Exploring the market predictive power of company comparative relationships from public information

Kun Chen, Peng Luo, Xiaowen Ji

South University of Science and Technology, Shen Zhen, 510008, China

Abstract: In this paper, we want to investigate the financial impacts of comparative relations between companies from public information. By constructing a company comparison network based on sentiment analysis, we apply the vector autoregression (VAR) model to investigate the underlying relationships between the novel network construct and stock performance. The results indicate an immediately bidirectional relationship between the company comparison network and stock performance. They also show that the company comparison network provides better predictive power than the co-citation network. Moreover, supporting relationships have stronger predictive power than opposing relationships on stock returns.

Keywords: Company comparison network, Stock market performance, Vector autoregression

1. Introduction

Public news, blogs, and online discussion boards contain rich and real-time information about companies. With the help of natural language processing and text-mining techniques, this information can be transformed into indicators for stock prediction and risk management. Academic researchers also employ various analysis methods on public information to predict stock price movements, volatility and traded volume. Most of these methods use textual features about a company to predict its stock performances (Oh & Sheng, 2011; Schumaker & Chen, 2009). In these studies, companies or stocks are analyzed individually without considering their interactions. Another notable research stream, however, extracts the co-citation relationships between companies to construct a market network; network metrics are then used to predict stock performances (Jindal & Liu, 2006; Z. Ma, Sheng, & Pant, 2009).

Inspired by the network idea, we wish to construct a company comparison network based on sentiment analysis. The comparison network contains not only company interactions but also interaction sentiments. With the network, we investigate 3 research questions: (1) Do company comparison networks show better predictive power than co-citation networks for stock performances? (2) What are the bidirectional relationships between company comparison networks and stock performance? (3) Which sentiment relationship (supporting or opposing) between companies has a stronger predictive power on stock performances?

To resolve these research questions, we use a vector autoregression (VAR) model, which is a time-series technique. VAR is suitable for examining the dynamics of company comparison network metrics and stock performances with cross effects and reverse causality. It can also measure the relative contributions of different variables.

The paper has implications for both theory and practice. On the theoretical front, it provides a new method to measure company relationships based on public information. It extends the studies that use co-citation and business trade to construct stock networks in IS and finance research. Also it examines the dynamic relationships between network metrics and asset price. On the practical front, the paper provides new suggestions for investors and managers using company comparative relationships for decision making.

2. Research Framework

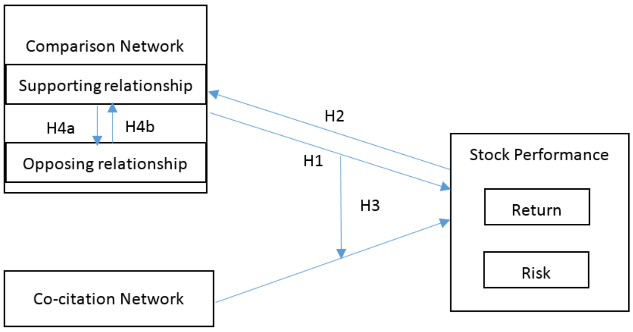

Figure 1 Research Framework

As shown in Figure 1, our research framework integrates various time-varying relationships among comparison network variables and stock market performances. In this study, we wish to focus on company comparison networks and investigate the bidirectional relationships between the comparison network and stock performance (H1 and H2); the market predictive power of comparison networks and co-citation networks (H3); the predictive power of supporting and opposing relationships in comparison networks (H4a and H4b). Previous work has shown that company co-citation networks can affect company stock performance, and we therefore expect that the comparative analysis of companies may also provide predictive power for stock performance. Furthermore, we expect comparative analysis to provide better company interaction (network) metrics than co-citation analysis.

H1: Company comparison network metrics have significant predictive power for company stock performance.

H3: Comparative analysis provides a stronger network indicator than co-citation metrics.

In addition to the known predictive relationship from company comparison networks to stock performance, our research framework also suggests a reverse predictive relationship from stock performance to company comparison network metrics. This means the market performance of a company may affect public reports about it. If the company's market performance is representative, then the company may be mentioned more frequently (along with other companies) in the news. Based on this assumption, we make the following assertion:

H2: Company stock performance has significant predictive power for company comparison network metrics.

To further investigate the sentiment of comparative opinions, we divide the competitive relationships into two categories, namely, supporting or positive relationships and opposing or negative relationships. Sentiment analysis has been widely adopted in financial studies to predict stock prices.

H4a: Supporting (positive) competitive relationships have a stronger predictive power than opposing (negative) competitive relationships.

H4b: Opposing (negative) competitive relationships have a stronger predictive power than supporting (positive) competitive relationships.

3. Data, Network, and Measurements

The raw data set consists of one year (2013) of business news for 300 companies in the Shanghai-Shenzhen 300 index. News stories are collected from a general search portal that covers 3000+ online sources. To identify intercompany relationships, we first filtered news items that mention only one company or mention more than 5 companies in a document. The next task was to find target companies in a news story. For example, if a company name appears in the title, it is the target company. If no company names appear in the title, we count the occurrence times of company names, and label the companies with the highest frequencies the target companies. In the next step, we wish to identify comparative opinions between companies using machine-learning methods. The basic idea is in line with the work of (Z. G. Zhang, Chenhui; Goes, Paulo, 2013).

In the comparison network, assume each sentence for target company c1 along with a mention of company c2 is mapped onto a comparison tuple $\mathrm{t} = \{\mathrm{c1}, \mathrm{c2}, \mathrm{P/N}\}$, where P/N means that the comparative opinion from c2 to c1 is supporting (positive) or opposing (negative). We also define N_{pd} as the number of supporting tuples and N_{nd} as the number of opposing tuples. Second, in the supporting network, an edge from node c1 to c2 is introduced when $N_{pd} > 0$ and the weight is $w = N_{pd}$. Similarly, when $N_{nd} > 0$, we can introduce a link from c1 to c2 and set the weight as $w = N_{nd}$ in the opposing network. Third, An edge between node c1 and c2 is introduced in undirected networ when $(N_{pd} + N_{nd}) > 0$, and the weight of the link is $w = (N_{pd} + N_{nd})$. Table 1 give a brief introduction of all variables.

Following previous research (Luo, Zhang, & Duan, 2013), we use two common measures to measure stock performance, abnormal return and risk. Abnormal return refers to stock valued beyond what is expected by the average stock market. Risk, referring to the vulnerability of stock value, can be measured as the standard deviation of the residuals of return:

$$R_{it} - R_{ft} = \alpha_i + \beta_i \left(R_{mt} - R_{ft} \right) + \varepsilon_{it} \quad (2)$$

where t is the subscript for the time period, R_{it} is the return of stock i in time t, R_{mt} is the average market returns represented by the Shanghai Composite Index, R_{ft} is the risk-free rate of return, α_i is the intercept and ε_{it} is the model residual. Equation (2) is processed for a rolling window of 250 trading days prior to the target day. The abnormal return of stock i (AR_i) is measured as the difference between the observed return and the expected return in Equation (3).

$$AR_{it} = \left(R_{it} - R_{ft} \right) - \left(\alpha_i + \beta_i \left(R_{mt} - R_{ft} \right) \right) \quad (3)$$

The network measures on the node are both degree- and centrality-based. Degree-based variables include in-degree and out-degree. Centrality-based variables include PageRank (Brin & Page, 1998), HITS (Kleinberg, 1999), Bonacich centrality (Bonacich, 1987) and closeness centrality (Sabidussi, 1966). The degree-based attributes measure the number of a node's connections to its neighbor nodes. Centrality attributes consider the entire network and represent the importance (or position) of a node in the entire network. Table 3 summarizes all of the measurements in this study.

Table 1 A Summary of Variables

Measures	Variables	Description
Return	$return_{it}$	The abnormal return of stock price for company i at day t.
Risk	$risk_{it}$	The idiosyncratic risk of the stock price for company i at day t.
Degree	cd_{it}	The degree of company i at day t in the co-citation network.
In-degree	sid_{it}	The in-degree of company i at day t in the supporting network.
	oid_{it}	The in-degree of company i at day t in the opposing network.

Out-degree	sod_{it}	The out-degree of company i at day t in the supporting network.
	ood_{it}	The out-degree of company i at day t in the opposing network.
PageRank	spc_{it}	The PageRank centrality of company i at day t in the supporting network.
	opc_{it}	The PageRank centrality of company i at day t in the opposing network.
HITS	shc_{it}	The HITS centrality of company i at day t in the supporting network.
	ohc_{it}	The HITS centrality of company i at day t in the opposing network.
Bonacich Centrality	cbc_{it}	The Bonacich centrality of company i at day t in the co-citation network.
Closeness Centrality	ccc_{it}	The Closeness centrality of company i at day t in the co-citation network.

4. Econometric Model and Results

We adopt the VAR model for empirical investigation. In our paper, the VAR model track the dynamic cumulative effects of company comparison network variables through generalized impulse response functions (GIRFs) (Pesaran & Shin, 1998), and vice versa. The model can also assess the relative contribution of various metrics of comparison networks through generalized forecast error variance decomposition (GFEVD) (Pesaran & Shin, 1998). To estimate the VAR model for individual companies, we use the companies in the banking industry for the experiment. The names of 14 banks appear more than 80 days during a year-long period (238 days). We conduct stationary and unit-root tests to examine the stability of stock performance metrics and network metrics. The augmented Dickey-Fuller (ADF) test is used to check stationarity.

4.1 The Bidirectional Relationships between Comparison Networks and Stock Performance

Table 2 below reports the immediate impulsive response and the average across firms of stock performance to the comparison network metrics through the simulation of GIRF. The immediate impact is defined as the effect derived from the estimates of the VAR model for the first three time periods. We note the impact on stock return and risk at basis points (one basis point is one hundredth of a percentage). Among these metrics, the shc (HITS centrality in a supporting network) shows a significantly positive impact on returns (0.41 basis points $p < 0.1$) and a significantly negative impact of 0.002 basis points ($p <0.1$) on risk. That is, an unexpected increase in shc values will predict a rise in daily stock returns by 0.41 basis points and a drop in stock risk by 0.002 basis points. Although these effects seem to be small relevant to the basis points, they have a substantial impact in terms of the dollar value. In monetary terms, the relationships between company comparison networks and stock performances could translate into a significant impact on a company's market capitalization. For example, other factors remaining equal, a unit increase in HITS centrality in a supporting network could erode approximately $0.41 million from the average market capitalization in the short term. Although the oid (in-degree in opposing network) has no significant impact on returns, it has a significant negative influence on risk (0.011 basis points $p < 0.1$). On the contrary, the opc (PageRank centrality in an opposing network) significantly predicts a decrease of returns (-3.90 basis points, $p < 0.1$) and is insignificantly related to risk. Thus, the results suggest that company comparison network metrics have significant predictive powers for company stock performances; thus, H1 is supported.

Table 2 Impulse Response of Stock Performance to Comparison Network Metrics

	sid	sod	Shc	spc	oid	ood	ohc	opc
return	-1.02	-5.68	**0.41***	-4.80	-9.84	-7.10	-13.40	**-3.90***
risk	-0.014	-0.017	**-0.002***	0.0063	**-0.011***	-0.043	0.032	-0.058

Notes. The coefficients of return and risk are in basis points (1 basis point = hundredth of a percentage). Significant values are represented in bold. * p < 0.1, ** p < 0.05, *** p < 0.01

Furthermore, we also investigate the immediate responses of company comparison network metrics on stock performance. Table 3 suggests that the return has a significantly positive relationship with in-degree (0.89, $p < 0.1$) and PageRank centrality (0.0000097, $p < 0.1$) in a supporting network. This shows that a change in returns is associated with an increase of the values of in-degree and PageRank centrality in a supporting network. Conversely, the firm returns also significantly predict a decrease in HITS centrality (-0.017, $p < 0.1$) and PageRank centrality (-0.00000015, $p < 0.1$) in opposing networks. The risk further shows a strong positive predictive value with PageRank centrality in a supporting network (0.000019, $p < 0.1$). Accordingly, these results reflect strong empirical evidence for H2, that company stock performance has a significant predictive power on company comparison network metrics.

Table 3 Impulse Response of Comparison Network Metrics to Stock Performances

	sid	sod	shc	Spc	oid	ood	ohc	opc
return	**0.89***	1.05	0.011	**0.0000097***	-0.021	-0.047	**-0.017***	**-0.00000015***
risk	1.36	2.07	-0.016	**0.000019***	0.016	-0.023	0.0071	-0.0000016

Notes. Significant values are represented in bold. * p < 0.1, ** p < 0.05, *** p < 0.01

4.2 Comparison Networks versus Co-citation Networks

With the two VAR models described in equations (4) and (5), we compare the R^2 of the two models: model 1 (stock performances and co-citation network metrics) and model 2 (stock performance and comparison network metrics). As shown in Table 4, the comparison network model explains stock returns with an R^2 of 0.56, outperforming the co-citation network model with an R^2 of 0.28 ($F = 48.94$ and $p < 0.01$). Further, the comparison network model also significantly outperforms the co-citation network model in explaining the stock risk with an R^2 of 0.48 ($F = 16.41$ and $p < 0.01$). Therefore, the H3 is well supported.

Table 4 A Comparison of Model 1 and Model 2

	R^2 (model 1)	R^2 (model 2)	R^2 (model 1)> R^2 (model 2)	
return	0.56	0.28	F-test	48.94***
risk	0.71	0.48		16.41***

* p<0.1, ** p<0.05, *** p<0.01

4.3 Supporting Relationships versus Opposing Relationships

The variance decomposition of the GFEVD results in Table 5 provides the relative importance of comparison network metrics in explaining the variance of stock performance. The results suggest the order of contributions as sid (8.59%), sod (7.84%), oid (7.57%), ood (6.43%), spc (4.52%), shc (4.38%), ohc (3.93%), and opc (3.18%) in predicting returns. Similarly, in forecasting risk, the results order the contributions of social network metrics as sid (8.34%), ood (7.49%), oid (6.91%), sod (6.06%), ohc (5.32%), shc (4.84%), opc (4.03%), and spc (3.72%). Then, in explaining returns, the total supporting network metrics account for a significantly greater proportion of the variance than the total opposing network metrics (25.33% versus 21.11% and $F = 2.39$, $p < 0.1$). In the variance decomposition of risk, the total opposing network metrics occupy a larger proportion of variance than the total supporting network metrics (23.75% versus 22.96%); however, it is not significant, according to the F statistics ($F = 0.072$, $p > 0.1$). Thus, it partially supports H4a, implying that supporting competitive relationships have stronger predictive power than opposing (negative) competitive relationships in predicting stock returns.

Table 5 Variance Decomposition of Stock Performance as Explained by Network Metrics

	sid	sod	Shc	spc	Total supporting network metrics
return	8.59	7.84	4.38	4.52	25.33
risk	8.34	6.06	4.84	3.72	22.96
	oid	ood	Ohc	opc	Total opposing network metrics
return	7.57	6.43	3.93	3.18	21.11
risk	6.91	7.49	5.32	4.03	23.75
return	Total supporting network metrics > Total opposing network metrics				
F-test	2.39*				
risk	Total opposing network metrics > Total supporting network metrics				
F-test	0.072				

* p < 0.1, ** p < 0.05, *** p < 0.01

5. Discussion

From the current results, three major theoretical insights can be drawn. First, we propose a unique company social network construct. Further, we find there are bidirectional impacts between the network and stock performance. The supporting links between companies in public reports lead to positive returns for the target stock, and opposing links between companies lead to negative returns. Conversely, a positive shock on stock returns would increase supporting links to the company, and a negative shock on stock returns would increase opposing links to the company. This finding reveals the relationship between public information and stock performance. Unlike previous work seeking to predict stock performance based on textual features—e.g., Li et al (Li, Xie, Chen, Wang, & Deng, 2014); Yu, Y., Duan, W. and Cao, Q. (Yu et al., 2013); and Tetlock, Saar-Tsechansky, and Macskassy (Tetlock et al., 2008)— our work discovers a two-way impact on underlying company relationships mined from public information and stock performance.

Second, although it has been suggested in marketing research that comparative analysis would improve prediction in product sales, we extend the idea into the finance domain and find that comparative analysis has a better predictive power for stock performance than the previously used co-citation analysis. Moreover, this paper is inspired by work in the finance literature that investigates company interactions based on trade data (e.g., Aobdia, Caskey, and Ozel (Aobdia et al., 2013); Ahern and Harford (Ahern & Harford, 2014)) because public news may include more business relationships between two companies, such as investments or debt. To an extent, the work also inspires traditional asset pricing theory (Sharpe, 1964) by providing a special perspective for investigating the structural features of stocks based on an overview of the market.

Third, in terms of supporting versus opposing relationships, we find that supporting relationships explain significantly more than opposing relationships about return. With respect to risk, although opposing relationships explain more than supporting relationships, the difference is not significant. These results are partially supported by the research of Chan, W.S. (Chan, 2003) and Van, P.N. (Van, 2015) and extend the stream of research from textual sentiment analysis of individual companies to the comparative analysis of multiple companies.

In addition to the theoretical implications discussed above, this study has practical implications for both investors and managers. For investors, the predictive model suggests that company comparative networks mined from public news would be quite useful in portfolio and risk management. Therefore, daily network metrics should be taken into consideration when making an investment decision. For managers, the findings suggest that companies should strengthen efforts to improve their public exposure. At the same time, they should monitor the dynamics of social networks from public information, especially when negative interactions occur.

Reference

Ahern, K. R., & Harford, J. (2014). The importance of industry links in merger waves. *The Journal of Finance, 69*(2), 527-576.

Aobdia, D., Caskey, J., & Ozel, N. B. (2013). Inter-industry network structure and the cross-predictability of earnings and stock returns. *Review of Accounting Studies, 19*(3), 1191-1224.

Bonacich, P. (1987). Power and centrality: A family of mesures. *American Journal of Sociology, 92*(5), 1170-1182.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*(1-7), 107-117.

Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. Journal of Financial Economics, 70(2), 223-260.

Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. Paper presented at the AAAI.

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems, 69*, 14-23.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of ACM, 46*(5), 6004-6632.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research, 24*(1), 146-163.

Ma, Z., Sheng, O. R., & Pant, G. (2009). Discovering company revenue relations from news: A network approach. Decision Support Systems, 47(4), 408-414.

Oh, C., & Sheng, O. R. L. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. Paper presented at the Thirty Second International Conference on Information Systems, Shanghai, China.

Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. Economics letters, 58(1), 17-29.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika, 31*(4), 581-603.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: the AZFin text system. ACM Transactions on Information Systems, 27(2), 12.

Sharpe, W. F. (1964). Captiical asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425-442.

Tetlock, P. C., Saar‐Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance, 63*(3), 1437-1467.

Van, P. N. (2015). A good news or bad news has greater impact on the Vietnamese stock market? : Banking Academy of Vietnam, State Bank of Vietnam.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems, 55*(4), 919-926.

Zhang, Z. G., Chenhui; Goes, Paulo. (2013). Product comparison networks for competitive analysis of online word-of-mouth. ACM Transactions on Management Information Systems (TMIS), 3(4), 20:21-20:22.