

Study of the applicability of an itemset-based portfolio planner in a multi-market context

Luca Cagliero

Dipartimento di Automatica e Informatica
Politecnico di Torino
Turin, Italy
luca.cagliero@polito.it

Paolo Garza

Dipartimento di Automatica e Informatica
Politecnico di Torino
Turin, Italy
paolo.garza@polito.it

ABSTRACT

Planning stock portfolios for long-term investments is a well-known financial problem. Many data mining and machine learning strategies have been proposed to automatically predict the set of uncorrelated stocks maximizing long-term portfolio returns. Among others, the use of scalable itemset-based strategies has recently been studied. Potentially, they can analyze large sets of historical prices corresponding to thousands of stocks in the worldwide market indexes. However, the current studies are still limited to single markets.

This paper investigates the applicability of itemset-based strategies for planning stock portfolios in a multi-market context. Scaling the analyses towards multi-market scenarios poses a number of research questions, among which the choice of the diversification strategy, the influence of inter-market correlations among stock prices, and the profitability of multi-market strategies compared to single-market ones. This paper aims at answering to the aforesaid questions by considering a state-of-the-art itemset-based approach.

The experimental results show that itemset-based strategies focus the generated portfolios on the outperforming markets. Furthermore, the performance of multi-market strategies with sector-based diversification is on average superior or comparable to single-market ones.

1 INTRODUCTION

Forecasting the stock markets is a well-known financial problem. It entails predicting the future prices of a set of stocks to drive investments in the short-, medium-, or long-term. Predictions are commonly driven by fundamental or technical analyses [8]. The former studies analyze the overall state of a company or a business (e.g., earnings, production, manufacturing), whereas the latter analyze the historical stock prices, which are assumed to reflect all the external influences. Technical analyses often consider both statistics-based indicators, computed on the sampled stock prices, and graphical patterns, recognized from the price time series, that are likely to be related to specific trends [20].

In this work, we focus on the analysis of the historical stock prices to make long-term predictions. The aim is to generate a portfolio consisting of a subset of market stocks whose prices are likely to increase. To spread bets across multiple assets, thus minimizing the losses in case forecasts turn out to be wrong, portfolios are asked to be *diversified*, i.e., they should comprise stocks from different sectors, markets, or geographical areas [4, 14].

In recent years, the diffusion of machine learning and data mining techniques has prompted the financial sector and the

research community to investigate their application to solve the portfolio generation problem. For example, classification and regression algorithms such as Neural Networks [15, 24], Decision trees [2, 18], and Support Vector Machines [5, 12] have been exploited to predict the future stock directions and prices, respectively, based on the values of multiple dependent variables. Alternative strategies entail the use of

- (i) Time series analyses, to pinpoint significant temporal trends in continuous stock signals [9, 11, 13, 25],
- (ii) Clustering algorithms, to group stocks characterized by similar behaviors [16, 21],
- (iii) Pattern recognition techniques, to recognize graphical patterns coming from technical analyses [17], and
- (iv) Particle swarm optimization and evolutionary algorithms, to identify the stocks that maximize a given objective function [1, 7].

Itemset mining is an exploratory data mining technique that focuses on discovering recurrent co-occurrences among items in large transactional dataset [3]. For example, let us consider a transactional dataset collecting the baskets of the customers of a market, where each transaction (basket) consists of a set of distinct items. Frequent itemset mining algorithms have been exploited to discover combinations of items that are frequently purchased together. Since items may have different importance within the analyzed datasets (e.g., different prices and purchased amounts) their occurrences in each transaction can be weighted [19].

Recently, in [6] a first attempt to apply itemset mining techniques to generate diversified stock portfolios has been made. Stocks are represented as distinct items in the dataset. A transactional dataset collects the historical stock prices within a time range. Each transaction corresponds to a distinct timestamp within the given time range and contains all the quoted stocks weighted by their price at the corresponding timestamp. According to the data model described above, itemsets represent candidate stock portfolios consisting of sets of stocks of arbitrary size. In [6] the most interesting itemsets are generated and ranked according to the average return of the contained stocks as well to their level of diversification in the portfolio.

The main advantages of itemset-based approaches are (i) the interpretability of the generated model and (ii) the scalability of the extraction algorithms, which can be applied to very large datasets [22]. On the one hand, the interpretability of the mined itemsets allows domain experts to manually explore the top ranked itemsets to make appropriate decisions. On the other hand, the scalability of the itemset mining process makes the portfolio generation process portable to multi-market domains. The algorithm can analyze large stock datasets acquired from multiple markets and automatically recommend diversified worldwide investments with limited human effort. However, to the best of our knowledge, the application of itemset-based strategies in multi-market contexts has not been investigated yet.

© 2018 Copyright held by the owner/author(s). Published in the Workshop Proceedings of the EDBT/ICDT 2018 Joint Conference (March 26, 2018, Vienna, Austria) on CEUR-WS.org (ISSN 1613-0073). Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0.

2 CONTRIBUTION

This paper investigates the applicability of itemset-based strategies for planning diversified stock portfolios in a multi-market context. Extending the scope of the stock data analysis from single markets to multiple ones poses the following research questions:

Choice of diversification strategy. Stocks can be categorized based on different strategies, such as the industrial sector of the underlying company, the market index of the stock, the nationality of the company, or the country/continent associated with the market index. These categorizations can be exploited to diversify investments across uncorrelated assets. In multi-market contexts, the choice of the diversification strategy is not trivial, as it could relevantly affect the performance of the stock portfolio planner. The research questions we would like to address in this study can be formulated as follows: **Which type of diversification strategy better preserves the portfolio profits? Which type of diversification strategy allows optimally spreading investments across multiple assets?**

Influence of inter-market stock correlations. Studying the correlation between the prices of multiple stocks is crucial for professional traders and private investors to take appropriate decisions. However, analyzing the influence between the stocks belonging to multiple markets is potentially challenging, because the number of stocks indexed in worldwide markets is very large. Itemset-based approaches [6] allow domain experts to set the desired levels of average return and diversification according to the chosen stock categorization. Under these constraints, a set of candidate portfolios is generated. The top ranked portfolios include the stocks with maximal average return and with a diversification level at least equal to the set least diversification level. Therefore, the outperforming markets are likely to be over-weighted, while the stocks indexed in under-performing markets are likely to be under-weighted. The research question we would like to address in this study can be formulated as follows: **Does the majority of the portfolio stocks belong to outperforming markets? Are the stocks in the portfolio correlated in terms of membership index?**

Comparison between different scenarios. Extending the scope of the analysis towards multiple markets gives professional traders and private investors new opportunities of investment on foreign markets. Considering a larger number of considered stocks not only simplifies the process of diversification of the investments, but also allows traders to move investments towards most profitable markets. However, a quantitative evaluation of the benefits for itemset-based approaches of considering multiple markets at the same time compared to single-market analyses is still missing. The research question we would like to address in this study can be formulated as follows: **Are the portfolios generated from multiple markets more profitable than those generated from single markets?**

In this study we investigated the use of different diversification strategies in multi-market scenarios to gain insights into the effectiveness of itemset-based strategies on large stock data. Furthermore, we analyzed the generated portfolios to understand to what extent inter-market stock correlations are considered in the recommended portfolios. Finally, we empirically compared the performance of single- and multi-market recommendations in different scenarios.

The rest of the paper is organized as follows. Section 3 summarizes the main steps of the diversified stock portfolio planner [6]. Section 4 describes the experimental design. Sections 5, 6, 7 discuss the choice of the diversification strategy, the influence of inter-market stock correlations, and the comparison between multiple and single market strategies, respectively. Finally, Section 8 draws conclusions and discusses the future research perspectives of this work.

3 THE DIVERSIFIED STOCK PORTFOLIO PLANNER

DISPLAN (Diversified stock portfolio planner) [6] is an itemset-based strategy for generating diversified stock portfolios based on the analysis of historical stock prices. It relies on the following steps:

Stock data collection and preprocessing. This step focuses on crawling historical stock prices and collecting them into a unique dataset. It takes as input a list of stocks and a time range. It acquires the daily closing prices of all the considered stocks and stores them into a weighted transactional dataset. Each row in the dataset (called *transaction*) corresponds to a different timestamp and contains the prices of all the considered stocks at the corresponding timestamp. Each pair $\langle \text{stock}, \text{price} \rangle$ occurring in the dataset is denoted as *weighted item*.

Weighted itemset mining. This step analyzes the correlations between stock prices based on weighted itemset mining techniques [19]. It takes as input the weighted transactional dataset prepared at the previous step and a taxonomy aggregating stocks into higher-level categories. For example, to each stock the corresponding industrial sector can be assigned.

It extracts interesting patterns, called frequent weighted itemsets, from the weighted transactional dataset. A weighted itemset is a set of stocks of arbitrary length, which represents a candidate stock portfolio. The extracted weighted itemsets satisfy the following properties:

- (i) The average daily return of all the stocks in the itemset is above a given minimum return threshold *minret*.
- (ii) The percentage of stocks belonging to different categories (according to the input taxonomy) is above a given diversification threshold *mindiv*.

Portfolio generation. This step analyzes the extracted itemsets to identify the best candidate stock portfolios satisfying all the user requirements. To make the extracted patterns promptly usable by investors for stock portfolio planning the mined itemsets are first ranked in order of (i) decreasing length (i.e., number of contained stocks) and (ii) average daily return. The top ranked itemsets are deemed as the most appropriate hints for buy-and-hold (long-term) investors. More specifically, the itemsets containing the maximal number of stocks are selected as best candidate portfolios, because they satisfy all user requirements (by construction) and contain the maximal number of stocks thus allowing investors to spread their bets over the largest number of different assets. In case of ties, the itemset with maximal least average return is considered as the best candidate stock portfolio because it achieved maximal profit on historical data. On equal terms (i.e., same length and average daily return) the analyst is asked to decide which itemset is deemed to be the most appropriate stock portfolio to consider based on his personal judgment and experience.

4 EXPERIMENTAL DESIGN

We analyzed stock data acquired by means of the Yahoo! Finance APIs [23]. To crawl data, we performed several API requests to retrieve the closing stock prices of several stocks from different market indexes. Each request produces a different stock dataset, which consists of the closing prices of the requested stocks within the considered time period sampled at the desired frequency. In our analyses, we considered the daily closing prices of the stocks in two representative years, i.e., 2008 and 2013. Year 2008 is representative of an unfavorable condition for worldwide financial markets, i.e., the rise of the global financial crisis, whereas year 2013 represents a favorable market condition, i.e., the boom of U.S. markets. Analyzing opposite market scenarios allows us to perform a fair assessment of the portfolio generator with different settings.

To study the performance of the itemset-based portfolio planner in a multi-market context, we classified market indexes based on the corresponding opening timezone as follows: (i) Europe, with approximate opening times from 9am to 5.30pm CET. (ii) Asia and Oceania, with approximate opening times from 2am to 9.30am CET. (iii) North and South America: with approximate opening times from 3pm to 10pm CET. As European indexes we considered the following ones: the Bruxelles Stock exchange (BEL 20) (20 stocks), the Paris market (CAC 40) (40 stocks), the London Financial Times Stock Exchange (FTSE 100) (98 stocks), the Italian stock exchange (FTSE MIB 40) (40 stocks), the General Athens Composite Index (GD) (59 stocks), the HDAX Deutscher Aktien index (GDAXI) (109 stocks), the OMX Stockholm 30 (OMX) (25 stocks), and Oslo Bors All Share Index (OSEAX) (127 stocks). As Asian and Oceania Indexes we considered the following ones: the BSE Sensex (BSESN) Based on the Bombay Stock Exchange (India), the FTSE Straits Times Index (STI) based on the Singapore Exchange (29 stocks), the Hang Seng Index (HSI) based on the Hong Kong Stock Exchange (50 stocks), the NIFTY 50 (NSEI) consisting of companies listed on the Bombay Stock Exchange (BSE-India) (50 stocks), the NZX 50 (NZ50) based on the New Zealand Stock Exchange (NZSX) (39 stocks), the S&P/ASX 200 (AXJO) based on the Australian Securities Exchange from Standard & Poor's (199 stocks), and the Taiwan Capitalization Weighted Stock Index TAIEX Index (TWII) (898 stocks). As North and South America Indexes we considered the following ones: the Brazil Broad-Based Index (IBRA) (116 stocks), the US Dow Jones Industrial Average (DJI) based on the New York Stock Exchange (30 stocks), the Brazilian Indice Bovespa (BVSP) (63 stocks), the IPC Index (MXX) based on the Mexican Stock Exchange (34 stocks), the IVBX 2 Brazilian Index (IVBX) (50 stocks), the Merval Index (MERV) based on the Stock Exchange of Buenos Aires (Argentina) (12 stocks), the US Nasdaq Stock Market index NASDAQ-100 (NDX), the US S&P 500 (GSPC) (502 stocks), and the Canadian S&P/TSX Venture Composite Index (SPCDNX) (338 stocks). The Europe timezone comprises 518 stocks, the Asia and Oceania timezone 1118 stocks, while the North and South America 1147. Hereafter we will denote as *market-based* categorization the stock categorization based on the considered indexes. We considered also a *sector-based* stock categorization according to the Industry Classification Benchmark (<http://www.icbenchmark.com/>).

To simulate long-term stock investments, we applied the following procedure: (i) We trained the itemset-based model and generated the diversified stock portfolio on the first 7-month time period. (ii) We tested the model by virtually buying the stocks at the beginning of August and selling the whole portfolios in

the following year. We varied the selling date in the date range between August, 1st 2013 and August, 1st 2014.

We simulated both *long* and *short* selling investing positions. *Long* selling entail buying the stocks because its price is likely to increase thus yielding a profit in case the price has increased when the stock is sold. Conversely, *short* selling is the practice of selling stocks that are not currently owned, and subsequently repurchasing them at the end of the investment. If the price decreases, the short seller profits, since the cost of (re)purchase is less than the proceeds received upon the initial (short) sale. Conversely, the short selling position closes out at a loss if the stock price rises prior to repurchase [10].

In the following section we will analyze also the impact of the algorithm parameters on the quality of the training models. Notice that the portfolio profits are lower bound estimates of the actual profits as stocks may be sold one by one rather than altogether and the investments can be reconsidered during the whole period (not only at the end). Furthermore, payoffs produced by intermediate sells can be reinvested during the same time period. Finally, transactions costs and local taxes and fees were not considered.

5 CHOICE OF THE DIVERSIFICATION STRATEGY

We compared the performance of the itemset-based stock portfolio planner (hereafter denoted as DISPLAN for the sake of brevity) by using two alternative diversification strategies:

- (i) a *market-based* strategy, where stocks are picked from different indexes to limit intra-market stock correlations (independently of the industrial sector of the considered stocks), and
- (ii) a *sector-based* strategy, where stocks picked from different sectors (independently of the underlying market index).

As representative examples, in Figures 1, 2, and 3 we plotted the relative returns achieved in year 2013 by the portfolio generated by DISPLAN on the markets of the Asia and Oceania, Europe, and North and South America timezones with both market- and sector-based diversification. For all the considered timezones we reported the results achieved by using long selling positions (see Section 4) and by setting the minimum diversification threshold to 70% (i.e., at least seven stocks out of 10 must belong to different categories). The minimum return threshold values enforced to discard non-profitable sets of stocks are 11%, 9%, and 12%, respectively. To compare DISPLAN performance with that of the considered indexes for each configuration, we plotted also the percentage variation of the benchmark indexes as well as those of an aggregate index consisting of all the underlying indexes the same timezone.

The achieved results show that applying a sector-based diversification yields significantly higher profits compared to applying a market-based strategy (e.g., in the Asia and Oceania timezone on July, 1 2014 the average variation of the sector-based strategy is 120% vs. the 18% of the market-based one). The motivations behind the achieved results are the inherent sparsity of the market-based categorization compared to the sector-based one and the stronger influence of sector-driven stock price movements. Specifically, by applying sector-based diversification within each category the algorithm can choose among a quite large number of stocks. Among the per-sector candidate stocks, some of them are likely to outperform the benchmarks. Therefore, the stocks that under-perform the corresponding sector can be discarded. Conversely, to satisfy the least diversification level, the

market-based strategy could be forced to pick stocks that underperform the corresponding markets index as well. Furthermore, the number of candidate indexes per timezone is still limited (8 for Europe and Asia and Oceania, 10 for South and North America). Thus, in large portfolios to achieve high diversification levels the stocks belonging to low-performing indexes cannot be neglected. On the other hand, the intra-sector correlations among stock prices appear to be stronger than intra-market ones. For example, a drop of the oil price negatively influences all the correlated stocks independently of the market index. In summary, based on the achieved results it turns out that considering a stock categorization based on sectors prevents the DISPLAN algorithm from making inappropriate decisions.

6 ANALYSIS OF INTER-MARKET STOCK CORRELATIONS

The stock portfolios generated by the DISPLAN algorithm may include stocks belonging to multiple markets. Hence, it is interesting to investigate how the inter-market correlations among stock prices could affect the performance of the DISPLAN algorithm.

Figure 4 shows, as representative case study, the relative returns achieved in year 2013 by the portfolio generated by DISPLAN (with long selling position) in both multi- and single-market scenarios. Specifically, as representative study, we reported the performance of the portfolio generated by DISPLAN from the analysis of the stock data related to all the markets in the Asia and Oceania timezone as well as the performance of the portfolios generated by DISPLAN from each index in the same timezone. Notice that since the AXJO, HSI, and STI indexes did not produce any single-market portfolios satisfying the minimum return and diversification constraints, the corresponding curves were omitted.

By enforcing a minimum diversification level among sectors of 70% ($mindiv=70\%$), the multi-market portfolio consists of the same stocks selected by the best performing single-market portfolio, i.e., the one generated by the TWII index of Taiwan (see Figure 4(a)). Conversely, while setting the highest possible value of sector-based diversification level ($mindiv=100\%$), all the stocks in the portfolios must belong to a different sector. The maximally diversified portfolio differs from the former one because a stock from the AX index (JBH) was selected. The sector of stock JBH under-performed the benchmark index in year 2013. However, stock JBH performed better than the benchmark sector (approximately +15%). For this reason, despite the higher diversification of the new portfolio its relative returns are still relatively high. On the contrary, by setting the diversification threshold to its maximal value the portfolio generated from the single-market TWII index significantly decreases its relative returns, because the added stocks under-performed the benchmark sector index (see Figure 4(b)).

7 COMPARISON BETWEEN MULTI- AND SINGLE-MARKET STRATEGIES

We compared multi- and single-market strategies to assess the applicability of the DISPLAN system in a multi-market scenario. The results, which were summarized in Figure 4 for a representative case study (Asia and Oceania, long selling position, year 2013), show that the multi-market approach performed better than most single-market strategies while it performs as good as the best performing single-market one. Therefore, applying the itemset-based portfolio generation strategy is particularly

appealing, as it allows us to diversify investments across multiple market indexes without significantly degrading the portfolio returns.

Figure 5 shows the performance of the DISPLAN algorithms by setting a relatively high minimum return threshold (16%). The results show that the multi-market strategy performed as good as the best single-market strategy (the one corresponding with the best performing index). Conversely, many single-market strategies appear to be less effective because few candidate stocks are selected. The motivation behind is that, given a large number of candidate stocks from multiple markets, the likelihood that a set of highly profitable stocks diversified over sectors is found is higher. The more stocks the algorithm can analyze, the most likely a profitable and diversified stock portfolio can be discovered. To avoid data overfitting the number of stock data samples should be at least on the order of the number of analyzed stocks.

8 CONCLUSIONS AND FUTURE WORK

In this paper, the application of itemset-based approaches to generating diversified stock portfolios in a multi-market scenario has been studied. Given a stock categorization and a dataset collecting the historical prices of a potentially large set of stocks, itemsets representing profitable yet diversified stock portfolios can be automatically extracted and recommended to investors, professional and not. The scalability of itemset-based techniques prompted their application in a multi-market scenario, where the following issues have been addressed.

(i) The choice of the diversification strategy is not immediate, because in multi-market contexts investors could spread bets across either markets or sectors. Based on the achieved results, sector-based diversification yielded significantly better results due to the good balancing between the stocks across sectors.

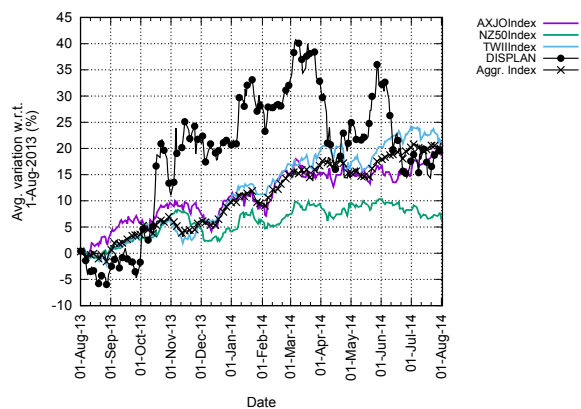
(ii) The inter-market correlations among stocks are properly handled by the itemset-based strategy, as the most profitable stocks are selected independently of the underlying market index.

(iii) Multi-market strategies performed better than or as good as single-market ones in most of the performed experiments.

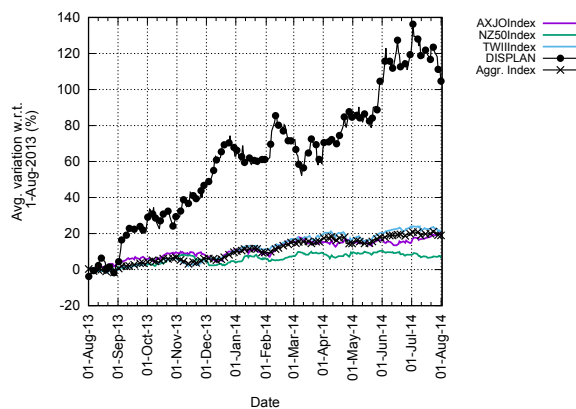
Future works will entail applying itemset-based strategies on datasets collecting historical stock prices at finer time granularities. The aim is to apply itemset-based models to drive medium- and short-term investments (e.g., intra-day trading). Furthermore, we will try to apply more advanced itemset mining techniques, such as utility and probabilistic itemset mining, in order to (i) shape investments according to the amounts of stocks already in the portfolio, and (ii) take stock volatility and risk levels into account.

REFERENCES

- [1] M.E. Abdual-Salam, H.M. Abdul-Kader, and W.F. Abdel-Wahed. 2010. Comparative study between Differential Evolution and Particle Swarm Optimization algorithms in training of feed-forward neural network for stock price prediction. In *Informatics and Systems (INFOS), 2010 The 7th International Conference on*. 1–8.
- [2] S.S. Abdullah and M.S. Rahaman. 2012. Stock market prediction model using TPWS and association rules mining. In *Computer and Information Technology (ICCIT), 2012 15th International Conference on*. 390–395. DOI: <http://dx.doi.org/10.1109/ICCITechn.2012.6509756>
- [3] R. Agrawal, T. Imieliński, and A. Swami. 1993. Mining Association Rules between Sets of Items in Large Databases. In *ACM SIGMOD Record*, Vol. 22. ACM, New York, 207–216.
- [4] Elena Baralis, Luca Cagliero, and Tania Cerquitelli. 2016. Supporting stock trading in multiple foreign markets: a multilingual news summarization approach. In *Proceedings of the Second International Workshop on Data Science for Macro-Modeling, DSM@SIGMOD 2016, San Francisco, CA, USA, June 26 - July 1, 2016*. 3:1–3:6. DOI: <http://dx.doi.org/10.1145/2951894.2951897>

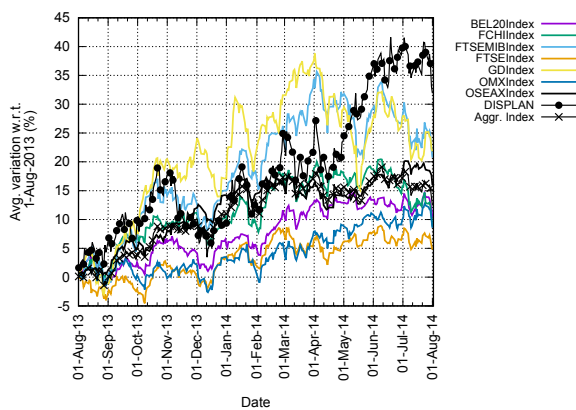


(a) Market-based strategy

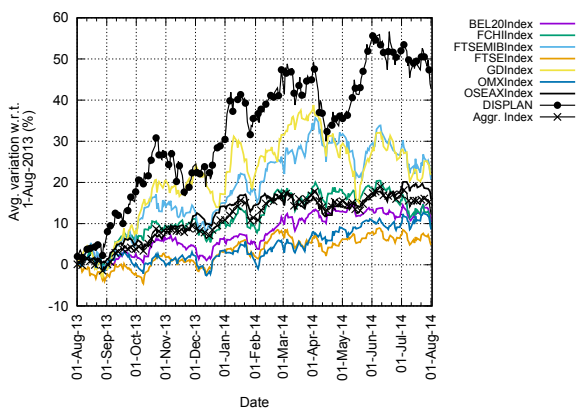


(b) Sector-based strategy

Figure 1: Market- vs sector-based diversification: Average percentage variation w.r.t. August, 1 2013. Asia and Ocean markets, long strategy, $mindiv=70%$, $minret=11%$

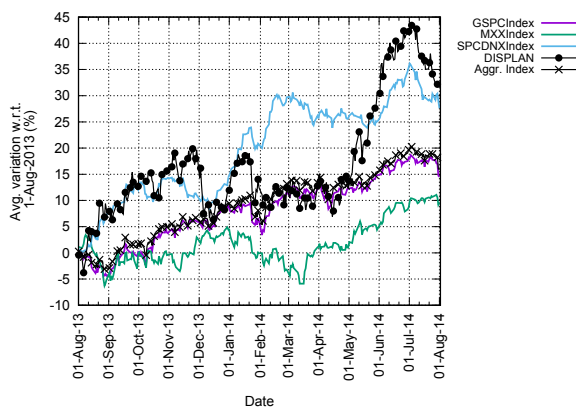


(a) Market-based strategy

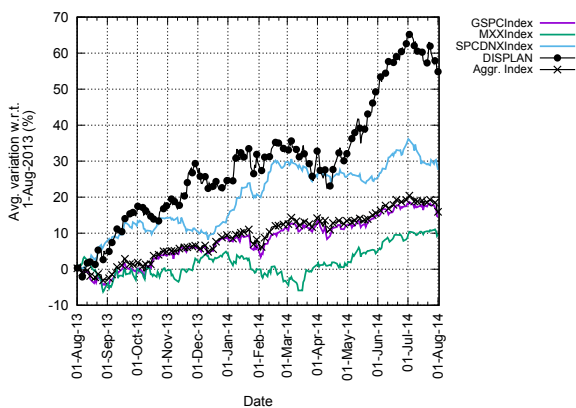


(b) Sector-based strategy

Figure 2: Market- vs sector-based diversification: Average percentage variation w.r.t. August, 1 2013. Europe timezone. Long selling position. $mindiv=70%$, $minret=9%$



(a) Market-based strategy



(b) Sector-based strategy

Figure 3: Market- vs sector-based diversification: Average percentage variation w.r.t. August, 1 2013. North and South America timezone. Long selling position. $mindiv=70%$, $minret=12%$

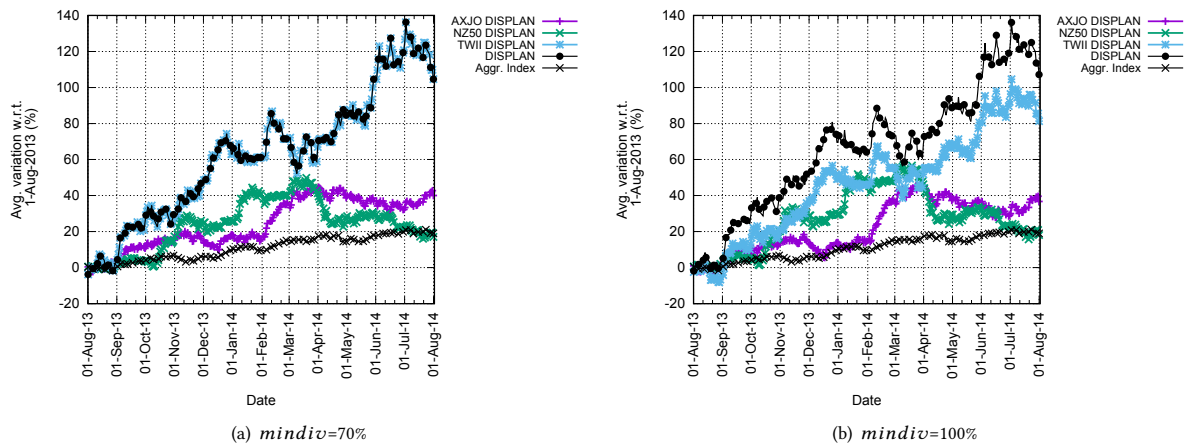


Figure 4: Multi-market vs single-market: Average percentage variation w.r.t. August, 1 2013. Asia and Oceania timezone. Long selling position. $minret=11\%$.

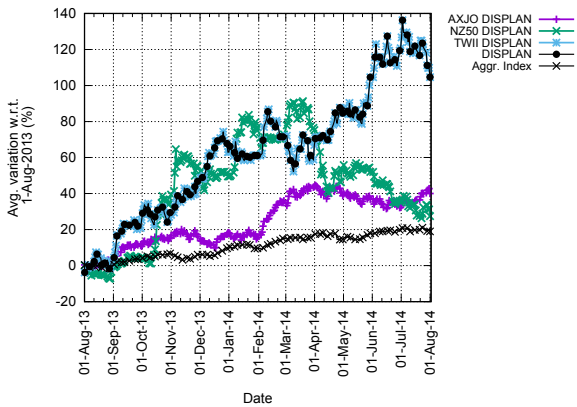


Figure 5: Effect of high minimum relative return threshold: average percentage variation w.r.t. August, 1 2013. Asia and Oceania timezone. Long selling position. $mindiv=70\%$, $minret=16\%$

[5] Elena Baralis, Luca Cagliero, Tania Cerquitelli, Paolo Garza, and Fabio Pulvirenti. 2017. Discovering profitable stocks for intraday trading. *Information Sciences* 405 (2017), 91 – 106. DOI : <http://dx.doi.org/https://doi.org/10.1016/j.ins.2017.04.013>

[6] Elena Baralis, Luca Cagliero, and Paolo Garza. 2017. Planning stock portfolios by means of weighted frequent itemsets. *Expert Syst. Appl.* 86 (2017), 1–17. DOI : <http://dx.doi.org/10.1016/j.eswa.2017.05.051>

[7] Diego J. Bodas-Sagi, Pablo Fernández, J. Ignacio Hidalgo, Francisco J. Soltero, and José L. Risco-Martín. 2009. Multiobjective Optimization of Technical Market Indicators. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers (GECCO '09)*. ACM, New York, NY, USA, 1999–2004. DOI : <http://dx.doi.org/10.1145/1570256.1570266>

[8] Gen-Huey Chen, Ming-Yang Kao, Yuh-Dauh Lyuu, and Hsing-Kuo Wong. 1999. Optimal Buy-and-hold Strategies for Financial Markets with Bounded Daily Returns. In *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing (STOC '99)*. ACM, New York, NY, USA, 119–128. DOI : <http://dx.doi.org/10.1145/301250.301284>

[9] Chonghui Guo, Hongfeng Jia, and Na Zhang. 2008. Time Series Clustering Based on ICA for Stock Data Analysis. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCom '08. 4th International Conference on*. 1–4. DOI : <http://dx.doi.org/10.1109/WiCom.2008.2534>

[10] Larry Harris. 2002. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press. <https://EconPapers.repec.org/RePEc:oxp:obooks:9780195144703>

[11] Fu lai Chung, Tak chung Fu, Robert Luk, and V. Ng. 2002. Evolutionary time series segmentation for stock data mining. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*. 83–90. DOI : <http://dx.doi.org/10.1109/ICDM.2002.1183889>

[12] Yuling Lin, Haixiang Guo, and Jinglu Hu. 2013. An SVM-based approach for stock market trend prediction. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*. 1–7. DOI : <http://dx.doi.org/10.1109/IJCNN.2013.6706743>

[13] Chao Luo, Yanchang Zhao, Longbing Cao, Yuming Ou, and Chengqi Zhang. 2008. Exception Mining on Multiple Time Series in Stock Market. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03 (WI-IAT '08)*. IEEE Computer Society, Washington, DC, USA, 690–693. DOI : <http://dx.doi.org/10.1109/WIIAT.2008.302>

[14] Harry M. Markowitz. 1991. *Portfolio Selection: Efficient Diversification of Investments* (2 ed.). Wiley.

[15] Amin Hedayati Moghaddam, Moein Hedayati Moghaddam, and Morteza Esfandyari. 2016. Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science* 21, 41 (2016), 89 – 93. DOI : <http://dx.doi.org/https://doi.org/10.1016/j.jefas.2016.07.002>

[16] C. Rostoker, A. Wagner, and H. Hoos. 2007. A Parallel Workflow for Real-time Correlation and Clustering of High-Frequency Stock Market Data. In *Parallel and Distributed Processing Symposium, 2007. IPDPS 2007. IEEE International*. 1–10. DOI : <http://dx.doi.org/10.1109/IPDPS.2007.370216>

[17] Chih-Fong Tsai and Zen-Yu Quan. 2014. Stock Prediction by Searching for Similarities in Candlestick Charts. *ACM Trans. Manage. Inf. Syst.* 5, 2, Article 9 (July 2014), 21 pages. DOI : <http://dx.doi.org/10.1145/2591672>

[18] Huacheng Wang, Yanxia Jiang, and Hui Wang. 2009. Stock return prediction based on Bagging-decision tree. In *Grey Systems and Intelligent Services, 2009. GSIS 2009. IEEE International Conference on*. 1575–1580. DOI : <http://dx.doi.org/10.1109/GSIS.2009.5408165>

[19] Wei Wang, Jiong Yang, and Philip S. Yu. 2000. Efficient mining of weighted association rules (WAR). In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'00*. 270–274.

[20] T. Williams and V. Turton. 2014. *Trading Economics: A Guide to Economic Statistics for Practitioners and Students*. Wiley.

[21] Zeng Xiu, Peng Hong, and Zeng Zhen. 2009. Clustering in stock market based on fractal theory. In *Machine Learning and Cybernetics, 2009 International Conference on*, Vol. 1. 161–164. DOI : <http://dx.doi.org/10.1109/ICMLC.2009.5212496>

[22] Y. Xun, J. Zhang, X. Qin, and X. Zhao. 2017. FiDooP-DP: Data Partitioning in Frequent Itemset Mining on Hadoop Clusters. *IEEE Transactions on Parallel and Distributed Systems* 28, 1 (Jan 2017), 101–114. DOI : <http://dx.doi.org/10.1109/TPDS.2016.2560176>

[23] YahooFinance. 2016. Yahoo Finance Website. Last access September 2016. (2016). <https://it.finance.yahoo.com/>

[24] Defu Zhang, Qingshan Jiang, and Xin Li. 2004. Application of Neural Networks in Financial Data Mining. In *International Conference on Computational Intelligence (2005-02-01)*, Ali Okatan (Ed.). International Computational Intelligence Society, 392–395.

[25] Zhe Zhang, Jian Jiang, Xiaoyan Liu, Ricky Lau, Huaiqing Wang, and Rui Zhang. 2010. A Real Time Hybrid Pattern Matching Scheme for Stock Time Series. In *Proceedings of the Twenty-First Australasian Conference on Database Technologies - Volume 104 (ADC '10)*. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 161–170.