proceed to consider three applications of the Fisher distance. First, we compare the Kullback Leibler divergence to the Fisher distance in the case of the normal distribution model to develop intuition for differences in these similarity measures. Second, we consider a nearest neighbor clustering application where we initially fit generalized Pareto distributions to equity return time series of members of the NASDAQ 100 index and consider their associated loss distributions. We then consider a single stock, AAPL, and identify which securities are nearest AAPL in a Fisher distance sense; thus finding the stocks whose risk profiles most closely resemble that of AAPL. Finally, we extend this idea to a hierarchical clustering application. Instead of focusing on a single stock, we iteratively cluster groups of stocks with a bottom-up hierarchical clustering algorithm that utilizes the Fisher distance as a similarity measure between their pairwise loss distributions along with Ward's linkage criteria for measuring the distance between clusters of stocks. This results in a risk focused hierarchical grouping of stocks.

5.1. Comparison between KL divergence and Fisher Distance. We first examine differences between the Kullback Leibler (KL) divergence, which is widely used as a similarity measure between probability distributions, and the Fisher information distance in the case of a univariate normal distribution through an empirical example. A related comparison was developed in [18] which we extend by considering an empirical example focused on equity returns from NASDAQ 100 stocks. To construct the dataset for this example, we download daily historical index components and end of day closing price data for the NASDAQ 100 index from Bloomberg between 2010-01-01 and 2015-12-31 and select stocks that remained in the index over this entire timeframe. Data was downloaded with Bloomberg's Python API package blpapi. We forward fill missing data at the price level so that a stock whose price was filled for a single day will have a corresponding zero daily return. We note that the price data is split and dividend adjusted by Bloomberg and that, with a few exceptions, each time series is fully populated over the timeframe being considered.

Our aim is to study differences between the Fisher distance and widely used KL divergence similarity measure. It is known that these two distance measures agree to second order when measuring the disparity between infinitesimally close distributions [28, 33]; however, significant differences arise on larger distance scales. Part of this comparison will examine how these distances differ when applied to the maximum likelihood parameters of each stock in the NASDAQ100 dataset. In particular, we fit the distribution to each of the return series using the unbiased MLE estimators for a normal distribution.

The general KL divergence for two probability distributions f, g is defined by

$$(33) \quad D_{KL}(f||g) = \int_{\mathbb{R}} f(x) \log \frac{f(x)}{g(x)} dx,$$

which for a univariate Gaussian distribution can be calculated explicitly

$$(34) \quad D_{KL}((\mu_1, \sigma_1) || (\mu_2, \sigma_2)) = \ln \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

This is not a distance function as it is neither symmetric nor satisfies the triangle inequality. However, it has been widely used as a similarity measure between distributions, in particular, in information theory applications [56]. We first provide a graphical comparison in Figure 1 for pairwise distances between all models in this NASDAQ 100 dataset as well as a visualization of the distance functions to a standard normal distribution of both similarity measures. Here, we compute all pairwise distances between the maximum likelihood model parameters of the 107 stocks which remained in the index over the timeframe of consideration. In the left most plot of Figure 1, we display histograms of pairwise distances for both similarity measures. The right tail of the KL divergence histograms was truncated at an upper limit of 3.0; however, we note that the KL divergence has several outlier distances. In particular, 13 values greater than 10.0 with

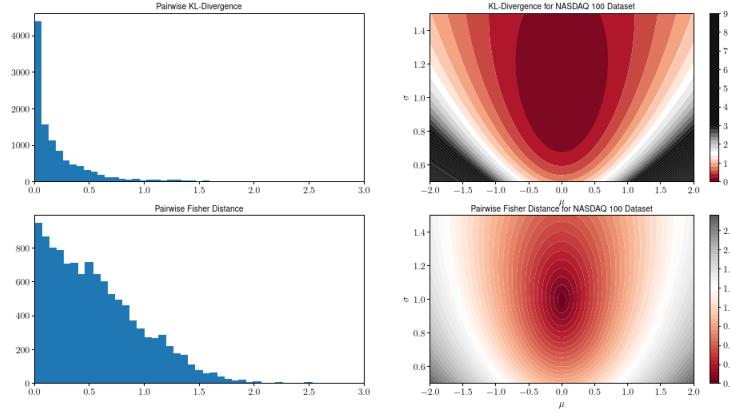


FIGURE 1. Comparison between pairwise KL-Divergence and Fisher information metric values for NASDAQ 100 parameters and distance functions to a $\mathcal{N}(0,1)$ distribution. Note that the KL divergence concentrates a number of distance values near zero and also has several large outliers whereas the Fisher distance distribution decreases roughly linearly for increasing distance.

a maximum value of 19.2 were observed. The pairwise Fisher distance histogram contains all data points, and by comparison, the maximum value is 2.7. Next, note that the KL divergence histogram has roughly an exponential decay whereas the Fisher distance histogram declines linearly. The KL divergence tends to concentrate pairwise distances near zero values when compared with the Fisher distance; this provides an indication that the Fisher distance may distinguish subtle differences in the return distributions between the equity series more strongly than the KL divergence. In addition, the greater dispersion of Fisher distances, in relation to those corresponding to the KL divergence, motivates its utilization in clustering algorithms. Specifically, one would expect a more informative cluster structure using the Fisher distance as opposed to having many data points tightly grouped together in the case of the KL divergence.

In the right two plots of Figure 1, we display contour plots that measure the similarity between all normal models in the domain and a standard $\mathcal{N}(0,1)$ distribution, i.e. we plot level sets of the function $f(\mu, \sigma) = d((0,1), (\mu, \sigma))$ for each (μ, σ) in the domain of consideration. Here we can see that the KL divergence initially concentrates around zero for models close to the standard Gaussian, but rapidly expands for further away models which have increasingly large values in comparison to the Fisher distance. In addition, the Fisher distance has slower growth and is more granular for models near the standard Gaussian.

5.2. Generalized Pareto Nearest Neighbor Example. We now consider an application of identifying the nearest neighbors of a given stock based on the Fisher distance between the loss distribution of the stock and the securities to which it is being compared. We use the generalized Pareto distribution as a model for the loss distributions of these stocks.

There are many ways to fit a generalized Pareto distribution to data, c.f. [32] for a comparison of a subset of such methods. Maximum likelihood estimation is notoriously difficult as the likelihood function is undefined for portions of the parameter domain. Multiple alternative fitting methods for the generalized Pareto distribution have been developed [8, 9, 31, 35, 41] which are generally more stable and robust when compared with typical estimation techniques such as the maximum likelihood estimation or the method of moments. Although any of these fitting methods may be used in this application, since our focus is model comparison and not estimation, we use a hybrid estimation procedure developed in [56] based on minimizing the Anderson-Darling statistic of this model which is both straightforward to implement and yields robust and intuitive results which we now briefly describe.

Electronic copy available at: <https://ssrn.com/abstract=3182914>

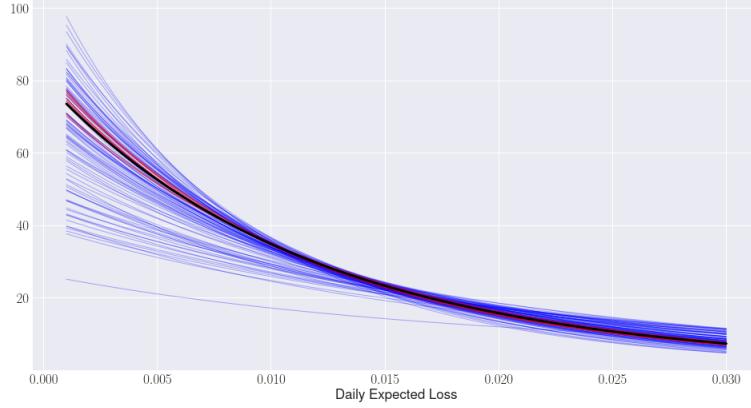


FIGURE 4. Generalized Pareto distribution fits for APPL (black), its ten nearest neighbors (red), and the remaining NASDAQ 100 stocks (blue).

to AAPL with respect to the Fisher distance. The remainder of model distributions are displayed in blue. This graph demonstrates that the Fisher distance groups stocks according to similarities in their distribution functional forms which was the original design intention of identifying stocks with common tail risk behavior.

5.3. Application to Clustering Based on Worst Annual Loss for S&P 500 Stocks. Next, we expand upon the nearest neighbor example by considering a clustering application using the Fisher distance based on an example of estimating the worst daily loss over an annual period provided in [29]. This application stems from the Fisher-Tippett-Gnedenko theorem [22, 27] which roughly states that for a set of independent identically distributed samples x_i from a probability distribution, the maximum $M_m = \max(\{x_1, \dots, x_n\})$ converges to a generalized extreme value distribution modulo a shift and scale sequence. Our application uses the GEV distribution as a model for the distribution of the maximum return of the loss distribution (worse single daily loss) for a series of stock returns.

We expand the number of securities in this application by extracting the components of the S&P 500 as of 12-31-2016 and keep stocks that remained in the index over the prior thirty years since 1-1-1987. For each stock, we find its minimum daily return over the prior calendar year and fit a three dimensional GEV distribution with shift and scale parameters using maximum likelihood estimation to each set of thirty maximum loss values. The result is a set of 272 GEV model parameters which estimate the distribution of the maximum yearly loss for each stock.

Our aim will be to apply a hierarchical clustering method to these models in order to group stocks by their worst annual losses. Specifically, we use an agglomerative clustering method which requires two inputs. First, one must specify a notion of distance between two objects being clustered. In our example, these objects correspond to the max loss distributions of distinct stocks and the distance between them is given by the Fisher distance. Second, one must specify a notion of distance between clusters consisting of multiple points called the linkage function. We use Ward's linkage which is designed to minimize the intra-cluster variance in each fusion step of the method [44, 57].

This clustering technique consists of initially grouping together the two securities with the smallest Fisher distance and then iteratively clustering the closest two security/cluster pairs until all distributions have been grouped into a single cluster. One way of visualizing this clustering algorithm is through the dendrogram of the clustering procedure, portions of which we display in Figure 5.

Several dendrogram subclusters contain stocks belonging to the same sector. For example, in the upper left plot, the four stocks WMT, HRL, BMS, and ABT are part of the large right subcluster which is itself a subcluster of the entire group. In addition, we find that some clusters are composed of companies with similar businesses such as DE, SWK, and LEG which concentrate in the design and manufacturing of engineered

Partial Mutual Information Analysis of Financial Networks

Then **partial mutual information** $I(X, Y|Z)$ denotes the part of mutual information $I(X, Y)$ that is not in Z and is defined as:

$$I(X, Y|Z) = H(X, Z) + H(Y, Z) - H(Z) - H(X, Y, Z). \quad (5)$$

PMI is symmetric so that $I(X, Y|Z) = I(Y, X|Z)$ and $0 \leq I(X, Y|Z)$. MI and PMI are only equal to 0 when X and Y are strictly independent.

To estimate PMI we need an estimator of Shannon's entropy. There is an abundance of estimators [43–48], in this study we use the Schurmann–Grassberger estimate of the entropy of a Dirichlet probability distribution, which is thought to be the best choice outside very specific conditions (particularly small samples) [49]. The Schurmann–Grassberger estimator is a Bayesian parametric procedure which assumes samples distributed following a Dirichlet distribution:

$$\hat{H}(X) = \frac{1}{m + |\chi|N} \sum_{x \in \chi} (\#(x) + N)(\psi(m + |\chi|N + 1) - \psi(\#(x) + N + 1)), \quad (6)$$

where $\#(x)$ is the number of data points having value x , $|\chi|$ is the number of bins from the discretization step, m is the sample size, and $\psi(z) = d \ln \Gamma(z)/dz$ is the digamma function. The Schurmann–Grassberger estimator assumes $N = 1/|\chi|$ as the prior [50].

For comparison we have also used Pearson's correlation and partial correlation, and networks based on it, as defined in [39].

We now turn to the networks we are creating based on partial mutual information. First we refine the structural view of the market offered by the networks based on correlation. **It is refined first by swapping correlation with mutual information, which adds nonlinearity.** We may further refine this by removing the mediated parts of the interrelations between financial instruments by controlling for a third variable with partial mutual information. We take the minimum of the PMI calculated controlling for all other financial instruments, **to show only the part of the mutual information not contained in other studied time series.** Thus taking in mind the standard mutual information based metric [36, 37] the distance used for the network topology is defined as:

$$d(X, Y) = H(X, Y) - \min_{Z \neq X, Y} I(X, Y|Z). \quad (7)$$

On this basis we may create a network with topological restraints of our choosing, we calculate minimal spanning trees and planar maximally filtered graphs which we call PMIMST and PMIPMFG. These trees and planar graphs are created based on a list of $d(X, Y)$ sorted in increasing order. By starting from the first entry of the list, we add a corresponding link if and only if the resulting network is still a tree or a forest (PMIMST) or is still planar, i.e. it can be drawn on the surface of a sphere without link crossing (PMIPMFG).

Second we may refine the partial correlation planar graph as defined in [38] by directly swapping partial

correlation with partial mutual information. For this purpose we need a measure of MI influence or influence of an element Z on the pair of elements X and Y . This quantity is large only when a significant fraction of the PM $I(X, Y)$ can be explained in terms of Z . This measure is defined as:

$$d(X, Y|Z) = I(X, Y) - I(X, Y|Z). \quad (8)$$

We define the average influence $d(X|Z)$ of element Z on the MIs between element X and all the other elements in the system as

$$d(X|Z) = \langle d(X, Y|Z) \rangle_{Y \neq X, Z}. \quad (9)$$

In order to construct a planar graph based on PMI we list the $N(N - 1)$ values of the average MI influence $d(X|Z)$ in decreasing order. The construction protocol of the network begins by considering an empty network with **N vertices**. By starting from the first entry of the list we put a link between them **if and only if the resulting network is still planar**. Similarly we can create a related minimal spanning tree by only adding a link if and only if the resulting network is still a tree or a forest.

Here we note that instead of filtering the information by network topology as we are doing above we may also filter the information by using a threshold to find which links should be entered into the network. Finding an appropriate threshold is not trivial however, but one can test for statistical significance based on the fact that $I(X, Y|Z)$ when X and Y are independent conditioned on Z follows a Gamma distribution with shape parameter $\kappa = |Z|(|X| - 1)(|Y| - 1)/2$ and scale parameter $\Theta = 1/N$ [51].

We also note a few areas in which this methodology can be extended. First we note that **one could add market (index) as a second variable which is being controlled in PMI, thus eliminating the effects of general market trends on the relationships between financial instruments.** Further we note that another approach to extending the analysis may be taken using **transfer entropy**, which is a measure closely related to partial mutual information [52]. Transfer entropy is a measure **quantifying causal information transfer between systems evolving in time, based on appropriately conditioned transition probabilities**, thus it uses time lags. Time lagged causal analysis can also be performed using correlation [53] or mutual information [37] based method for filtering similarity measures into a network of statistically-validated directed links.

3. Results

To find the networks presented above we have taken log returns for 91 securities out of 100 which constitute the NYSE 100, excluding those with incomplete data. These log returns are based on daily closing prices. The data has been downloaded from Google Finance database available at <http://www.google.com/finance/> and was up to date as of the 11th of November 2013, going 10 years back. The data is transformed in the standard way for analysing price movements, that is so

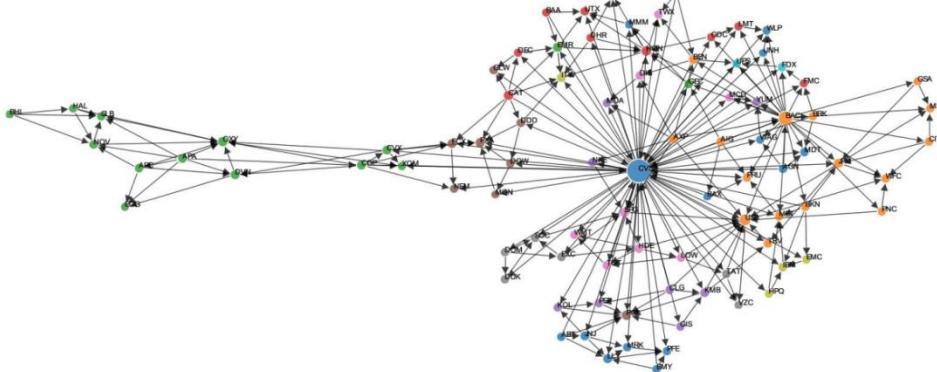


Fig. 3. Partial mutual information PMFG.

On cluster level we confirm that in all created networks the clusters have been aligned largely according to economic sectors, which means that this important characteristic of the network approach to financial markets has been preserved. Preserving this information is important because it cannot be reproduced by simulating a virtual market [57]. This can be shown numerically as the ratio of arcs between stocks of the same sector to all nodes as presented in Table. As can be seen without any topological restraints the full market has 11.58% links within sectors, but for our networks it is between 42% and 67%, thus we see that this important feature is preserved. In general we can also see that mutual information suits this task better, which further corroborates the usefulness of our approach.

TABLE
Network comparision.

Network	Tree ratio	Graph ratio	Clustering
1	62.22%	49.06%	17.60%
2	66.67%	55.81%	20.80%
3	64.44%	51.69%	14.60%
4	65.56%	54.68%	17.70%
5	48.89%	42.70%	5.30%
6	57.78%	47.19%	14.70%
Ref.	11.28%	11.28%	50.00%

On network level there is a limited possibility of investigation as most network-wide measures would be constrained by the common topological restraints we are using. Nonetheless we have calculated clustering coefficients (ratio of the number of triangles observed to the number of possible triangles in the network) for the planar graphs (there can not be any triangles in trees), which are presented in Table below. As can be seen, mutual information produces significantly more clustering than correlation. Nonetheless, partial correlation/mutual information analysis creates slightly less clustered networks, hinting that some clustering happens due to mediating noise, but also due to nonlinearity.

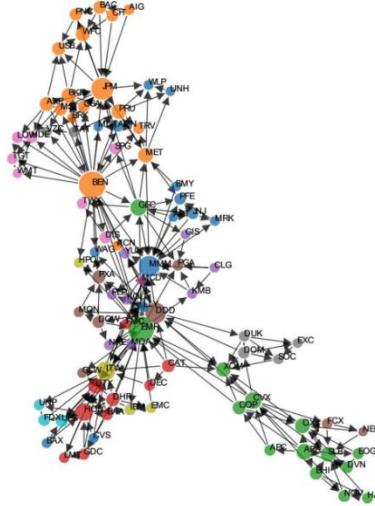


Fig. 4. Partial mutual information planar graph.

The reference values in Table are calculated for a network with no topological restraints.

Strictly as an example we have provided network number 4 in Fig. 3 and 6 in Fig. 4. The node size is based on its centrality in the network.

4. Conclusions

We have analysed a method for producing networks of financial markets based on partial mutual information and how are those different from networks based on correlation, partial correlation and mutual information. The analysis leads us to believe that mutual information should be used in network analysis of financial markets as it provides different and likely more accurate networks, and that partial mutual information may be

Causal Non-Linear Financial Networks

matrices an empirical transfer entropy matrix C is constructed using the partial mutual information of columns from A and B ,

$$C_{m,n} = I(A_m, B_n | A_n), \quad (2)$$

where $I(X, Y)$ is the mutual information between X and Y , $I(X, Y | Z)$ is the conditional (partial) mutual information [54] between X and Y conditioned on Z , and A_m denotes column m of A . Note that $C_{m,m} = 0$.

Here we show that the above is measuring generalised Granger causality. Transfer entropy is a non-parametric statistic measuring the amount of information (in Shannon's sense) transferred between two random processes (transfer entropy is directional). Transfer entropy is thus a measure of Granger causality, but a more general one, sensitive to non-linear interactions [55]. Assuming the data as presented above we can define transfer entropy as:

$$T_{m \rightarrow n} = H(B_n | A_n) - H(B_n | A_n, A_m), \quad (3)$$

where $H(X)$ is Shannon's entropy and $H(X | Y)$ denotes conditional Shannon's entropy. Transfer entropy is equivalent to a specific conditional (partial) mutual information [55]:

$$T_{m \rightarrow n} = I(A_m, B_n | A_n). \quad (4)$$

To apply the above in practice we need an estimator of entropy (mutual information can be defined in terms of entropy). We note that for easy estimation we need discrete data, and while stock returns are discrete, their resolution is much too high for practical purposes, thus we need to discretise them. For discussion of this step see below and Refs. [6, 20]. There is a large number of estimators of entropy, for details please see Refs. [56–60]. In this study we continue to use the Schurmann–Grassberger estimate of entropy, which we have applied in our previous study [61]. The Schurmann–Grassberger estimator is a Bayesian parametric procedure, which assumes samples distributed according to Dirichlet distribution:

$$\begin{aligned} \hat{H}(X) &= \frac{1}{m + |\chi|N} \\ &\sum_{x \in \chi} (\#(x) + N)(\psi(m + |\chi|N + 1) - \psi(\#(x) + N + 1)), \end{aligned} \quad (5)$$

where $\#(x)$ is the number of data points with value x , $|\chi|$ is the number of bins from the discretisation step, m is the sample size, and $\psi(z) = d \ln \Gamma(z) / dz$ is the digamma function. The Schurmann–Grassberger estimator assumes $N = 1/|\chi|$ as the prior [62].

The matrix C can be seen as a weighted adjacency matrix for a fully connected, directed graph. As stated above, such matrix needs to be filtered. To find a threshold of statistical significance Curme et al. [29] apply a shuffling technique [63]. The rows of A are shuffled repeatedly without replacement in order to create a large

number of surrogate time series. These are then validated by p -value adjusted to account for multiple comparisons. Curme et al. [29] use the conservative Bonferroni correction (p/N^2). For $N = 100$ and the unadjusted p -value equal to 0.01 it gives $0.01/100^2$, which requires the construction of 10^6 independently shuffled surrogate time series. The same can be done for methodology based on mutual information and transfer entropy. But we find the computational requirements of this procedure to be prohibitively large for studies of large networks (at least for practical, real-time applications), and financial markets usually consist of hundreds of stocks. Nonetheless, a less computationally expensive method has been presented in our previous study, without introducing very strong assumptions (here we note that the bootstrap method is, at least in principle, better, if computation time is not an issue, as it takes into account the heterogeneousness of the studied time series). It has been shown that mutual information between independent random variables (X & Y), when estimated from relative frequencies, follows a very good approximation of Gamma distribution with parameters $a = (D)/2$ and $b = 1/(N \ln 2)$ [64, 65]:

$$I(X, Y) \sim \Gamma\left(\frac{D}{2}, \frac{1}{N \ln 2}\right), \quad (6)$$

where N is the sample size and D denote the number of degrees of freedom (dependent on the alphabet used for the studied discrete time series). This has been explained in Ref. [30].

Therefore, to determine the significance of $I(A_m, B_n)$ from a sample study of length N at a significance level p , we check the condition:

$$I(A_m, B_n) \geq \Gamma_{1-p}\left(\frac{D}{2}, \frac{1}{N \ln 2}\right), \quad (7)$$

where $\Gamma_{1-p}(a, b)$ denotes the $(1 - p)$ -quantile of the Gamma distribution. The same holds for conditional mutual information, or transfer entropy, but obviously the quantiles themselves will have different values as the number of degrees of freedom is higher in transfer entropy as it is in mutual information. We note that besides bootstrapping and validating based on Gamma distribution, a threshold can also be set by the analysts based on their experience and practical needs. This being a less formal approach is not considered here. Here we also note that most of the results presented in this study would remain virtually unchanged (with the exception of the actual number of significant links) by a choice of a slightly different threshold value, thus this step is of relatively mild importance to most applications (where analysts would concentrate on the most significant links and not the ones slightly above threshold).

III. EMPIRICAL APPLICATION

To compare the results of this study with our previous study investigating lead–lag effect [30] we use the same

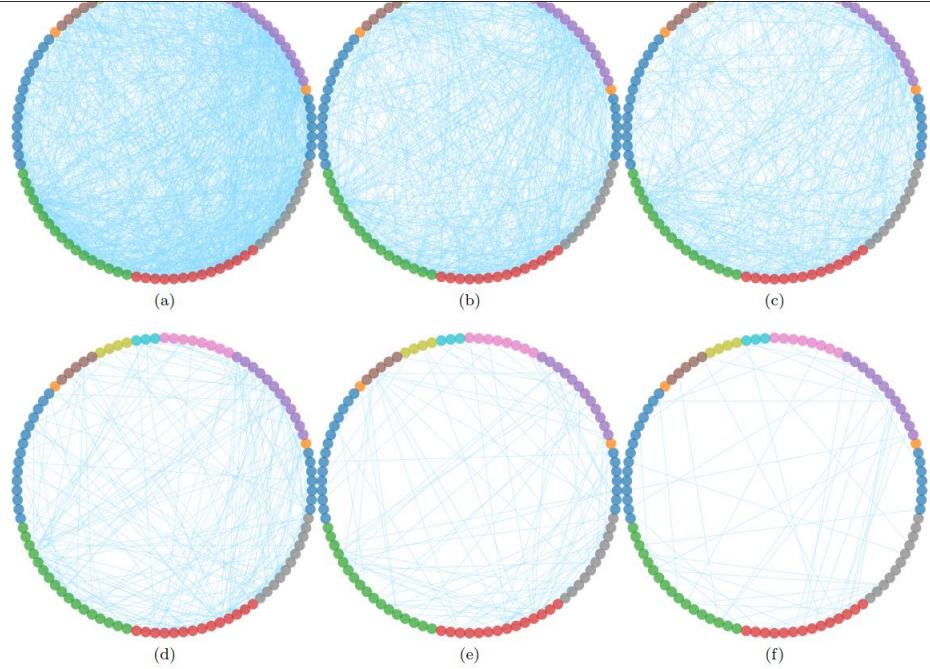


FIG. 5. Bonferroni networks of statistically significant transfer entropy between pairs of NYSE 100 stocks for time lag λ of: (a) one minute; (b) five minutes; (c) ten minutes; (d) twenty minutes; (e) thirty minutes; and (f) forty minutes. The decay in statistically significant relationships with increasing λ is clearly visible.

sectors of economic activity). We can clearly observe the decay of links with increasing time lag when looking at the networks themselves. At this point we note that for small time lags the number of statistically significant links extends the number of links that standard topological methods would provide (e.g. Planar Maximally Filtered Graph). One may therefore be tempted to use such method, instead of applying threshold, to create more cohesive networks. There is a problem with such an approach however. We have created networks based on transfer entropy, which are roughly equivalent to Partial Correlation Planar Graphs introduced in Ref. [19] (these are not relevant in themselves, thus will not be presented). In the best case scenario we get a network where only 77.08% of links are statistically significant. In most cases this number would drop to a much lower value. By adopting a topological method of creating filtered networks in this setting we would therefore be creating networks in which the amount of statistical noise would be very high. Thus we propose, as is proposed in other studies on similar topics, to use threshold for creating asynchronous financial networks.

The method of applying threshold does have a drawback however, in that the networks are not as cohesive as Minimally Spanning Trees or Planar Maximally Filtered Graphs. In other words the networks based on threshold are rarely small world networks, and rather often resemble random graphs. This in itself is not necessarily a bad thing, since we would not expect a causal structure to be the same as a static structure of financial markets. Nonetheless, in Fig. 6 we present degree distributions for Bonferroni networks at different time lags, as specified above. In degrees are presented with dots, and out degrees as crosses. All distributions are presented on log-log scales. We see that for small time lags λ the networks approximate random graphs. In other words, for small lags the causal structure is not highly cohesive (the resulting distributions aren't well approximated by theoretical fat-tailed distributions such as power law or log-normal distributions). But we also observe that with increasing time lag the degree distributions move closer to fat-tailed distributions such as log-normal distribution or even power law. We postulate that this is due to the fact that for small time lags there is a lot of statistically

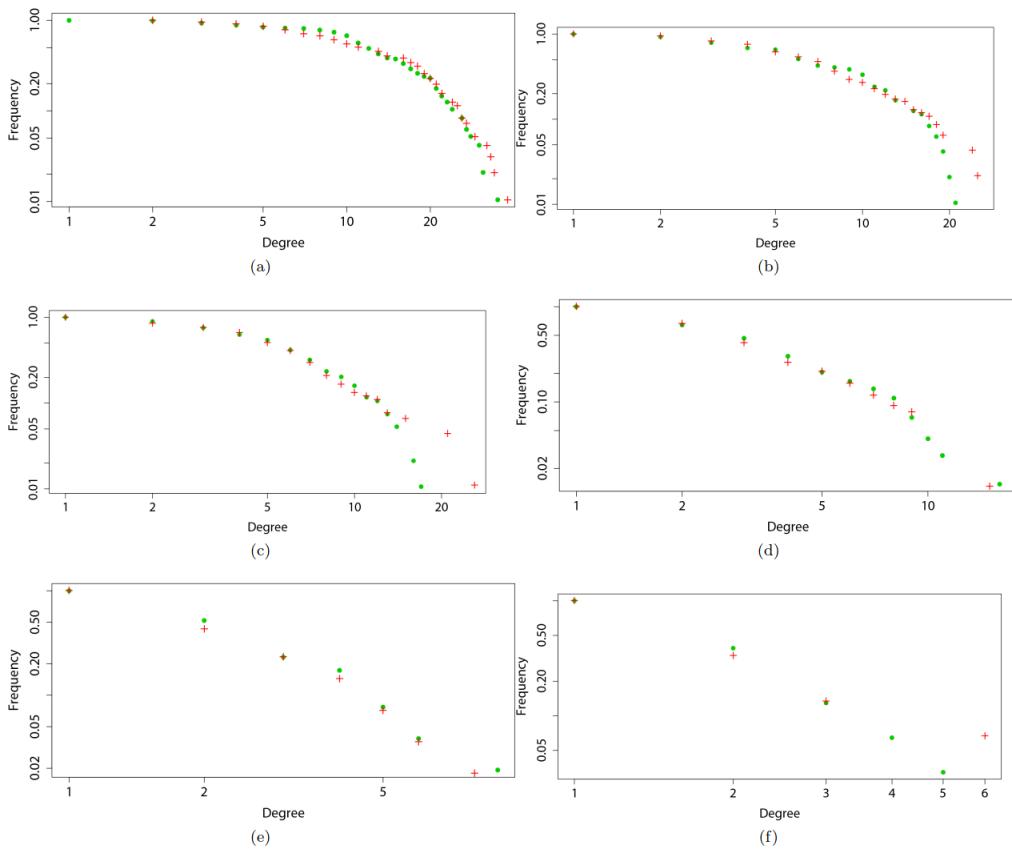


FIG. 6. Degree distributions (in degree as dots, out degree as crosses) for Bonferroni networks based on significant transfer entropy between pairs of NYSE 100 stocks for time lag λ of: (a) one minute; (b) five minutes; (c) ten minutes; (d) twenty minutes; (e) thirty minutes; and (f) forty minutes. The distributions for small values of λ are close to random graphs, but move close to scale free networks for large values of λ .

significant links due to the synchronisation in the market, and these tend to be scattered around the market. The causal relationships which operate at higher time lags are usually strongly grounded in some economic basis, and thus create a more cohesive structure, similar to the standard synchronous scale free financial networks.

While we do not observe a scale free characteristic in the degree distributions of Bonferroni networks, we do observe a power law in the statistically significant transfer entropy values themselves. The distributions of these for the above-mentioned time lags are presented in Fig. 7 on log-log scales. We see that for all time lags the distribution is well approximated by either a power law or a log-normal distribution (both fitted and plotted). Thus we see that within the statistically significant causal re-

lationship in New York's market only a handful are very strong, while most of them are only slightly above the threshold of statistical significance. This is desirable, as analyst may concentrate on the highly significant links, which they can be sure are not included in the networks due to badly chosen threshold. This also underscores that the choice of the threshold is not as important as it may appear at first glance, as any reasonable value will still retain the most important information within the constructed networks.

Finally, in Fig. 8 we present the percentage of pairs belonging to the same economic sector for all studied time lags λ within two groups: all 9506 studied pairs (dashed line) and only the statistically significant pairs (solid line). We see that, unlike in the standard analysis