Journal of Banking and Finance 91 (2018) 49–69

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Qualitative similarity and stock price comovement

Travis Box[1]

School of Business Administration, University of Mississippi, 340 Holman, P.O. Box 1848, University, MS 38677-1848, United States

ARTICLE INFO

Article history:
Received 29 November 2016
Accepted 14 April 2018
Available online 21 April 2018

JEL classification:
C33
D81
D83
G12
G14

Keywords:
Information consumption
Information-driven comovement
Textual analysis
Correlation

ABSTRACT

I introduce a method for gauging the qualitative similarity of firm-specific information based on linguistic commonality in newswire text. I show that this new qualitative similarity measure predicts future cross-firm return correlation even after accounting for the pair's contemporaneous price comovement, common exposures to systematic risk, firm liquidity, price, index membership, text volume, headquarters location, product similarity, shared mutual fund or institutional ownership, common analyst following and newswire co-mentions. I also demonstrate that content produced solely by journalists cannot predict an economically meaningful portion of future comovement. Out-of-sample tests confirm that knowledge of qualitative similarity can also reduce portfolio risk.

The primary determinant of equity portfolio risk is the likelihood that pairs of stock prices will rise and fall together. Both in research and in practice, our expectations regarding future price comovement, and our appraisals of portfolio risk, have relied on lengthy series of historical stock returns. Yet, the mechanism generating these returns depends on a flow of firm-specific information that changes throughout time. As new sources of opportunity and uncertainty are revealed to the market, the links between distant historical prices and future stock price comovement become weaker. Therefore, accurate predictions of comovement must also consider the similarity of contemporaneous information flows across firms. I develop a proxy for this similarity and test whether this new measure can improve predictions of future stock price comovement.

The field of finance is replete with simple quantitative descriptors designed to identify similarities in firm characteristics. However, contemporaneous changes in the flow of information are not reflected in these quantitative measures until firms announce earnings or publish financial reports. To identify contemporaneous changes in firm similarity, investors must rely on softer, more qualitative, sources of information. In real-time, this content is often delivered to the market through financial newswires. These services act as information conduits by compressing a vast array of firm-specific material into a digestible sequence that investors can use to make portfolio decisions. This paper examines whether the qualitative information circulated on one such newswire, the Reuters Integrated Data Network, can predict how future equity payoffs are correlated across firms.

During each six-month period from 2003 to 2013, I measure the similarity of firm-specific newswire text written about different companies. I propose that the contemporaneous information flows for two firms are qualitatively similar if there is commonality in their newswire text. In support of this hypothesis, I find that the qualitative similarity of the newswire items written about a firm pair predicts their stock return correlation during the following six-month period. Furthermore, this new measure of qualitative similarity can predict future price comovement even after accounting for the pair's contemporaneous return correlation. Thus, qualitative similarity describes similarity in information flows that cannot be inferred from historical stock prices.

Prior literature has recognized that characterisitics such as firm beta (Ledoit and Wolf 2003), size (Pindyck and Rotemberg 1993), book-to-market (Bekaert et al., 2009), momentum (Asness et al., 2013) and industry (Campbell et al., 2001; Irvine and Pontiff, 2009;

E-mail address: tbox@bus.olemiss.edu

[1] I thank the University of Mississippi for their research support. I am grateful for the helpful comments of Eric Kelley, Volkan Muslu, Adam Kleinbaum, Andrew Lynch, Ryan Davis, Robert Van Ness, Seong Byun, Angela Box and all the participants at the Ole Miss Finance Department Seminar, the Midwest Finance Association Annual Meeting, and the Arrowstreet Capital, LP Practitioner Seminar. I thank Paul Tetlock from Columbia University and Richard Brown and Maciek Pomalecki from Thomson Reuters Machine Readable News for data considerations. All mistakes in this article are my own.

https://doi.org/10.1016/j.jbankfin.2018.04.010
0378-4266/

Brandt et al., 2010) proxy for common sources of systematic variance that generate price comovement between firms. The literature also offers many alternative explanations for stock price comovement that are based on some type of market friction. Specifically, cross-sectional variation in information diffusion (Barberis et al., 2005), as well as the categorical trading of assets (Barberis and Schleifer 2003), have been shown to cause higher levels of stock price comovement. To ensure that qualitative similarity does not proxy for one of these other documented sources of return correlation, I show that my measure's predictability remains after controlling for similarities in exposure to systematic risk as well as firm liquidity, price, index membership, text volume and headquarters location. Thus, commonality in the information flow across firms predicts return correlation that cannot be accounted for with standard asset pricing models and alternative explanations for stock price comovement

The newswire text appearing on the Reuters IDN originates from a variety of sources and sources and spans a broad range of topics. To better understand how the contemporaneous flow of information predicts future stock price comovement, I divide my sample of newswire text along two dimensions. First, I consider whether the qualitative similarity of text produced by journalists is more or less informative than content generated by the firms themselves. Second, I determine if the relation of interest depends on whether the newswire content describes the financial results of the firm.

Most of the text circulated on the Reuters Integrated Data Network is generated by the firms themselves in the form of press releases and legal disclosures. However, I give special attention to content written by journalists because a great deal of prior literature focuses specifically on the role of text produced by the traditional press (see Barber and Loeffler, 1993; Tetlock, 2007; Fang and Peress, 2009; Tetlock et al., 2008; Tetlock, 2011; Peress, 2014). If, as suggested by Ahern and Sosyura (2014), journalist-produced content merely summarizes primary sources that are written by the firms, then a thorough examination of their output will not provide much in the way of meaningful insights. Accordingly, I find only weak evidence that the qualitative similarity of newswire text produced by the financial press can predict an economically meaningful portion of future cross-firm comovement.

Pindyck and Rotemberg (1993) propose that the comovements of individual stock prices should depend only on expectations about future earnings. However, when I divide my newswire sample by topic, I find that text related to corporate financial results is a weaker predictor of future return correlation. Nevertheless, the predictive performance of newswire content focused on results does not imply that these stories contain less information. If most of the information revealed in these newswire items is communicated through a numerical value, such as an earnings level, then the text accompanying this release might contain less important qualitative information. In either case, the most notable result from this analysis is that truncation, whether between journalist and firm or earnings and non-earnings, leads to a significant loss of qualitative information relative to the full sample.

Next, a series of closely related projects attempt to quantify qualitative information produced by either the firms themselves or by some other information producer. With an eye toward predicting return correlation, Israelsen (2015) and Muslu et al. (2014) study common analyst coverage and Anton and Polk (2014) look at shared ownership among actively managed mutual funds. While Hoberg and Phillips (2010a, 2010b) do not forecast stock price comovement directly, they propose a text-based measure of product differentiation that should be well suited to the task. Tetlock (2007) and Tetlock et al. (2008) show that tonal measures of news text, like pessimism, can predict stock prices and accounting earnings. Finally, Scherbina and Schlusche (2015) identify cross-firm predicatability in stock returns for companies that are mentioned together in certain types of news stories.

Additional tests confirm that newswire text from the Reuters Integrated Data Network contains at least some information about future return correlations that is orthogonal to the other sources of qualitative information highlighted by these related projects. These tests also demonstrate that the Hoberg and Phillips (2010a, 2010b) product similarity measures are strong predictors of future comovement. However, the results are less encouraging with regards to the remaining sources of qualitative information. First, I find little evidence that textual tone contributes positively to future return correlation. Furthermore, the variables measuring shared mutual fund or institutional ownership, common analyst following and newswire co-mentions appear to be correlated with persistent firm-pair panel effects. It is not surprising that such connections are persistent enough to be subsumed by panel effects if specific analysts and reporters follow, or institutions and mutual funds hold, firms with similar characteristics.

In recent years, another growing body of research has examined return predictability arising from interfirm linkages and investor inattention. Cohen and Frazzini (2008) propose that stock prices do not promptly incorporate news about economically related firms when investors are subject to information constraints. In support of their hypothesis, they find evidence of return predictability across groups of firms that are linked through customer-supplier relationships. Likewise, Menzly and Ozbas (2010) find that stocks in economically related supplier and customer industries cross-predict each other's returns. Also, Cohen and Lou (2012) posit that limited information processing capacity, not just inattention, can lead to a significant delay in the impounding of information into asset prices. They demonstrate that the returns of stand-alone firms predict the returns of more complex conglomerate firms that conduct some their business in the same industry. Finally, Cao et al. (2016) find evidence of return predictability between firms engaging in strategic alliances.

To confirm that qualitative similarity is not a proxy for these relationships, I perform my analysis on subsets of firms that are less likely to have identifiable economic linkages. Thus, I remove all observations for firm-pairs mentioned in the same newswire item and all firm-pairs that are in the same industry according to the Hoberg and Phillips (2010a, 2010b) product similarity measure. I also remove all firm-pairs with "second-tier" linkages whereby two firms are not linked directly but are linked through their respective direct linkages to some other firm. These filters remove combinations where the companies sell a similar product, announce a strategic partnership or have a public supply chain relationship at any point during my sample period. My results demonstrate that qualitative similarity is still able to predict future comovement between firm pairs that lack these types of direct or indirect economic linkages.

Finally, to evaluate the economic significance of the relation between qualitative similarity and future stock price comovement, I test whether forecasts of rolling correlations can reduce the out-of-sample volatility of an equity portfolio. In general, I find that portfolios based on forecasted correlations have dramatically lower standard deviations than passive strategies such as market- and equal-weighted portfolios. Furthermore, out-of-sample correlation forecasts benefit when qualitative similarity is included in the regression specification. Ultimately, my results indicate that investors may reduce the out-of-sample volatility their portfolios by incorporating the qualitative similarity of firm-specific information into their covariance predictions.