9-2021

Media connection and return comovement

Zilin CHEN

Li GUO

Jun TU

Jun TU
Singapore Management University, tujun@smu.edu.sg

Published in Journal of Economic Dynamics and Control. September 2021, pp. 1-18.
https://doi.org/10.1016/j.jedc.2021.104191

Media connection and return comovement☆

Zilin Chen[a], Li Guo[b,c], Jun Tu[d,*]

[a] *School of Finance, Southwestern University of Finance and Economics, 555 Liutai Avenue, Chengdu, 611130, China*
[b] *School of Economics, Fudan University, 600 Guoquan Road, Shanghai, 200433, China*
[c] *Shanghai Institute of International Finance and Economics, China*
[d] *Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore, 178899*

ARTICLE INFO

Article history:
Received 4 January 2021
Revised 3 July 2021
Accepted 8 July 2021
Available online 13 July 2021

JEL classification:
G11
G12
G41

Keywords:
News linkages
Fundamental comovement
Return comovement
Journalists collective opinions
Complex economic linkages

ABSTRACT

Media news may cover multiple firms in one article, which establishes a media connection across firms. We propose a media connection strength (MCS) measure between two given firms, which is defined as the number of news articles co-mentioning these two firms. We show that the MCS measure can significantly explain and forecast return comovement of media-connected firm-pairs. Further analyses show that our results are robust to various alternative explanations. We argue that the MCS measure can capture comprehensive and complex correlated fundamental information among media-connected firms and hence may provide a new mechanism for return comovement beyond the existing rational- and behavioral-based explanations.

1. Introduction

Understanding the driving forces of stock return comovement has important implications for many applications, such as risk management (Philippe (2001)), portfolio allocation (Qian et al. (2007)), asset price dynamics (Rosenberg and Schuermann (2006) and Brooks et al. (2002)), and trading strategies (Gatev et al. (2006); Papadakis and Wysocki (2007); Chen et al. (2019)). Theoretically, the stock return comovement of two firms can be driven by common shocks to their fundamentals, such as future cash flows or discount rates. Therefore, empirically explanatory variables based on the fundamental information of a firm, usually quantitative information, have been proposed to explain comovement (e.g., Shiller (1989); Fama and French (1993); Chen et al. (2019); Green and Hwang (2009)). In addition, recent studies document æexcess comovementg that cannot be explained by fundamental information and propose alternative behavioral bias-based ex-

☆ We thank Dashan Huang, Jianfeng Hu, Roger Loh, Weikai Li, Chishen Wei, Weina Zhang, and participants at 2019 SMU Summer Camp and Conference on the Theories and Practices of Securities and Financial Markets 2019 for their helpful suggestions. We also thank Eugene Fama and Kenneth French for sharing the Fama-French factors returns, Malcolm Baker and Jeffrey Wurgler for sharing the sentiment index data, and Gerard Hoberg and Gordon M. Phillips for sharing the text-based network industry classifications (TNIC) data. Li Guo acknowledges the financial support sponsored by Shanghai Pujiang Program. Jun Tu acknowledges that the study was funded through a research grant from Sim Kee Boon Institute for Financial Economics. All errors remain our responsibility.

* Corresponding author.
E-mail addresses: chenzilin@swufe.edu.cn (Z. Chen), guo_li@fudan.edu.cn (L. Guo), tujun@smu.edu.sg (J. Tu).

planations for return comovement (e.g., Karolyi and Stulz (1996); Barberis and Shleifer (2003); Barberis et al. (2005); Bekaert et al. (2005); Kumar and Lee (2006); Kallberg and Pasquariello (2008); Anton and Polk (2014); Kumar et al. (2016)).

In this study, we investigate a new potential channel for return comovement; namely, media connection, which is beyond the existing fundamental information channels and is not based on behavioral bias. In particular, we examine whether correlated fundamental information contained in multi-firm news articles could be an important driver for return comovement. Multi-firm news articles covering two or more firms may reflect journalists collective opinions on a certain complicated fundamental connectivity between the firms that are cited. In addition to well-documented fundamental links including customer-supplier links (Cohen and Frazzini (2008)), text-based links (Hoberg and Phillips (2016)), and EDGAR co-search links (Lee et al. (2015)), multi-firm news articles may be a proxy of other complex economic linkages, such as strategic partnerships, intra- and inter-sectoral competitive links, and credit, financing, banking, and subsidiary relations (Schwenkler and Zheng (2019)). Therefore, we propose a media connection strength (MCS) measure based on the news data collected by Thomson Reuters News Analytics (TRNA), which is defined as the number of news articles covering two firms.[1] We expect that the MCS measure can offer incrementally better explanatory power regarding the comovement of fundamental performance and return comovement for media-connected firm pairs that goes beyond those existing variables measuring correlated fundamentals across firms.

We conduct empirical tests to verify our hypothesis. First, we examine the determinants of the MCS. We run cross-sectional regressions of our MCS measure on the firm-pair characteristics that are documented in the literature as associated with return comovement. We find that variables such as the mutual fund common ownership and common analyst coverage are significantly and positively associated with MCS. Further tests show that many fundamental similarities between co-mentioned stocks can significantly explain MCS, suggesting that fundamental similarities of stock pairs are indeed captured by our MCS.

Next, we examine whether the MCS measure can predict fundamental similarities of media-connected firm pairs. Following Livnat and Mendenhall (2006), we define the standardized unexpected earnings (SUE) as a proxy for the fundamental information of a firm and find that the correlation of SUE, when controlling for other well-known return comovement sources and other specific fundamental similarities, can be positively and significantly predicted by the MCS among the co-mentioned stock pairs. Specifically, one standard deviation of the increase in the MCS is associated with an increase of 0.6% in the correlation of SUE in quarter t and an increase of 1.0% in quarter t+1. Meanwhile, to further validate our fundamental argument for a media connection, we employ the correlation of sales growth (SALEG) as an alternative measure of fundamental comovement, and the results are also consistent with our expectation; that is, a one standard deviation increase in the MCS of quarter t is associated with an increase of 1.8% in the correlation of SALEG among firm pairs in quarter t and an increase of 1.9% in quarter t+1, when controlling for other well-known return comovement sources and other fundamental similarities. Finally, we also measure the correlated fundamentals of stock pairs as the negative value of the absolute difference in percentile ranking of SUE or SALEG for the media-connected stock pairs, and the results remain similar. This suggests that media connection contributes to an increasing fundamental news comovement for media-connected firm pairs beyond what can be explained by other stock-pair similarities documented in the literature.

Subsequently, we conduct a simple univariate analysis to examine if the predictive power of our MCS on fundamental similarities can be extended to an association between MCS and excessive return comovement. We first split all firm pairs into three groups based on the MCS and calculate the average return correlation of these three groups; we then form a high-minus-low group that captures the return correlation difference between the high-MCS group and the low-MCS group. We find that the return correlation of the high-MCS group is significantly higher than that of the low-MCS group. In addition, the average return correlation of the stock pairs with no-zero MCS is significantly higher than that of those with zero MCS, indicating that our MCS is indeed positively associated with a stronger return comovement. We then perform cross-sectional regressions to formally investigate if the contemporaneous and future excess return comovement of media-connected firm pairs can be significantly explained through the MCS measure. Consistent with our hypothesis, the MCS is positively and significantly associated with not only cross-sectional variation in the strength of fundamental similarities but also a stronger return comovement among firm pairs in the current month and next month, when controlling for other well-known return comovement sources and a list of firm-pair fundamental similarities. A one standard deviation increase in MCS is associated with an increase of 0.9% in return comovement and an increase of 1.0% in excess return comovement (under the Fama and French (2015) five-factor model (FF5)) in the current month, and the results are similar to that of the next month. Moreover, the predicted variation in FF5 residual correlation is in the average range of 5.33% to 16.75%, with a mean abnormal correlation of 7%, suggesting that the effect of MCS is also economically significant.

Kumar and Lee (2006), Kumar et al. (2013), and Kumar et al. (2016) find that stock prices tend to co-move if they share the same sentiment. Given that stocks mentioned in the same news article may share a similar sentiment, we examine whether the return comovement predictability of our MCS is independent of investor sentiment. Through defining high- and low-sentiment periods based on the sentiment index of Baker and Wurgler (2006), we find that the predictability of the MCS is robust and similar in magnitude under different sentiment regimes, suggesting that investor sentiment is unlikely to explain the results. In addition, we find that the MCS measure is positively associated with trading activity comovement,

[1] Thomson Reuters not only provides the number of news articles (i.e., news coverage) related to a firm but also provides the positive (or negative) value-relevant information (i.e., news tone score) in each article based on their specialized algorithms.

measured as the correlation of order flow or turnover ratio between paired firms. This is consistent with the hypothesis that the correlated fundamentals are captured through the MCS measure, which may induce correlated trading activities and the corresponding return comovement.

Nevertheless, investors may simply trade on media-connected stock pairs due to category trading habitat instead of the fundamental similarities between those stock pairs. Barberis and Shleifer (2003) argue that investors allocate capital at the level of asset categories rather than individual stocks. Firms are mentioned frequently in the same news article because they are viewed as being in the same asset category by market participants, including news media and investors. Alternatively, investors view of asset categories could be influenced by the frequency of stocks being mentioned together in news articles. In either case, investors trading in and out of media-connected stocks that are viewed as being in the same category could induce an excessive comovement in their returns.

However, this is less likely to explain our results, given that the coefficients of MCS on return comovement remain similar during both low- and high-sentiment periods. We further rule out this alternative hypothesis in two ways. Firstly, we replace the correlation of order flow (OFCOR) or the correlation of turnover (TOCOR) in the current month with the similarity in turnover (TO) and the correlation of TO in the past 60 months (VOLCOR) in the cross-sectional regressions of the return comovement so as to control for the correlated flow-induced trading induced by category trading habitat. The results show that the effect of our MCS is still there after the inclusion of these control variables. Secondly, as is detailed in the latter part of the paper, we conduct portfolio analysis to further show that category trading habitat is unlikely an explanation of our results.

Finally, we conduct robust checks to examine whether other fundamental similarities documented in the literature can be used to explain our results. Pirinsky and Wang (2006) find that stock prices tend to co-move if their headquarters are located in the same place. Thus, we add a dummy variable to our regression indicating whether two stocks are headquartered in the same U.S. county or not. Consistent with Pirinsky and Wang (2006), firm pairs located in the same county exhibit an excess return comovement in the next month. Nevertheless, the addition of the same-county dummy variable does not affect the return correlation predictability of our MCS measure.

Furthermore, Hoberg and Phillips (2010, 2016) construct a text-based industry classification based on the production description of 10,000 annual reports and find that this classification is more informative than traditional ones such as SIC or NAICS. We control this possible channel of similarity in the production of firm pairs for return comovement by including the text-based industry similarity score in our regression. The economic magnitude of the coefficient of our MCS measure is slightly weaker but remains statistically significant at the 1% level. Finally, Lee et al. (2015) find that stocks that investors co-search in the EDGAR system exhibit higher fundamental similarities than traditional industry classifications. As a measure of the attention paid by investors to EDGAR searches, the co-search measure could subsume our MCS in predicting the excess return comovement, so we include it in our regression model. Again, the return correlation predictability of our media connection measure remains significant. Overall, our results are robust to these alternative sources of stock return comovement.

Our paper contributes to several strands in the literature. First, we contribute to the debate on whether stock comovement is caused by information or noise. On the one hand, stock return comovement is driven by similarities in fundamentals such as size and book-to-market (Fama and French (1993), cash flows (Chen et al. (2019), price level (Green and Hwang (2009)), and analyst coverage (Chan and Hameed (2006); Muslu et al. (2014); Hameed et al. (2015)). On the other hand, some studies attribute the excess return comovement that cannot be explained by fundamental commonalities to behavioral biases such as categorical trading (Barberis and Shleifer (2003); Barberis et al. (2005)), sentiment (Kumar and Lee (2006); Kumar et al. (2013); Kumar et al. (2016)), investor attention (Peng and Xiong (2006); Huang et al. (2019)), forecast errors (Israelsen (2016)), and trading pressure from institutional investors (Anton and Polk (2014)). Our paper shows that stock return comovement for media-connected stock pairs is driven by the common fundamental information conveyed by media reports. Moreover, our results are consistent with the predictions based on the models of profit-maximizing information producers (Veldkamp (2006)). In particular, information production is non-rival with a high fixed cost of discovery and a low marginal cost of replication. Therefore, competitive producers such as news media tend to provide news that can maximize their profits from investors. Moreover, the information useful for predicting a subset of stocks could attract more investors than information that is useful for predicting only a single stock. As a consequence, given the high per-unit cost of information production, information useful for more stocks would induce news media to extend their coverage to these stocks, especially those whose fundamentals correlate more with each other.

Second, this study contributes to the literature on lead-lag effects in the returns of economically related stocks. Early studies tend to group firms according to their fundamental characteristics so as to find economically related firms, such as Lo and MacKinlay (1990), Brennan et al. (1993), Badrinath et al. (1995) and Chordia and Swaminathan (2000), while recent studies focus on specific economic links to identity firm connections, such as Cohen and Frazzini (2008), Lee et al. (2019), and Ali and Hirshleifer (2019). In this paper, we propose another way of identifying firm peers based on media connection, which can drive return lead-lag effects among firms via information spillover across media connection.

Third, we contribute to the literature that investigates the role of news media in financial markets ((Huang et al., 2021). Existing studies largely focus on two perspectives, one of which is to show that news tone or sentiment can be used to predict a firms future performance. Tetlock (2007) argues that language, especially negative language, could be used to predict excess market returns. Tetlock et al. (2008) analyze firm-specific news to explore the predictability of cross-sectional return. Another perspective examines whether media coverage can strongly affect future stock returns. For example, Fang and Per-

Table 1
Descriptive Statistics. This table presents the summary statistics of our sample over the sample period from 1996:01 to 2014:12. The sample is restricted to common stocks listed on NYSE, AMEX, and NASDAQ. We consider a stock that is covered by media news if there is more than one news article covering this stock during the past 12 months. Reported are the proportion of stocks with common news articles in a month, the average number of days with common news articles for connected stocks in a month, the average number of common news articles for connected stocks in a month, the proportion of stock-pairs with common news articles in a month, the average number of days with common news articles for connected stock-pairs in a month, the average number of common news articles for connected stock-pairs in a month, the proportion of news articles mentioning two or more firms, and the average number of firms mentioned in those common news articles.

	Mean	Median	Std. Dev.	Min.	Max.
% of stocks with common news articles	23.23	21.63	9.13	13.31	46.92
# of days with common news articles for connected stocks	3.64	3.56	0.29	2.97	4.47
# of common news articles for connected stocks	13.80	13.02	3.03	9.06	23.99
% of stock-pairs with common news articles	1.32	1.12	0.64	0.65	3.56
# of days with common news articles for stock-pairs	1.51	1.53	0.19	1.13	3.11
# of common news articles for stock-pairs	3.87	3.73	0.92	2.46	21.45
% of news articles covering two or more firms	15.49	14.30	5.54	9.15	31.27
# of firms covered in common news articles	2.92	2.60	0.66	2.24	4.74

ess (2009) find that stocks with no media coverage can earn higher returns than stocks with high media coverage, even after controlling for well-known risk factors. Engelberg and Parsons (2011) find that local media coverage can strongly predict local trading. However, there is limited evidence for the interactive effect induced by common media coverage on return comovement. In this paper, we construct a proxy for fundamental similarities between two media-connected firms through media connection, and our empirical results suggest that our MCS measure is significantly associated with fundamental similarities and return comovement.

The rest of the paper is organized as follows. Section 2 describes the data used in this paper and explains how the main variables are constructed. In Section 3, we present the main empirical results. Alternative explanations are examined and robustness tests are conducted in Section 4 to rule out alternative hypotheses. Section 5 concludes.

2. Data and main variables

We obtain stock return data from CRSP, accounting data from Compustat, and analyst coverage data from IBES. Institutional holding data is obtained from Thomson Reuters Institutional Holdings (13F) database and mutual fund holdings data is collected from the Thomson Reuters Mutual Fund Holdings database. Our data sample includes all common stocks listed on the NYSE, AMEX, and NASDAQ.

The media news data used in this paper is collected from TRNA over the period from January 1996 to December 2014. This data set contains the specific stock IDs of TRNA and RICs, which can be matched to the ticker symbols of stocks and then PERMNOs, the dates on which news articles are released, unique news IDs, along with three news tone scores indicative of whether and to what extent a news event may produce a positive, neutral, or negative effect on the stock prices in each news article.[2]

Table 1 presents the summary statistics of our sample stocks covered in the media. We consider stocks covered by the media during the past 12 months. The time-series average proportion of individual stocks with common news articles is around 23.23%, suggesting that it is common for the stocks in our sample to be mentioned with others in the same news article. For these stocks, the average number of days with common news articles is around 3.64 in a certain month, with around 13.80 common news articles reported per month. For media-connected stock pairs, the average proportion of stock pairs with common news articles relative to all possible pairs of media-covered stocks (including those with zero common news) is around 1.32%. Further, the average number of common news (days) per month for each stock pair is 3.87 (1.51). Finally, we find that the average proportion of those common news articles is about 15.49%, and the average number of firms covered in these articles is 2.92, suggesting that the connected stock pairs are a non-trivial part of the overall sample.

In our empirical analysis, we define the media connection strength (MCS) of each stock pair as the natural logarithm of the number of co-mentioned news articles in a month. In the event that multiple stocks may be mentioned in the same news article, this piece of news is defined as common media coverage and the co-cited stocks are treated as stocks linked to this piece of news.[3]

Panel A of Table 2 reports the summary statistics of the MCS. Overall, there are 2,369,731 stock pair observations in our sample with valid measures of MCS. On average, the value of MCS is 1.34, which is the log-transformed number of 3.87 from Table 1. Moreover, the MCS occurs on a frequent basis with more than half of stock-pairs have three or four common news articles per month.

[2] These three probabilities are calculated through a neural-network sentiment engine from Thomson Reuters; the sum of scores equals to 1.

[3] In our entire sample, the proportions of news articles mentioning one firm, two firms, three firms, and more than three firms are 84.6%, 9.1%, 3.2%, and 3.1%, respectively. Our results remain similar if we restrict the news articles to those that cover only two firms.

Table 2

Stock-level Descriptive Statistics. This table reports the summary statistics of our main variables over the sample period from 1996:01 to 2014:12. Panel A presents the summary statistics for media connection strength (MCS). Panel B presents the summary statistics for fundamental similarities, correlations, and stock return correlations of stock-pairs. Panel C reports the summary statistics for other control variables. The correlation of fundamental information is calculated as the correlation between firm-pairs' fundamental information over a future 20-quarter period. The fundamental similarity is defined as the negative of the absolute difference in the percentile ranking for the fundamental information. The return correlations in Panel B are computed monthly using daily raw returns (RAW), daily residuals (FF3) from the Fama and French (1993) three-factor model, and daily residuals (FF5) from the Fama and French (2015) five-factor model.

	N	Mean	Median	Std. Dev.	Min.	Max.
Panel A: Main Variables						
Media Connection Strength (MCS)	2,369,731	1.34	1.31	0.20	0.89	3.04
Panel B: Fundamental Similarities, Correlations, and Stock Return Correlations						
Correlation of Raw Returns (COR_RAW)	2,346,795	0.32	0.33	0.26	-0.67	0.88
Correlation of Abnormal Returns under FF3 (COR_FF3)	2,346,795	0.07	0.07	0.28	-0.80	0.81
Correlation of Abnormal Returns under FF5 (COR_FF5)	2,346,795	0.07	0.06	0.29	-0.82	0.81
Correlation of SUE (COR_SUE)	2,346,891	0.06	0.03	0.24	-0.91	0.93
Similarity in SUE (SAMESUE)	2,248,823	-26.58	-22.30	20.19	-96.74	0.00
Correlation of sales growth (COR_SALEG)	2,282,071	0.15	0.15	0.42	-0.89	0.91
Similarity in sales growth (SAMESALEG)	2,226,701	-28.05	-23.24	21.45	-97.13	0.00
Panel C: Control Variables						
Common Mutual Fund Ownership (FCAP)	2,369,731	0.03	0.02	0.03	0.00	0.21
Common Analyst Coverage (CAC)	2,369,731	2.86	0.15	5.64	0.00	41.79
Similarity in Book to Market (SAMEBEME)	2,107,890	-25.16	-19.95	20.42	-94.30	0.00
Similarity in Size (SAMESIZE)	2,357,672	-12.53	-7.09	14.81	-93.62	0.00
Similarity in Momentum (SAMEMOM)	2,252,786	-28.60	-23.37	22.29	-97.19	0.00
Similarity in Turnover (SAMETO)	2,369,528	-19.66	-15.94	15.80	-90.02	0.00
Similarity in Forecast Dispersion ($Simi_{dispersion}$)	1,957,647	-28.81	-24.32	21.62	-96.92	0.00
Similarity in Forecast Consensus ($Simi_{consensus}$)	2,281,857	-28.57	-22.36	23.55	-97.71	0.00
Similarity in Forecast Revisions ($Simi_{revision}$)	2,290,967	-28.02	-23.59	21.40	-96.57	0.00
Similarity in $CAR_{t-252,t-31}$ ($Simi_{CAR11M}$)	2,369,349	-30.57	-25.42	23.31	-98.10	0.00
Similarity in $CAR_{t-30,t-3}$ ($Simi_{CAR30D}$)	2,218,205	-28.82	-24.39	21.46	-97.31	0.00
Similarity in AR_{t-2} ($Simi_{AR2}$)	2,227,763	-29.45	-25.37	21.34	-96.41	0.00
Past Return Correlation (RETCOR)	1,983,742	0.30	0.29	0.20	-0.38	0.94
Past Profitability Correlation (ROECOR)	2,081,521	0.07	0.06	0.32	-0.76	0.94
Past Abnormal Trading Volumn Correlation (VOLCOR)	1,899,638	0.19	0.18	0.22	-0.50	0.92
Absolute Value Diff. in Leverage Ratio (DLEVER)	2,201,816	0.16	0.12	0.15	0.00	0.80
Absolute Value Diff. in Log Price (DPRC)	2,357,669	0.88	0.69	0.76	0.00	5.77
Same S&P 500 Index Dummy (DINDEX)	2,369,731	0.42	0.30	0.48	0.00	1.00
Same Exchange Listing Dummy (DLIST)	2,369,731	0.65	1.00	0.47	0.00	1.00
Industry Similarity (DINDUSTRY)	2,369,731	0.20	0.00	0.39	0.00	1.00

The main dependent variable in our paper, the stock excess return comovement, is defined as the co-variation among stock pairs beyond what can be accounted for based on fundamental or systematic factors. In this study, the fundamental or systematic factors include the FF3 (Fama and French (1993)) and FF5 factors (Fama and French (2015)). In each month, the excess comovement is calculated as the correlation of regression residuals of daily excess returns on the factor models among stock pairs. We require at least 15 non-missing daily returns to be present in each month for the calculation of return comovement.

More specifically, excess return comovement, $\rho_{i,j}$, is calculated as the correlation of the daily residuals between two stocks based on the asset pricing model:

$$R_{i,t} - r_{f,t} = \alpha + \sum_{k=1}^{N} \beta_{i,k} \sum_{k=1}^{N} f_{k,t} + \epsilon_{i,t}, \tag{1}$$

where $\epsilon_{i,t}$ is the daily residual of stock i on day t in terms of the FF3 or FF5-factor model. Next, the correlation is calculated as

$$\rho_{i,j} = \frac{\sum_t^T \epsilon_{i,t}\epsilon_{j,t}}{\sqrt{\sum_t^T \epsilon_{i,t}^2 \sum_t^T \epsilon_{j,t}^2}}, \tag{2}$$

To draw a comparison, a raw return comovement is also calculated as the correlation of raw returns in excess of the risk-free rate between two connected firms. Panel B of Table 2 presents a summary of the statistics for the three return comovement measures. Based on the raw returns, the average stock return comovement is approximately 0.32. The excess return comovement shows a sharp decline to an average value of 0.07 based on the FF3-factor model and 0.07 based on the FF5-factor model, suggesting that the asset pricing models we use capture the comovement of fundamental characteristics to some extent. When the raw returns are used, roughly 15% of the stock pairs experience negative return comovement, while the number increases to 44% for FF3 abnormal returns and 43% for FF5 abnormal returns, respectively.

Tetlock et al. (2008) argue that information contained in the news is able to predict the corporate earnings and stock returns controlling for traditional accounting measures of firms' fundamentals. Therefore, to verify whether our MCS indeed captures fundamental comovement between media-connected firm pairs, we employ two kinds of fundamental comovement measures; namely the standard unexpected earnings surprise (SUE) and the corporate sales growth (SALEG). Following Livnat and Mendenhall (2006), SUE is defined as the earnings at quarter t minus that at quarter t-4, scaled by the price at quarter t. A firm's SALEG is defined as the quarterly growth of sales (SALEQ from Quarterly Compustat) of the firm. Then, the fundamental comovement is calculated as the correlation between the fundamental information of two connected firms over the future 20 quarters. We require at least 15 non-missing observations to calculate the correlations of fundamentals. We also define alternative fundamental similarities as the negative of the absolute difference in the percentile ranking of SUE or SALEG for each stock pair in each quarter. By means of construction, the smaller these measures are, the more fundamentally similar the two media-connected firms are in respect of SUE or SALEG. Panel B of Table 2 presents the summary statistics of these fundamental measures. Overall, the magnitude of the fundamental correlations COR_SUE and COR_SALEG are similar to the return correlations. The average correlation of SUE (COR_SUE) is roughly 0.06, while that of sales growth (COR_SALEG) is approximately 0.15.

In previous studies, it has been demonstrated that the common ownership by a mutual fund also contributes to the excess comovement of stock returns. At the end of each quarter of the sample period, following Anton and Polk (2014), we measure common ownership as follows:

$$FCAP_{i,j,t} = \frac{\sum_{f=1}^{F}(s_{i,t}^{f}p_{i,t} + s_{j,t}^{f}p_{j,t})}{s_{i,t}p_{i,t} + s_{j,t}p_{j,t}}$$

where $s_{i,t}^{f}$ denotes the number of shares held by fund f at time t, $p_{i,t}$ indicates the trading price at time t, and $s_{i,t}$ represents the total shares outstanding of stock i at time t. The denominator $s_{i,t}p_{i,t} + s_{j,t}p_{j,t}$ refers to the total market capitalization of stock i and stock j, while the numerator $s_{i,t}^{f}p_{i,t} + s_{j,t}^{f}p_{j,t}$ is defined as the total value of shares held by fund f of the stock pair. For stock-] pairs with missing FCAP, we also define the FCAP for such stock pairs as zero. Panel B of Table 2 presents the summary statistics of FCAP. In general, the common ownership of mutual funds (FCAP) is relatively frequent as about 77% of the media-connected stock pairs have at least one common mutual fund owner. On average, 3% of the combined market capitalization of the media-connected stock pairs is owned by common mutual funds, which is slightly larger than the figure reported (0.77% on average) in Anton and Polk (2014), which is consistent with our expectation, considering that the number of mutual funds tends to have risen sharply over the most recent decades. In addition, our paper covers a longer sample period through 2014 compared with that of Anton and Polk (2014), which ended in 2008.

Analysts tend to cover similar stocks, and the errors made by them in earnings forecasts tend to be correlated among stocks within their portfolios (Israelsen (2016)), which results in an excess return comovement. Specifically, common analyst coverage (CAC) is measured as the number of analysts that have issued at least one annual earnings forecast for both stock i and stock j during the 12-month period preceding each month t. For stock pairs with missing CAC, we impose the CAC as zero. The second row in Panel C in Table 2 reports the summary statistics of this CAC. On average, this variable is about 2.86. Nevertheless, the distribution of this value is highly skewed with a standard deviation of 5.64.

We also calculate various similarities in the fundamental information to examine whether MCS is capable of predicting the fundamental or return comovement between connected firms when controlling for these fundamental similarities. Since style characteristics are not completely consistent with regression loadings (Daniel and Titman (1997)), similarity measures such as SAMESIZE, SAMEBEME, and SAMEMOM are also constructed by calculating the negative of the absolute difference in the percentile ranking for the market capitalization, book-to-market ratio, and momentum, respectively. In addition, we follow Tetlock et al. (2008) and choose variables such as TO and three measures of recent abnormal stock returns ($CAR_{t-252,t-31}$, $CAR_{t-30,t-3}$ and AR_{t-2}), analysts forecast dispersion, their forecast consensus, and their earnings forecast revisions. $CAR_{t-252,t-31}$ is defined as the cumulative abnormal return of a firm over the previous calendar year, with the most recent month ignored. The benchmark return is defined as the Fama and French (1993) three-factor model over an estimation window of [252,31] trading days prior to the earnings announcement. $CAR_{t-30,t-3}$ and AR_{t-2} represent the cumulative abnormal return from the window of [30,3] trading days and the abnormal return on day 2 based on the same benchmark return over an estimation window of [252,31] trading days prior to the earnings announcement. Analysts forecast dispersion is defined as the standard deviation of analysts earnings forecasts over the most recent time period prior to the announcement scaled by the average absolute value of the earnings forecasts. Analysts' forecast consensus is calculated as the median value of analyst earnings forecasts in each month. Analysts earnings forecast revision is defined as the three-month earnings forecast revision, which is calculated as the three-month sum of changes in the median forecast scaled by the firms stock price in the prior month. The fundamental similarities between connected stocks are calculated as the negative of the absolute difference in the percentile ranking for the aforementioned firm-specific characteristics.

Finally, we control for variables that are potential sources of return comovement beyond what is mentioned above. Specifically, the following variables are included in our regression: RETCOR, which is defined as the monthly return correlation during the past five years; ROECOR, a profitability correlation covering the last 20 quarters, with ROE measured as the ratio of the 12-month earnings per share to the book value of the equity per share; VOLCOR, which measures the correlation of volume between the media-connected stocks based on abnormal trading volume in the past five years; DLEVER, which refers to the absolute value difference in leverage ratio and is defined as the total long-term debt (Compustat item

LTT) scaled by the total assets (Compustat item AT); DPRC, which is the absolute value difference between two media-connected stocks in their log share price; DINEX, which is a dummy variable used to decide whether two media-connected stocks fall into the S&P 500 index; DLIST, which determines whether two media-connected stocks are listed on the same stock exchange; and DINDUSTRY, which indicates whether two media-connected stocks are in the same Fama-French 48 industry or not.

Panel C of Table 2 presents the summary statistics of those control variables. Basically, the statistics of these variables are consistent with the variables in Anton and Polk (2014). The above controls are updated on a quarterly basis if calculated using quarterly Compustat data or on a monthly basis if calculated using monthly CRSP data. It is worth noting that these control variables are not used to directly perform regression analysis in our study, except for dummy variables. Instead, they are normalized in such a way that control variables are made comparable across cross-sections.[4]

3. Empirical results

3.1. Cross-Sectional determinants of media connection strength

We start our empirical analysis by examining whether stocks with similar fundamentals are more likely to be mentioned in the same news article. Therefore, we investigate what determines the MCS to understand the mechanism between common media coverage and firm-pair characteristics. Specifically, we run Fama and MacBeth (1973) regressions of MCS on firm-pair characteristics mentioned in data section. To see the improvement of R^2, we gradually add independent variables from Column (1) to Column (4), all of which are rank-transformed and normalized to zero mean and a standard deviation of one.

Table 3 presents the time-series average coefficients and Newey and West (1987) adjusted t-statistics of these variables. Not surprisingly, in Column (1), we find that the common ownership by mutual fund (FCAP) and common analyst coverage (CAC) are positively and significantly associated with MCS, suggesting that firm pairs held by the same investor or covered by the same analyst are more likely to be cited in the same news article. In Column (2), the similarities in size (SAMESIZE), momentum (SAMEMOM), and sales growth (SAMESALEG) are also positively and significantly associated with MCS, while that in turnover (SAMETO) is negatively and significantly associated with MCS. In Column (4), when we include the correlation of trading volume (VOLCOR) in the regression, the coefficient becomes positive and significant, suggesting that the positive effect of the VOLCOR using past 5-year historical information on the MCS is more consistent with our intuition. In addition, in Column (4), when we include all the independent variables, the signs of some variables are negative and inconsistent with our expectation. Given that some of these variables are likely to be correlated, multicollinearity might exist in this regression. We also run univariate regressions of the MCS on each of independent variables to examine the relationship between the MCS and these firm-pair characteristics. In Column (5), the results show that all these variables explain the MCS in a way consistent with our intuition. For example, the coefficient of the similarity of the earnings surprise (SAMESUE) in Column (4) is negative and significant, while its coefficient in Column (5) is positive and significant. Overall, these results suggest that our MCS indeed captures the fundamental comovement of stock-pairs.

3.2. Forecasting fundamental comovement

In this section, we examine whether or not the MCS can predict fundamental correlations or similarities when controlling for other characteristics at the firm-pair level. Specifically, we perform quarterly Fama and MacBeth (1973) regressions of fundamental correlations or similarities for the stock pairs on MCS and a list of control variables. We use SUE or SALEG to measure fundamental performance, as each of them is effective in capturing the fundamental changes in a firm without requiring measurement of return.

Table 4 presents the results of the Fama and MacBeth (1973) regressions of the correlations of earnings surprise (COR_SUE) calculated using future five-year information and the similarity of the earnings surprise (SAMESUE) on the lagged MCS between co-mentioned stocks and other controls over the sample period from 1996 to 2014. Apart from that, in order to control the potential persistence of fundamental similarities, lagged COR_SUE or lagged SAMESUE ($Dep._{t-1}$) is included in our regressions. With SUE correlations as dependent variables, the first two columns in Table 4 present the results of contemporaneous and predictive regressions. As expected, the earnings surprise of stock pairs become more significantly correlated as a result of an increased MCS in both the current and following months. A one standard deviation increase in the MCS leads to an increase of 0.6% in COR_SUE with a t-value of 7.533 in the current month, and an increase of 1.0% with a t-value of 5.655 in the following month. Given that the cross-sectional median value of COR_SUE is 0.03, the impact of MCS on the earnings comovement is of significance both statistically and economically. The next two columns report the results of contemporaneous and predictive regressions with SAMESUE as a dependent variable. Again, the positive and significant coefficients of MCS suggest that it is able to predict the fundamental comovement of stock-pairs controlling for comprehensive fundamental similarities. With regard to other fundamental similarities, the results are consistent with our

[4] We transform these variables following Anton and Polk (2014). First, the percentile ranking of each stock-pair is calculated on a particular stock pair characteristic in each cross section. Then, the rankings are normalized to obtain zero mean and a standard deviation of one.

Table 3
Cross-sectional Determinants of Media Connection Strength (MCS). This table reports the Fama and MacBeth (1973) estimates of monthly cross-sectional regressions of MCS on various control variables. Column (1) to (4) present results of multivariate regressions and Column (5) presents results of univariate regressions of the MCS on each independent variable. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

	DepVar: MCS				
	Multivariate regression				Univariate regression
	(1)	(2)	(3)	(4)	(5)
FCAP	0.833***	0.397***	0.387***	0.180***	0.984***
	(17.518)	(10.131)	(9.532)	(3.931)	(19.591)
CAC	1.024***	1.069***	1.099***	0.960***	1.138***
	(21.888)	(20.856)	(22.414)	(19.934)	(22.481)
SAMEBEME		0.002	-0.002	-0.038	0.305***
		(0.083)	(-0.088)	(-1.496)	(12.081)
SAMESIZE		0.697***	0.774***	0.700***	0.987***
		(19.577)	(22.375)	(20.218)	(23.482)
SAMEMOM		0.062***	0.041*	0.037*	0.347***
		(3.407)	(1.822)	(1.795)	(19.528)
SAMETO		-0.053***	-0.077***	-0.090***	0.275***
		(-2.971)	(-4.262)	(-5.342)	(10.314)
SAMESALEG		0.040**	0.034*	0.029*	0.263***
		(2.309)	(1.942)	(1.655)	(17.332)
SAMESUE		-0.017	-0.023	-0.045***	0.232***
		(-1.039)	(-1.498)	(-2.762)	(12.723)
$Simi_{dispersion}$			0.065***	0.051***	0.304***
			(4.006)	(3.008)	(15.715)
$Simi_{consensus}$			-0.097***	-0.031	0.342***
			(-4.888)	(-1.473)	(17.436)
$Simi_{revision}$			0.018	0.004	0.280***
			(0.955)	(0.227)	(14.541)
$Simi_{CAR11M}$			0.046**	-0.003	0.303***
			(2.297)	(-0.153)	(18.578)
$Simi_{CAR30D}$			0.042***	0.017	0.263***
			(2.843)	(1.138)	(17.409)
$Simi_{AR2}$			-0.062***	-0.053***	0.061***
			(-4.111)	(-3.320)	(4.549)
RETCOR				0.230***	0.967***
				(6.760)	(23.313)
ROECOR				0.169***	0.388***
				(9.978)	(13.627)
VOLCOR				0.405***	0.962***
				(12.731)	(16.635)
DLEVER				-0.175***	-0.444***
				(-10.251)	(-16.097)
DPRC				0.223***	-0.225***
				(7.866)	(-10.827)
DINDEX				0.651***	1.743***
				(8.485)	(19.152)
DLIST				-0.339***	0.589***
				(-4.123)	(11.018)
DINDUSTRY				0.361***	0.413***
				(5.468)	(7.265)
N	2,369,731	1,987,137	1,579,577	1,441,364	
Adj. R^2	0.036	0.045	0.050	0.062	

expectation that their SUE values tend to be closer to one another if these firms are more similar in certain specific fundamental information.

In a robustness test, SALEG is also treated as the proxy of a firm's fundamental performance. In particular, the sales growth correlation (COR_SALEG) is calculated over the future 20 quarters. In addition, a SAMESALEG analogous to SAMESUE is calculated. Then, Fama and MacBeth (1973) regressions are repeated by replacing COR_SUE and SAMESUE in Table 4 with COR_SALEG and SAMESALEG, respectively. Table 5 details the results. Similar to the results shown in Table 4, the MCS shows a significantly positive association with both current and future fundamental comovement for media-connected stocks. Overall, the evidence presented so far substantiates our view that the complex correlated fundamentals contained in the media can provide incremental information about fundamental comovement for media-connected stock pairs when controlling for a list of fundamental similarities.

Table 4
MCS and Stock SUE Comovement. This table reports the Fama and MacBeth (1973) estimates of quarterly cross-sectional regressions of SUE correlation or similarity in SUE on MCS and other controls. The SUE correlation (COR_SUE) at quarter t is computed quarterly over the future 5-year window from quarter t. 15 valid observations with non-missing information are required for calculation. The similarity in SUE (SAMESUE) is defined as the negative of the absolute difference in the percentile ranking for the SUE between connected stocks. SUE is calculated as earnings at quarter t minus earnings at quarter t-4, scaled by price at quarter t following Livnat and Mendenhall (2006). The first two columns present results of contemporaneous regressions and predictive regressions of the SUE correlation. The next two columns present results of contemporaneous regressions and predictive regressions of the similarity in SUE. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

	DepVar: COR_SUE		DepVar: SAMESUE	
	Quarter t	Quarter t+1	Quarter t	Quarter t+1
MCS	0.006***	0.010***	0.123**	0.035**
	(7.533)	(5.655)	(2.098)	(1.981)
$Dep._{t-1}$	0.017***	0.027***	1.982***	2.005***
	(15.455)	(11.307)	(37.632)	(33.869)
FCAP	0.003***	0.007***	0.180***	0.163**
	(4.832)	(6.060)	(2.835)	(2.558)
CAC	0.025***	0.028***	0.037	0.154**
	(21.078)	(16.628)	(0.619)	(2.259)
SAMESIZE	-0.005***	-0.002	1.567***	1.819***
	(-4.446)	(-1.168)	(14.040)	(14.241)
SAMEBEME	0.003***	0.004***	1.525***	1.432***
	(8.815)	(5.547)	(24.445)	(19.060)
SAMETO	0.002***	0.000	-0.407***	-0.341***
	(3.669)	(0.213)	(-10.553)	(-9.429)
$Simi_{dispersion}$	0.004***	0.006***	1.184***	1.084***
	(12.504)	(8.806)	(21.769)	(16.234)
$Simi_{consensus}$	0.005***	0.005***	0.837***	0.721***
	(5.948)	(2.725)	(7.699)	(7.666)
$Simi_{revision}$	0.003***	0.009***	3.058***	2.921***
	(5.882)	(10.578)	(22.290)	(24.001)
$Simi_{CAR11M}$	0.006***	0.008***	1.257***	1.179***
	(17.651)	(12.323)	(13.532)	(14.526)
$Simi_{CAR30D}$	0.002***	0.003***	0.496***	0.517***
	(5.015)	(4.081)	(10.783)	(11.413)
$Simi_{AR2}$	-0.000	-0.000	0.335***	0.412***
	(-0.449)	(-0.617)	(7.847)	(12.174)
Intercept	0.057***	0.072***	-26.353***	-26.556***
	(19.077)	(10.065)	(-55.919)	(-59.315)
N	1,563,212	1,496,823	1,530,349	1,419,965
Adj. R^2	0.029	0.023	0.095	0.090

3.3. Univariate analysis

The main hypothesis in our paper is that stocks mentioned in the same news article tend to comove in their stock returns. Therefore, we conduct simple univariate analysis to verify this argument. Specifically, we sort stock pairs with non-zero MCS into tercile groups based on the raw value of MCS in each month; we then compute an equal-weighted return correlation of these stock pairs in each tercile group in the same month. To examine if stock pairs with higher MCS tend to exhibit higher return comovement, we also calculate a time series of the differences in the return correlations between the group with the highest MCS and that with the lowest MCS. Finally, we also investigate if the return correlations among stock pairs with non-zero MCS are higher than those among stock pairs with zero MCS.

Table 6 reports the results. Panel A shows that stock pairs with the highest MCS tend to comove more in their stock returns compared to those with lower MCS. For example, the return correlation for stock pairs with the highest MCS (MCS3) is 0.39 for raw returns, 0.12 for excess returns under FF3, and 0.11 for excess returns under FF5, while for stock pairs with the lowest MCS (MCS1), these values are 0.28, 0.05, and 0.05, respectively. Moreover, all the differences in the correlations between connected stocks in MCS3 and MCS1 are statistically significant at the 1% level. Panel B of Table 6 shows that the return correlations for stock pairs with non-zero MCS are significantly higher than the return correlations for stock pairs

Table 5
MCS and Stock Sales Growth Comovement. This table reports the Fama and MacBeth (1973) estimates of quarterly cross-sectional regressions of sales growth correlation or similarity in sales growth between pair of stocks on MCS and other controls. The sales growth correlation (COR_SALEG) at quarter t is computed quarterly over the future 5-year window from quarter t. 15 valid observations with non-missing information are required for calculation. The similarity in sales growth (SAMESALEG) is defined as the negative of the absolute difference in the percentile ranking for the sales growth between connected stocks. Sales growth is calculated as quarterly change of sales. The first two columns present results of contemporaneous regressions and predictive regressions of the sales growth correlation. The next two columns present results of contemporaneous regressions and predictive regressions of the similarity in sales growth. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

	DepVar: COR_SALEG		DepVar: SAMESALEG	
	Quarter t	Quarter t+1	Quarter t	Quarter t+1
MCS	0.018***	0.019***	0.533***	0.010***
	(9.775)	(9.985)	(7.294)	(6.174)
$Dep._{t-1}$	0.045***	0.045***	1.369***	0.088***
	(34.236)	(31.350)	(14.388)	(37.326)
FCAP	0.003**	0.003**	1.002***	0.033***
	(2.153)	(2.151)	(14.142)	(9.145)
CAC	0.050***	0.050***	2.060***	0.094***
	(23.486)	(22.349)	(37.294)	(18.368)
SAMESIZE	-0.001	-0.001	0.206***	0.014***
	(-1.001)	(-1.047)	(3.534)	(4.242)
SAMEBEME	0.005***	0.005***	0.152***	0.004
	(4.992)	(4.417)	(2.867)	(1.599)
SAMETO	-0.002***	-0.002***	-0.260***	-0.008***
	(-4.298)	(-4.151)	(-5.834)	(-4.244)
$Simi_{dispersion}$	0.002***	0.002**	0.182***	0.012***
	(2.815)	(2.128)	(4.375)	(4.571)
$Simi_{consensus}$	0.002***	0.002**	0.447***	0.025***
	(2.710)	(2.353)	(8.093)	(8.587)
$Simi_{revision}$	0.010***	0.009***	1.008***	0.029***
	(9.794)	(9.110)	(16.290)	(11.719)
$Simi_{CAR11M}$	0.005***	0.005***	0.631***	0.025***
	(6.035)	(5.334)	(12.594)	(13.585)
$Simi_{CAR30D}$	0.005***	0.005***	0.428***	0.024***
	(6.450)	(7.507)	(16.612)	(10.919)
$Simi_{AR2}$	0.001**	0.001**	0.283***	0.013***
	(2.430)	(2.562)	(8.031)	(8.463)
Intercept	0.124***	0.125***	-28.880***	-0.003
	(22.838)	(21.516)	(-86.712)	(-1.316)
N	1,535,112	1,500,219	1,586,320	1,468,398
Adj. R^2	0.040	0.039	0.033	0.033

with zero MCS. Overall, these results indicate that media-connected stocks tend to exhibit stronger return comovement than those that are not connected with media news.

3.4. Forecasting excess return comovement

Considering that the MCS is positively associated with the fundamental comovement of stock pairs, it it expected to be associated with stock return comovement as well. To verify our argument, we conduct Fama and MacBeth (1973) regressions by regressing the monthly return comovement on the MCS between the paired stocks. In the regression model, the other alternative fundamental similarity measures discussed in Section 2 are included to examine whether our MCS is able to capture the excess return comovement beyond what can be explained by other fundamental similarities. In addition to contemporaneous regression analysis, we also conduct predictive regressions by lagging all independent variables one month to predict the return comovement of the following month. Table 7 shows the results of our regressions performed under various specifications, so as to understand the return comovement based on the excess returns and the risk adjusted returns under the FF3-factor and FF5-fator models.

The first three columns of Table 7 present the contemporaneous regression results. In all of the three cases, the coefficients of the MCS show a high level of positive significance. A one standard deviation increase in MCS is associated with

Table 6
Univariate Analysis. This table reports the monthly average return correlations for groups formed by the media connection strength (MCS). Panel A reports the average return correlations for the tercile groups sorted by the MCS, as well as the differences of correlations between tercile 3 (MCS3) and tercile 1 (MCS1). Panel B reports the average return correlations for stocks with non-zero MCS and zero MCS, as well as the differences of correlations between these two groups. The return correlations are computed monthly using daily raw returns (RAW), daily residuals (FF3) from the Fama and French (1993) three-factor model, and daily residuals (FF5) from the Fama and French (2015) five-factor model.

	Number of News	RAW	FF3	FF5
Panel A: Sample split into terciles by MCS				
MCS1	1.06	0.28	0.05	0.05
MCS2	2.32	0.32	0.07	0.07
MCS3	11.1	0.39	0.12	0.11
MCS3-MCS1	10.05***	0.11***	0.07***	0.06***
	(37.23)	(9.55)	(11.22)	(10.97)
Panel B: Sample split into zero MCS and non-zero MCS				
Zero MCS	0	0.24	0.01	0.01
Non-zero MCS	3.87	0.32	0.07	0.07
Difference	3.87***	0.08***	0.06***	0.06***
	(58.56)	(8.98)	(14.51)	(15.40)

an increase of 0.9% in return comovement and an excess return comovement of 1.0% (under the FF3) and 1.0% (under the FF5). The Newey and West (1987) adjusted t-statistics, which are shown in parentheses, indicate that the coefficients are statistically significant at the 1% level. Moreover, consistent with previous studies, both FCAP and CAC are positively related to the excess return comovement at the same time. For example, the coefficients of the FCAP are 1.0% for returns adjusted using Fama and French (1993) three-factor model and 0.9% for returns adjusted using Fama and French (2015) five-factor model, all of which are significant at the 1% level. As for CAC, the coefficients show a high level of positive significance in all three cases. A one standard deviation increase in CAC is associated with an increase of 2.1% in return comovement and an excessive comovement of 4.5% (under the FF3) and 3.9% (under the FF5), which are consistent with the results in Israelsen (2016).

In the next three columns in Table 7, the results of predictive regressions are presented. The coefficients of the MCS remain significantly positive regardless of the exact empirical settings. As for economic significance, with the FF3-factor model applied, the fitted value ranges from an average minimum of 5.51% to an average maximum of 16.85%, and the average of excess return correlation reaches 7%. In comparison, with the FF5-factor model applied, the fitted value ranges from 5.33% to 16.75% with an average excess return correlation of 7%, suggesting that the impact of MCS on the excess return comovement is of significance both statistically and economically. In the FF3-factor and FF5-factor models, the FCAP predicts excess return comovement with statistically significant coefficients ranging from 0.7% to 1.0% (t-statistics of 6.390 to 10.106), respectively, which is consistent with the finding in Anton and Polk (2014) that stocks owned by the same mutual fund tend to move at the same pace. The coefficients on CAC are also positive and statistically significant, forecasting return excess comovement with values of 4.6% and 4.2% (adjusted t-statistics of 27.697 and 28.630) under the Fama and French (1993) three-factor model and the Fama and French (2015) five-factor model, respectively.

Regarding the results of other control variables, most control coefficients exhibit significant signs, as expected. For example, the coefficients on SAMEMOM and SAMEBEME are all positive and highly significant, while the coefficients on SAMESIZE are negative and significant solely under the settings of the Fama and French (1993) three-factor model and Fama and French (2015) five-factor model. As for the characteristics-based correlations, RETCOR, ROECOR, and VOLCOR are all positively associated with excess return comovement at the 1% level. Finally, a pair of stocks in the same industry (DINDUSTRY) are more likely to co-move than those in different industries in the current and the following month, as all coefficients are positive and statistically significant.

If our hypothesis related to the correlated fundamentals contained in common media coverage is substantiated, then the return comovement predictability of the MCS is supposed to be independent of investor sentiment. We use the Baker and Wurgler (2006) sentiment index to distinguish between high- and low-sentiment periods. Specifically, the period of high (low) sentiment is defined as the sentiment index in the previous month that is above (below) the median value of sentiment index throughout the whole sample period. Then, the predictive regressions in these two sub-sample periods are re-examined to investigate whether or not the predictability of MCS is affected by the investor sentiment. Table 8 presents the regression results, which confirm that the coefficients of MCS barely change compared with the results throughout the whole sample period, suggesting that the return comovement induced by media connection is largely attributed to the correlated fundamentals contained in the common media coverage rather than the irrational biases held by investors.

Table 7
MCS and Stock Return Comovement. This table reports the Fama and MacBeth (1973) estimates of monthly cross-sectional regressions of the return correlation between pair of stocks on MCS and other controls. The return correlations are computed monthly using daily raw returns (RAW), daily residuals (FF3) from the Fama and French (1993) 3-factor model, and daily residuals (FF5) from the Fama and French (2015) 5-factor model. 15 valid observations with non-missing returns are required for calculation. The first three columns present results of contemporaneous regressions and next three columns present results of predictive regressions. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

	Month t			Month t+1		
	RAW	FF3	FF5	RAW	FF3	FF5
MCS	0.009***	0.010***	0.010***	0.011***	0.011***	0.011***
	(8.562)	(9.976)	(9.890)	(11.073)	(10.253)	(10.821)
FCAP	0.006***	0.010***	0.009***	0.007***	0.010***	0.009***
	(6.442)	(9.452)	(9.883)	(6.390)	(10.106)	(9.410)
CAC	0.021***	0.045***	0.039***	0.024***	0.046***	0.042***
	(11.334)	(26.365)	(28.332)	(12.609)	(27.697)	(28.630)
SAMEBEME	0.004***	0.002***	0.001*	0.004***	0.001***	0.001
	(5.737)	(3.541)	(1.732)	(6.649)	(3.047)	(1.400)
SAMESIZE	0.019***	-0.007***	-0.006***	0.017***	-0.007***	-0.006***
	(15.510)	(-6.992)	(-5.631)	(13.542)	(-6.405)	(-5.159)
SAMEMOM	0.016***	0.013***	0.011***	0.016***	0.013***	0.011***
	(20.736)	(19.556)	(17.463)	(20.736)	(19.556)	(18.114)
SAMETO	0.009***	0.011***	0.009***	0.010***	0.012***	0.008***
	(14.663)	(19.360)	(13.116)	(13.956)	(22.810)	(14.262)
RETCOR	0.062***	0.041***	0.033***	0.058***	0.036***	0.030***
	(33.662)	(27.425)	(23.871)	(31.728)	(26.681)	(24.017)
ROECOR	0.008***	0.009***	0.007***	0.008***	0.008***	0.007***
	(22.316)	(15.663)	(15.713)	(21.579)	(15.428)	(14.347)
VOLCOR	0.020***	0.011***	0.011***	0.021***	0.012***	0.011***
	(19.769)	(19.992)	(22.636)	(20.381)	(19.045)	(20.260)
DLEVER	-0.009***	-0.002***	-0.001***	-0.008***	-0.002***	-0.001***
	(-14.361)	(-3.453)	(-3.063)	(-12.295)	(-3.905)	(-2.864)
DPRC	-0.012***	-0.011***	-0.010***	-0.013***	-0.010***	-0.009***
	(-12.360)	(-14.112)	(-13.942)	(-13.546)	(-13.483)	(-14.884)
DINDEX	0.014***	-0.019***	-0.020***	0.010***	-0.018***	-0.018***
	(4.113)	(-12.964)	(-13.003)	(3.774)	(-10.709)	(-11.482)
DLIST	0.010***	0.015***	0.015***	0.001	0.016***	0.016***
	(3.426)	(10.674)	(12.336)	(0.497)	(9.986)	(12.086)
DINDUSTRY	0.040***	0.084***	0.084***	0.041***	0.085***	0.082***
	(15.113)	(23.068)	(24.446)	(14.261)	(21.681)	(22.735)
Intercept	0.331***	0.043***	0.039***	0.302***	0.039***	0.032***
	(31.625)	(14.663)	(13.876)	(29.230)	(12.620)	(12.078)
N	1,760,874	1,760,874	1,760,874	1,757,479	1,757,479	1,757,479
Adj. R^2	0.223	0.144	0.128	0.233	0.169	0.134

3.5. MCS And trading activities comovement

A potential mechanism behind our findings is that investors receiving news from Thomson Reuters may trade the stocks mentioned in the same news article, and their trading behavior on correlated fundamentals will lead to comovement in stock returns. This hypothesis is verified by examining whether the MCS can be used to predict the trading activity comovement of media-connected stock pairs. To measure trading activity comovement, the daily order flow is first calculated for each stock following Lee and Ready (1991), and then a monthly order flow correlation is constructed for each stock pair by calculating the correlations of order flows between two stocks. Similar to the regression setting indicated in Table 7, the dependent variable is replaced by the correlations of the order flow of stock pairs.

Panel A of Table 9 shows our Fama and MacBeth (1973) regression results with the correlations of the order flow as the dependent variables. In the first two columns, the contemporaneous correlations of the order flow are regressed on the same variables as those in Table 7. The coefficients of MCS are positive and statistically significant in explaining the current trading comovement with t-statistics above 5, thus confirming that common media coverage can attract attention from investors and prompt their correlated trading activities on media-connected stock pairs. In the last two columns, the correlations of the order flow are regressed on lagged MCS and other lagged control variables. The coefficients of MCS remain significantly positive in predicting the comovement of trading activities in the following month. The effects of an increase in MCS on the order flow comovement between two stocks are 0.7% without controls and 0.4% with controls, respectively.

Table 8
Return Comovement during High or Low Sentiment Period. This table reports the Fama and MacBeth (1973) estimates of monthly cross-sectional regressions of the future return correlation between pair of stocks on MCS and other controls during high or low sentiment periods. The return correlations are computed monthly using daily raw returns (RAW), daily residuals (FF3) from the Fama and French (1993) 3-factor model, and daily residuals (FF5) from the Fama and French (2015) 5-factor model. 15 valid observations with non-missing returns are required for calculation. The first three columns present results of regressions during low sentiment period and next three columns present results of regressions during high sentiment period. The low or high sentiment is defined based on the median value of the Baker and Wurgler (2006) sentiment index across whole sample period. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

	Low sentiment			High sentiment		
	RAW	FF3	FF5	RAW	FF3	FF5
MCS	0.010***	0.011***	0.011***	0.011***	0.011***	0.011***
	(6.945)	(8.532)	(8.275)	(8.160)	(7.273)	(7.740)
FCAP	0.007***	0.009***	0.009***	0.008***	0.010***	0.010***
	(4.512)	(7.808)	(6.813)	(5.131)	(7.535)	(7.075)
CAC	0.019***	0.042***	0.039***	0.027***	0.048***	0.045***
	(10.541)	(24.989)	(25.975)	(10.228)	(21.048)	(21.604)
SAMEBEME	0.002***	0.001	0.000	0.005***	0.002***	0.001
	(2.747)	(1.171)	(0.442)	(6.583)	(2.956)	(1.376)
SAMESIZE	0.017***	-0.005***	-0.004***	0.017***	-0.009***	-0.007***
	(8.126)	(-3.888)	(-2.710)	(9.845)	(-5.128)	(-4.381)
SAMEMOM	0.014***	0.011***	0.008***	0.018***	0.014***	0.013***
	(10.758)	(12.857)	(8.990)	(18.147)	(14.401)	(13.634)
SAMETO	0.009***	0.013***	0.010***	0.010***	0.011***	0.007***
	(10.147)	(17.557)	(11.934)	(9.025)	(17.898)	(10.416)
RETCOR	0.061***	0.036***	0.030***	0.055***	0.036***	0.031***
	(33.437)	(18.308)	(15.230)	(20.056)	(20.997)	(20.172)
ROECOR	0.009***	0.006***	0.005***	0.008***	0.010***	0.009***
	(13.596)	(9.070)	(7.883)	(16.879)	(14.842)	(13.986)
VOLCOR	0.021***	0.010***	0.009***	0.021***	0.014***	0.013***
	(13.281)	(15.527)	(16.898)	(17.153)	(15.645)	(15.583)
DLEVER	-0.008***	-0.003***	-0.003***	-0.009***	-0.001	-0.000
	(-9.452)	(-4.789)	(-5.225)	(-9.401)	(-1.623)	(-0.523)
DPRC	-0.014***	-0.010***	-0.009***	-0.012***	-0.010***	-0.010***
	(-12.056)	(-14.265)	(-13.974)	(-9.858)	(-8.917)	(-9.771)
DINDEX	0.007	-0.022***	-0.022***	0.013***	-0.015***	-0.015***
	(1.511)	(-9.767)	(-9.277)	(3.909)	(-6.708)	(-7.581)
DLIST	-0.002	0.014***	0.013***	0.003	0.018***	0.018***
	(-0.527)	(8.514)	(7.687)	(0.849)	(7.558)	(10.672)
DINDUSTRY	0.038***	0.090***	0.085***	0.044***	0.082***	0.081***
	(8.231)	(18.073)	(18.409)	(12.815)	(16.069)	(17.088)
Intercept	0.325***	0.034***	0.029***	0.286***	0.043***	0.034***
	(21.808)	(8.479)	(7.534)	(23.264)	(10.285)	(9.991)
N	889,696	889,696	889,696	867,783	867,783	867,783
Adj. R^2	0.221	0.153	0.119	0.241	0.180	0.144

Additionally, following Lo and Wang (2000), we calculate share turnover as a proxy of trading activities. Specifically, the turnover is calculated as the logarithm of the number of the shares traded divided by the total number of shares outstanding each day. Similarly, the correlation of the turnover ratio is calculated for each month using the daily turnover ratio. Panel B of Table 9 presents the Fama and MacBeth (1973) regression results with the correlations of the turnover ratio as the dependent variables. The MCS positively predicts the correlation of the turnover ratio of stock pairs in month t and month t+1. The coefficients of MCS range from 2.0% to 4.3%, with t-statistics above 15. Based on the significant association between the MCS and the comovement of trading activities, we can conclude that investors trade on the correlated fundamental information contained in news items, thus generating comovement in stock returns for the stocks mentioned in the same news items.

4. Alternative explanations

4.1. MCS And category trading habitat

Barberis and Shleifer (2003) argue that investors allocate capital at the level of asset categories rather than individual stocks. Firms are mentioned frequently in the same news articles, which may be because they are viewed as being in

Table 9
MCS and Trading Activity Comovement. This table reports the Fama and MacBeth (1973) estimates of monthly cross-sectional regressions of the trading activity comovement on MCS and other controls. The measure of trading activity comovement is the correlation of order flow in Panel A and correlation of turnover ratio in Panel B. Order flow is constructed following Lee and Ready (1991). Turnover ratio is defined as the logarithm of number of shares traded divided by total number of shares outstanding. All control variables are same as in Table 7. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

Panel A: DepVar: Correlation of order flow

	Month t		Month t+1	
MCS	0.006***	0.003***	0.007***	0.004***
	(9.367)	(5.929)	(9.116)	(6.339)
FCAP		0.004***		0.004***
		(5.809)		(5.751)
CAC		0.001***		0.001**
		(2.877)		(1.941)
Intercept	0.010***	0.005***	0.011***	0.005***
	(7.471)	(4.812)	(9.663)	(3.613)
Other Controls	NO	YES	NO	YES
N	1,961,979	1,468,970	1,892,066	1,457,878
Adj. R^2	0.001	0.011	0.002	0.012

Panel B: DepVar: Correlation of turnover

	Month t		Month t+1	
MCS	0.046***	0.026***	0.043***	0.020***
	(21.442)	(16.334)	(19.362)	(18.881)
FCAP		0.025***		0.013***
		(17.883)		(15.316)
CAC		0.006***		0.009***
		(6.997)		(10.432)
Intercept	0.161***	0.202***	0.111***	0.153***
	(22.346)	(31.633)	(22.563)	(24.362)
Other Controls	NO	YES	NO	YES
N	2,346,523	1,749,362	2,346,510	1,747,752
Adj. R^2	0.053	0.088	0.099	0.174

the same asset category by market participants including news media and investors. Alternatively, investors view of asset categories could be influenced by the frequency of stocks being mentioned together in news articles. In either case, investors trading in and out of media-connected stocks that are viewed as in the same category could induce excessive comovement in their returns beyond fundamentals. Given that our MCS can predict investors' correlated trading, one concern is that our findings are due to the alternative hypothesis that investors simply trade on media-connected stock pairs even when there is no many fundamental similarities among those stock pairs.

We conduct several analyses to address the concern about category trading habitat as the mechanism of the MCS measure, so as to predict return comovement. First, as is argued by Barberis et al. (2005), if some investors are noise traders that conduct category trading due to correlated sentiment, then their correlated trading should induce excessive comovement regardless of any fundamental similarities among stock pairs. This suggests that if our MCS indeed captures investors' category trading habitat, we should see a stronger effect of the MCS on return comovement during high-sentiment periods. However, the results in Table 8 show that the predictability of our MCS remains basically the same during periods of both low and high sentiment. Hence, the mechanism for the use of the MCS measure to predict return comovement is not likely to be category trading induced from non-fundamental biases, such as correlated sentiment.

In addition, the inclusion of FCAP in the regressions of Table 7, which captures investors' category trading habitat to some extent, suggests that the correlated trading does not affect the predictability of the MCS. Nevertheless, this measure may not be sufficient to exclude the category trading explanation. Thus, we repeat the regressions in Table 7 by replacing the correlations of order flow (OFCOR) or the correlations of turnover in one month (TOCOR) with the similarities in turnover (SAMETO) and the correlations of turnover in the preceding 60 months (VOLCOR) to control for the correlated flow-induced trading caused by category trading habitat. The correlated trading can be better captured in time by the construction of TOCOR than of VOLCOR, which is calculated based on historical information in the past 60 months. Table 10 shows that after controlling for these two direct measures of category trading, the regression coefficients of MCS for return comovement are still positive and highly significant, which is not different from the coefficients of MCS for return comovement in Table 7.

Table 10
MCS and Categorical Trading Habitat. This table reports the Fama and MacBeth (1973) estimates of monthly cross-sectional regressions of the return correlation between pair of stocks on MCS, proxies of categorical trading, and other controls. The measure of categorical trading is the correlation of order flow in Panel A and correlation of turnover ratio in Panel B. Order flow is constructed following Lee and Ready (1991). Turnover ratio is defined as the logarithm of number of shares traded divided by total number of shares outstanding. All control variables except SAMETO are the same as in Table 7. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

Panel A: Controlling for correlation of order flow (OFCOR)

	Month t		Month t+1	
	FF3	FF5	FF3	FF5
MCS	0.011***	0.011***	0.010***	0.011***
	(9.356)	(10.211)	(10.312)	(10.146)
FCAP	0.012***	0.011***	0.011***	0.011***
	(11.647)	(10.750)	(10.003)	(9.109)
CAC	0.048***	0.044***	0.045***	0.044***
	(29.509)	(30.409)	(27.498)	(25.452)
OFCOR	0.005***	0.005***	0.005***	0.005***
	(7.860)	(7.625)	(7.448)	(7.775)
Intercept	0.043***	0.036***	0.038***	0.036***
	(15.136)	(14.413)	(12.629)	(11.469)
Other Controls	YES	YES	YES	YES
N	1,467,213	1,467,213	1,454,835	1,454,835
Adj. R^2	0.171	0.136	0.182	0.144

Panel B: Controlling for correlation of turnover (TOCOR)

	Month t		Month t+1	
	FF3	FF5	FF3	FF5
MCS	0.011***	0.011***	0.010***	0.011***
	(10.114)	(10.103)	(10.234)	(10.006)
FCAP	0.012***	0.011***	0.011***	0.011***
	(11.788)	(10.952)	(9.621)	(9.336)
CAC	0.048***	0.044***	0.045***	0.044***
	(29.518)	(30.415)	(27.463)	(25.835)
TOCOR	0.001***	0.001***	0.001***	0.001***
	(4.061)	(3.444)	(4.145)	(3.733)
Intercept	0.043***	0.035***	0.038***	0.035***
	(15.136)	(14.185)	(12.263)	(11.332)
Other Controls	YES	YES	YES	YES
N	1,749,312	1,749,312	1,746,836	1,746,836
Adj. R^2	0.170	0.135	0.181	0.143

These results indicate that the category trading hypothesis is not likely to be the main mechanism of the predictive power of our MCS measure.

We conduct an additional analysis to show that a non-fundamental-based category trading hypothesis seems implausible. If a comovement is caused by media connection through category trading instead of correlated fundamental information, the category trading activity may temporarily lead the prices of two media-connected stocks to move above (below) their fundamental value at the same time through the comovement in price pressure, leading to overprice (underprice) at this period and a possible price reversal in the next period. Hence, we construct a trading strategy based on the principle that we buy (sell) a stock that has gone up (down) together with its peer stocks jointly covered by multi-firm news articles. More specifically, for a given month, we first define a media-connected portfolio return for each stock as the weighted average return of the media-connected peers, whose weight is the number of multi-firm news articles covering this focal stock and a given peer stock. Then at the end of each month, we independently sort stocks into quintiles based on the focal stock's own return and its media-connected portfolio return. We calculate the equally weighted returns on these 25 composite portfolios for the next month.

Table 11 presents the Fama and French (2015) five-factor alphas on these 25 composite portfolios, which generally increase as one moves from low to high connected returns within an own-return quintile. Anton and Polk (2014) argue that the excessive return comovement induced by common mutual fund ownership is driven by investors' correlated trading demands and exploit a similar trading strategy. As is shown in their Table V, alphas on the composite portfolios increase from high to low mutual-fund-connected stock returns, which strongly supports their argument that common ownership

Table 11
MCS-based Trading Strategies This table presents the profitability of a trading strategy exploiting media connection. We independently sort stocks into quintiles based on their own returns over the last month and the returns of their media-connected portfolio over the last month. The media-connected portfolio return is defined as the MCS-weighted return of the media-connected stocks. ALL portfolio returns are equal-weighted and rebalanced on a monthly basis. The table reports the Fama and French (2015) five-factor alphas on these 25 composite portfolios. We also report the average alphas on a connected-stock trading strategy (HH-LL) that buys the high own-return high connected-return composite portfolio (HH) and sells the low own-return low connected-return composite portfolio (LL). Newey and West (1987) adjusted t-statistics are presented in parentheses.

		Own return						
		Low	2	3	4	High	H-L	
Connected portfolio return	Low	-0.10	-0.21	-0.26	-0.34	-0.35	-0.25	
		(-0.35)	(-0.92)	(-1.12)	(-1.64)	(-1.50)	(-1.55)	
	2	-0.05	-0.14	-0.12	-0.30	-0.41	-0.36	
		(-0.19)	(-0.82)	(-0.62)	(-1.76)	(-1.34)	(-1.36)	
	3	0.21	0.20	0.24	0.05	-0.20	-0.41	
		(1.44)	(1.30)	(1.34)	(0.33)	(-0.67)	(-1.87)	
	4	0.36	0.22	0.15	0.20	0.19	-0.17	
		(2.48)	(1.68)	(0.80)	(1.61)	(1.26)	(-0.55)	
	High	0.63	0.23	0.34	0.33	0.31	-0.32	
		(2.52)	(1.83)	(2.05)	(1.96)	(2.07)	(-1.46)	
	H-L	0.73	0.44	0.60	0.67	0.66	0.41	HH-LL
		(2.66)	(1.88)	(2.33)	(2.53)	(2.31)	(2.08)	

causes comovement through price pressure. However, our result is opposite to that of Anton and Polk (2014), indicating that correlated trading demand caused by category trading is unlikely to be the underlying mechanism in our case.

In addition, within each connected-return quintile, the alphas on these composite portfolios generally decrease as one moves from low to high own-return portfolios. Our MCS-based trading strategy, which we denote by HH-LL, buys the composite high portfolio (high own return and high connected portfolio return) and sells the composite low portfolio (low own return and low connected portfolio return). The five-factor alpha of this strategy is 0.41% per month with a t-statistic of 2.08, which is inconsistent with the expected price reversal, as is mentioned above under the category trading mechanism. Overall, these results suggest that investor category trading habitat is unlikely to be the mechanism for the result of our MCS measure, even though we may not be able to completely rule it out.[5]

4.2. Robustness tests on other possible explanations

So far, the results show that stock returns of firms mentioned in the same piece of news exhibit a high degree of return comovement. In this section, based on the existing literature, we further conduct robustness tests to rule out other possible explanations of the stock return comovement.

Pirinsky and Wang (2006) reveal that the stock prices tend to co-move when their headquarters are located at the same place. Therefore, a dummy variable is introduced to our regression to verify whether our results will be affected if two stocks mentioned in the same piece of news are headquartered in the same U.S. county. In Column 1 of Table 12, consistent with Pirinsky and Wang (2006), firms located in the same county tend to exhibit excess return comovement in the following month, and the economic magnitude is comparable to that of the MCS. However, the addition of this dummy indicator has no effect on the predictability of the MCS in return correlation.

Hoberg and Phillips (2010, 2016) construct a new text-based industry classification according to the production description of 10,000 annual reports, which leads to a finding that this classification is more informative than such traditional industry classifications as SIC or NAICS when various firm characteristics are explained, including profitability, Tobin's Q and dividends. Our MCS is also able to capture such a dimension. Using this data, we control for two test-based industry similarity scores whose calibrations of the classifications are as granular as the two-digit and three-digit SIC codes in our regression models to investigate whether our findings can be accounted for by the similarities of products. As shown in Columns 2 and 3 of Table 12, the economic magnitude attached to the coefficient of our MCS measure is weaker, while the coefficient remains significant at the 1% level. Furthermore, consistent with the findings of Hoberg and Phillips (2010, 2016), the product similarity score is closely associated with excess return comovement.

Finally, Lee et al. (2015) reveal that stocks co-searched by the investors on the EDGAR website have a higher fundamental similarity than traditional industry classifications. As the measure of EDGAR co-search on firm pairs by investors can subsume our measure of MCS as the predictor of excess return comovement, this co-search measure is incorporated into

[5] Israelsen (2016) establishes the casual link between excess return comovement and common analyst coverage based on exogenous changes in commonality in analyst coverage around brokerage firm mergers. Unfortunately, in our case, we do not have the data for exogenous shocks in common media coverage.

Table 12
Robustness. This table reports the Fama and MacBeth (1973) estimates of monthly cross-sectional regressions of the abnormal return correlations between pair of stocks on MCS and other controls. The abnormal return correlation is computed monthly using daily residuals from regressing daily raw return on daily Fama and French (2015) five-factor model. 15 valid observations with non-missing returns are required for calculation. Same County is an dummy variable indicating if two stocks are located in a county using firm's headquarter location. TNIC3 (calibrated to be as granular as three-digit SIC codes) and TNIC2 (calibrated to be as granular as two-digit SIC codes) are the text-based network industry classifications from Hoberg and Phillips (2010) and Hoberg and Phillips (2016). EDGAR is the number of common searches on a pair of stocks from EDGAR log file. All independent variables except dummy variables are rank-transformed and normalized to have unit standard deviation. Newey and West (1987) adjusted t-statistics are presented in parentheses. ***, ** and * indicate significance at 1%, 5%, and 10% levels, respectively.

	DepVar: abnormal return correlation					
	(1)	(2)	(3)	(4)	(5)	(6)
MCS	0.011***	0.007***	0.006***	0.010***	0.008***	0.006***
	(10.578)	(4.598)	(3.271)	(9.829)	(4.809)	(3.940)
FCAP	0.009***	0.013***	0.018***	0.013***	0.012***	0.013***
	(9.401)	(7.510)	(7.923)	(13.648)	(6.446)	(4.960)
CAC	0.042***	0.047***	0.054***	0.048***	0.039***	0.045***
	(28.520)	(21.121)	(22.742)	(240.632)	(24.106)	(19.731)
Same County	0.012***				-0.010***	-0.010**
	(6.401)				(-2.830)	(-2.535)
TNIC2		0.020***			0.021***	
		(14.528)			(11.446)	
TNIC3			0.016***			0.016***
			(12.187)			(10.534)
EDGAR				0.008***	0.008***	0.008***
				(5.197)	(5.649)	(5.142)
Intercept	0.032***	0.072***	0.072***	0.048***	0.069***	0.068***
	(12.033)	(18.952)	(14.847)	(13.690)	(20.919)	(17.035)
Other Controls	YES	YES	YES	YES	YES	YES
N	1,757,479	397,597	269,566	1,555,418	335,790	227,519
Adj. R^2	0.134	0.171	0.174	0.112	0.164	0.170

our regression. However, the return correlation predictability of our MCS remains significant controlling for this EDGAR co-search measure. In the final two columns of Table 12, all variables are included in the regression model, which shows that our MCS remains statistically significant when all these variables are subject to control, although with a weaker economic magnitude. Overall, our results are robust to these alternative sources of stock return comovement.

5. Conclusions

In this study, based on news data from Thomson Reuters, we construct a media connection strength (MCS) measure across two given firms, which is defined as the number of multi-firm news articles covering those two firms. We find that the MCS measure is correlated with fundamental similarity and can significantly explain and forecast return comovement for media-connected firm pairs. Further analyses show that our results are less likely to be driven by investors' category trading activities or various other alternative channels. Overall, we illustrate that media news connection can be a proxy for (complex) fundamental similarities across firms not captured by various documented variables, and this may hence provide a new mechanism explaining return comovement beyond what has been proposed in existing studies.

References

Ali, U., Hirshleifer, D., 2019. Shared analyst coverage: unifying momentum spillover effects. J financ econ.
Anton, M., Polk, C., 2014. Connected stocks. Journal of Finance 69 (3), 1099–1127.
Badrinath, S.G., Kale, J.R., Noe, T.H., 1995. Of shepherds, sheep, and the cross-autocorrelations in equity returns. Review of Financial Studies 8 (2), 401–430.
Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61 (4), 1645–1680.
Barberis, N., Shleifer, A., 2003. Style investing. J financ econ 68 (2), 161–199.
Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. J financ econ 75 (2), 283–317.
Bekaert, G., Harvey, C.R., Ng, A., 2005. Market integration and contagion. Journal of Business 78 (1).
Brennan, M.J., Jegadeesh, N., Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. Review of Financial Studies 6 (4), 799–824.
Brooks, C., Henry, O.T., Persand, G., 2002. The effect of asymmetries on optimal hedge ratios. Journal of Business 75 (2), 333–352.
Chan, K., Hameed, A., 2006. Stock price synchronicity and analyst coverage in emerging markets. J financ econ 80 (1), 115–147.
Chen, H., Chen, S., Chen, Z., Li, F., 2019. Empirical investigation of an equity pairs trading strategy. Manage Sci 65 (1), 370–389.
Chordia, T., Swaminathan, B., 2000. Trading volume and cross-autocorrelations in stock returns. Journal of Finance 55 (2), 913–935.
Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. Journal of Finance 63 (4), 1977–2011.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52 (1), 1–33.
Engelberg, J.E., Parsons, C.A., 2011. The causal impact of media in financial markets. Journal of Finance 66 (1), 67–97.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J financ econ 33 (1), 3–56.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J financ econ 116 (1), 1–22.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81 (3), 607–636.
Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. Journal of Finance 64 (5), 2023–2052.
Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G., 2006. Pairs trading: performance of a relative-value arbitrage rule. Review of Financial Studies 19 (3), 797–827.
Green, T.C., Hwang, B.-H., 2009. Price-based return comovement. J financ econ 93 (1), 37–50.
Hameed, A., Morck, R., Shen, J., Yeung, B., 2015. Information, analysts, and stock return comovement. Review of Financial Studies 28 (11), 3153–3187.
Hoberg, G., Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. Review of Financial Studies 23 (10), 3773–3811.
Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. Journal of Political Economy 124 (5), 1423–1465.
Huang, S., Huang, Y., Lin, T.-C., 2019. Attention allocation and return co-movement: evidence from repeated natural experiments. J financ econ 132 (2), 369–383.
Huang, H., Sun, Y., Chu, Q., 2021. Can we-media information disclosure drive listed companies' innovation? — From the perspective of financing constraints. China Finance Review International.
Israelsen, R.D., 2016. Does common analyst coverage explain excess comovement? Journal of Financial and Quantitative Analysis 51 (4), 1193–1229.
Kallberg, J., Pasquariello, P., 2008. Time-series and cross-sectional excess comovement in stock indexes. Journal of Empirical Finance 15 (3), 481–502.
Karolyi, G.A., Stulz, R.M., 1996. Why do markets move together? an investigation of us-japan stock return comovements. Journal of Finance 51 (3), 951–986.
Kumar, A., Lee, C.M., 2006. Retail investor sentiment and return comovements. Journal of Finance 61 (5), 2451–2486.
Kumar, A., Page, J.K., Spalt, O.G., 2013. Investor sentiment and return comovements: evidence from stock splits and headquarters changes. Rev Financ 17 (3), 921–953.
Kumar, A., Page, J.K., Spalt, O.G., 2016. Gambling and comovement. Journal of Financial and Quantitative Analysis 51 (1), 85–111.
Lee, C.M., Ma, P., Wang, C.C., 2015. Search-based peer firms: aggregating investor perceptions through internet co-searches. J financ econ 116 (2), 410–431.
Lee, C.M., Ready, M.J., 1991. Inferring trade direction from intraday data. Journal of Finance 46 (2), 733–746.
Lee, C.M., Sun, S.T., Wang, R., Zhang, R., 2019. Technological links and predictable returns. J financ econ 132 (3), 76–96.
Livnat, J., Mendenhall, R.R., 2006. Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. Journal of Accounting Research 44 (1), 177–205.
Lo, A.W., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3 (2), 175–205.
Lo, A.W., Wang, J., 2000. Trading volume: definitions, data analysis, and implications of portfolio theory. Review of Financial Studies 13 (2), 257–300.
Muslu, V., Rebello, M., Xu, Y., 2014. Sell-side analyst research and stock comovement. Journal of Accounting Research 52 (4), 911–954.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Papadakis, G., Wysocki, P., 2007. Pairs trading and accounting information. SSRN Working Paper.
Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. J financ econ 80 (3), 563–602.
Philippe, J., 2001. Value at risk: the new benchmark for managing financial risk. NY: McGraw-Hill Professional.
Pirinsky, C., Wang, Q., 2006. Does corporate headquarters location matter for stock returns? Journal of Finance 61 (4), 1991–2015.
Qian, E.E., Hua, R.H., Sorensen, E.H., 2007. Quantitative equity portfolio management: Modern techniques and applications. CRC Press.
Rosenberg, J.V., Schuermann, T., 2006. A general approach to integrated risk management with skewed, fat-tailed risks. J financ econ 79 (3), 569–614.
Schwenkler, G., Zheng, H., 2019. The network of firms implied by the news. SSRN working paper.
Shiller, R.J., 1989. Comovements in stock prices and comovements in dividends. Journal of Finance 44 (3), 719–729.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. Journal of Finance 62 (3), 1139–1168.
Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. Journal of Finance 63 (3), 1437–1467.
Veldkamp, L.L., 2006. Information markets and the comovement of asset prices. Review of Economic Studies 73 (3), 823–845.