JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 55, No. 2, Mar. 2020, pp. 549–580
COPYRIGHT 2019, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109019000061

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

Googling Investor Sentiment around the World

Zhenyu Gao, Haohan Ren, and Bohui Zhang*

Abstract

We study how investor sentiment affects stock prices around the world. Relying on households' Google search behavior, we construct a weekly measure of sentiment for 38 countries during 2004–2014. We validate the sentiment index in tests using sports outcomes and show that the sentiment measure is a contrarian predictor of country-level market returns. Furthermore, we document an important role of global sentiment in stock markets.

I. Introduction

Investor sentiment has long been proposed to play an important role in explaining stock price variation (Keynes (1936), Shiller (1990), and Baker and Wurgler (2000)). This sentiment conjecture is supported by both the De Long, Shleifer, Summers, and Waldmann (1990) model, in which stocks are held predominantly by noise traders, and the Shleifer and Vishny (1997) theory that

*Gao, gaozhenyu@baf.cuhk.edu.hk, Chinese University of Hong Kong Business School; Ren (corresponding author), haohanren@fudan.edu.cn, Fudan University School of Management; and Zhang, bohuizhang@cuhk.edu.cn, Chinese University of Hong Kong Shenzhen Finance Institute and the School of Management and Economics. We thank Jennifer Conrad (the editor) and Byoung-Hyoun Hwang (the referee) for their insightful suggestions. We are grateful for valuable comments from Hendrik Bessembinder, Jie Cao, Zhanhui Chen, Zhiyao Chen, Joseph Cheng, Darwin Choi, Tarun Chordia, Oleg Chuprinin, Lili Dai, Stephen Geoffrey Dimmock, David Feldman, Huasheng Gao, Allaudeen Hameed, Chuan-Yang Hwang, Hyun Joong Im, Wenxi Jiang, Jun-Koo Kang, Paul Karehnke, Andrew Karolyi, Tiecheng Leng, Donghui Li, Weikai Li, Zhangxin Frank Liu, Jiang Luo, Chenghu Ma, Phong Ngo, David Ong, Heungju Park, Jerry T. Parwada, Peter Pham, Yaxuan Qi, Hong Ru, Konark Saxena, Jianfeng Shen, Qian Sun, Cameron Truong, Baolian Wang, Chunyang Wang, Cong Wang, Shaun Shuxun Wang, Chishen Wei, Takeshi Yamada, Hua Zhang, Jason Zein, Xianming Zhou, Qi Zhu, Xiaoneng Zhu, and the conference and seminar participants at the 2016 Symposium on Emerging Financial Markets: China and Beyond, the 2016 Financial Management Association Asia/Pacific Conference, the 2016 Financial Management Association Annual Meeting, the International Finance and Banking Society 2016 Conference, the 2016 Asian Financial Association annual meeting, the 2016 Vietnam International Conference in Finance, Australian National University, Fudan University, Jinan University, Macquarie University, Nanyang Technological University, Norwegian School of Economics, Peking University, the Royal Melbourne Institute of Technology, Sungkyunkwan University, Shanghai University of Finance and Economics, the Chinese University of Hong Kong, and University of New South Wales, Australia. We thank Simon Huang and Yun Xue for their excellent research assistance. Gao acknowledges research support from the Early Career Scheme (ECS) (24503515).

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

rational investors are limited by arbitrage constraints. Empirically, this conjecture has been tested in the influential works of Baker and Wurgler (2006) and Baker, Wurgler, and Yuan (2012), who construct a composite market-based sentiment index and document its return prediction in the United States and 5 non-U.S. developed markets.[1]

However, several challenges remain in the sentiment literature. First, investor sentiment is difficult to measure (Baker and Wurgler (2007)). For example, Sibley, Wang, Xing, and Zhang (2016) argue that the return prediction of the Baker and Wurgler (2006) sentiment index is mainly driven by business-cycle and risk components.[2] Second, there is a lack of direct evidence that the return prediction of sentiment indices is tailored to sentiment theories. For example, considerable anecdotal evidence of sentiment can be explained by rational models (e.g., Pástor and Veronesi (2003), (2005), (2006)). Third, previous evidence is constrained by time frequency and market coverage. For example, Huang, Jiang, Tu, and Zhou (2015) summarize that the return prediction of investor sentiment is mainly observed at horizons of 1 year or longer,[3] and Baker et al. (2012) find few studies that investigate sentiment outside the United States.

Given the increasing interest in investor sentiment among economists, we work on the preceding issues by using alternative sentiment measures and conducting out-of-sample analyses. We measure investor sentiment based on households' Google search behavior (via Google Trends) and explore the sentiment effect at a weekly frequency in a large sample of 38 countries between 2004 and 2014. As the most popular search engine around the world, Google provides an ideal platform for our cross-country study to measure investor sentiment and examine its price effect.[4] Google searches not only reflect the attitudes of market participants but also reveal information in a timely manner.[5] For example, scientists can detect influenza epidemics by using Google search queries (Ginsberg, Mohebbi, Patel, Brammer, Smolinski, and Brilliant (2009)).

Motivated by the Financial and Economic Attitudes Revealed by Search (FEARS) index of Da, Engelberg, and Gao (2015), we first construct an economics-related sentiment index by aggregating the volume of search queries

[1] Stambaugh, Yu, and Yuan (2012) study the effect of investor sentiment on a broad set of anomalies in cross-sectional stock returns. Their findings are confirmed by Stambaugh et al. (2014).

[2] Qiu and Welch ((2006), p. 29) raise a concise question: "How does one test a theory that is about inputs→outputs with an output measure?"

[3] However, theories make the common assumption that noise traders cannot survive in the long run (Kogan, Ross, Wang, and Westerfield (2006), (2017)).

[4] For example, according to NetMarketShare's statistics in 2014, Google holds a 67.5% global market share in household searches. Though censored or banned in some countries, such as China, Google dominates most developed and emerging markets, such as Australia (93.0%), Brazil (96.9%), India (97.0%), Japan (75.3%), and Poland (97.4%).

[5] According to Google statistics for 2014, the top search terms regarding China included "China largest economy," "China overtakes U.S. economy," "China passes U.S. economy," and so on. However, after the stock market rout in June 2015, the top search terms changed substantially to "China economy collapse," "China economy crash," and "China economy crisis." See Curran, E. "Googling China's Economy Shows Shifting Sentiment," *Bloomberg* (Dec. 20, 2015), available at http://www.bloomberg.com/news/articles/2015-12-20/googling-china-s-economy-shows-shifting-sentiment.

related to economics and finance in the predominant language for each country.[6] For example, we use "gold" as a search query for the United States, "金" for China, "or" for France, and "oro" for Spain. This economics-related index aims to capture investor sentiment triggered by economic factors, and the stock price deviates from fundemantals due to investors' underreaction or overreaction to these shocks.

Given that investor sentiment can also stem from nonfundamental influences, we advance the Da et al. (2015) approach by considering search terms unrelated to economics and finance, such as queries on weather, disasters, pollution, terrors/terrorists, holidays, and sports. These search terms are relatively less correlated with economic fundamentals. Taken together, we expect that economics-related and non-economics-related sentiment indices can complement each other and create a composite index that includes both measures to capture households' sentiment in a broader context.

Furthermore, whereas the FEARS index is based on negative sentiment, we take into account both positive and negative sentiment. We have two reasons for including positive sentiment in our sentiment measure. First, lower negative sentiment does not necessitate higher positive sentiment, given the dispersion of investor opinions (e.g., Diether, Malloy, and Scherbina (2002), Hong and Stein (2007)). Second, positive sentiment may be more important than negative sentiment in substantiating the market impact of noise traders (Yu and Yuan (2011)).

We formulate three objectives in this article: i) to validate our search-based index as a proxy for investor sentiment, ii) to examine whether the sentiment measure is a contrarian predictor of country-level market returns worldwide, and iii) to investigate whether and how sentiment travels across countries and study the return prediction of global sentiment.

We begin by providing a validation test of our search-based sentiment index. We connect our sentiment measure to international soccer results (i.e., the knockout stage match outcomes of World Cup and regional tournaments). We find that a country's sentiment index (excluding the sports category) is negatively and significantly correlated with its soccer losses. The effects mainly come from the non-economics-related component of the sentiment index and therefore validate the non-economics-based portion of our measure as a proxy for sentiment.

Further evidence supports the prevalence of sentiment in stock markets around the world. For 36 of 38 countries in our sample, we find a negative relation between sentiment and the next week's market returns. Among them, 20 countries display a significant pattern at the 5% level. In terms of economic significance, a 1-standard-deviation increase in investor sentiment predicts a decline in the weekly return of 26 basis points (bps) (an annual return of 13%). Following the return prediction of our sentiment indices, a trading strategy can generate an annual return of 11% with a Sharpe ratio of 2.0.

[6] According to Wikipedia, preceded by Chinese (14.1%) and Spanish (5.85%), English is the third most popular language in the world, with 5.52% of the world population speaking English as their native language in 2007. See https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

In addition, we show that both non-economics- and economics-related sentiment index components can independently predict market returns. Our results are robust to the use of subsamples of developed and emerging countries, the exclusion of the financial crisis period, and the use of local currency returns. Furthermore, our findings survive after controlling for a set of fundamental variables (e.g., changes in liquidity, Treasury bill rate, unemployment rate, consumer price index (CPI), term and default spreads, disposable income, consumption, industrial production, and the National Bureau of Economic Research (NBER) recession dummy in the United States) as suggested by Sibley et al. (2016), alleviating, though not eliminating, the concern that the return prediction of the sentiment index is driven by business-cycle and risk factors.[7] We also find that the difference in sentiment is positively associated with dual-listed firms' price parity deviations, which may not be explained by firm fundamentals.

Next, to mitigate potential endogeneity issues, we conduct difference-in-differences analyses around plausibly exogenous shocks to examine the two sentiment channels proposed by Baker and Wurgler (2006): *difficult-to-value channel* and *limits-to-arbitrage channel*. First, we study the implementation of the Markets in Financial Instruments Directive (MiFID) as a shock to the information environment. MiFID contains a series of harmonized rules that enhance market transparency and investor protection. Consistent with the difficult-to-value channel, we find that with the improvement in market transparency, asset valuation is less subjective, and investors are less likely to react to sentiment. Second, we employ the short-selling ban following the global financial crisis as evidence of limits to arbitrage. In response to market crashes during the financial crisis of 2007–2009, several countries banned the short selling of stocks for a while. In line with the limits-to-arbitrage channel, our sentiment measure has a larger impact on the markets affected by the short-selling ban because rational investors find it difficult to correct the mispricing caused by market sentiment, especially from the short leg. Our difference-in-differences tests further confirm that our measure captures investor sentiment across countries.

Finally, by aggregating country-level sentiment indices into a global sentiment measure, we find that global sentiment prevails in international markets. Specifically, we start by documenting the commonality of sentiment across countries. In this regard, global sentiment, on average, explains 4.7% of the sentiment variation for all countries and 6.0% for developed countries. The result supports the contagious characteristic of investor sentiment proposed by Baker et al. (2012). Economics-related sentiment indices are more contagious among countries than non-economics-related sentiment indices. One plausible explaination is that large commonality in economics-related sentiment is driven by globally linked economic activities. Furthermore, we use the adoption of International Financial Reporting Standards (IFRS) as an event study to test the effect of capital market integration on sentiment comovement. IFRS improves information sharing and capital flows across countries, which contributes to a higher degree of capital market integration (DeFond, Hu, Hung, and Li (2011),

[7]Note that we can include only the U.S. variables as a proxy for global fundamental and risk factors because other countries lack high-frequency macroeconomic information.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

De George, Li, and Shivakumar (2016)). Our results show that global sentiment has a larger positive impact on the sentiment of countries that adopted IFRS. Finally, we document that global sentiment significantly predicts future market returns in 29 countries, whereas local sentiment has significant power in predicting market returns in only 16 countries. Consistent with the impact of capital market integration on sentiment comovement, the return prediction of global sentiment is more pronounced in developed countries than in emerging countries.

Our article contributes to two streams of the literature. Our primary contribution is to the literature that quantifies investor sentiment and investigates the effect of sentiment on asset prices (Hirshleifer and Shumway (2003), Kamstra, Kramer, and Levi (2003), Brown and Cliff (2004), Qiu and Welch (2006), Baker and Wurgler (2006), (2007), Lemmon and Portniaguina (2006), Das and Chen (2007), Edmans, Garcia, and Norli (2007), Tetlock (2007), Barber, Odean, and Zhu (2009), Kaplanski and Levy (2010), Hwang (2011), Yu and Yuan (2011), Baker et al. (2012), Stambaugh et al. (2012), (2014), Huang et al. (2015), Da et al. (2015), Sibley et al. (2016), and Soo (2018)). Our findings support the market prevalence of sentiment with direct measurement, high frequency, and broad market coverage.

Our article also contributes to the literature that studies market participant behavior by using search engines. Specifically, search engine data can be used to predict flu outbreaks (Ginsberg et al. (2009)); forecast economic activities, such as automobile sales (Choi and Varian (2012)); evaluate investor attention (Mondria, Wu, and Zhang (2010), Da, Engelberg, and Gao (2011)); measure investor information demand (Drake, Roulstone, and Thornock (2012)); and identify economically related peer firms (Lee, Ma, and Wang (2015)). In contrast to these studies, we advance the Da et al. (2015) approach by constructing a composite search-based index of investor sentiment.

II. Data and Methodology

A. Google Search

As the largest search engine in the world, Google search captures 67.5% of the global market share (see https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0). Panel A of Table 1 reports the Google search volume from 2004 to 2014 by year, month, day, and second. Globally, 2.095 trillion search queries were performed through Google in 2014, that is, 66,440 every second. This figure has increased by almost 25 times since 2004. Panel B shows the popularity of Google search across both developed and emerging countries. For example, its market share is more than 90% in 25 of the 38 countries in the sample, and astonishingly, it is as high as 99% in countries such as Thailand. Overall, the data show that Google search can serve as a powerful platform to track households' search behavior across countries.

B. Sentiment Index

Our task is to track households' search activity and measure investor sentiment for each country in our sample. We follow and improve the approach of Da et al. (2015) to construct weekly sentiment indices for 38 countries.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 1
Summary Statistics of Google Search

Table 1 reports the summary statistics of Google search for 38 countries from 2004 to 2014. Panel A presents the total Google search volume around the world during the sample period (data available at http://www.internetlivestats.com/google-search-statistics/#trend and http://www.statisticbrain.com/google-searches). Panel B presents the details of Google usage for each country. Internet user (IT_USER) from World Bank is measured as the 11-year average of Internet users per hundred people, and Google market share (G_SHR) is Google's search market share (data available at http://returnonnow.com/internet-marketing-resources/2013-search-engine-market-share-by-country and http://www.mvfglobal.com) in each country at the end of 2013. LANGUAGE denotes the languages used to construct the sentiment measure for each country. LANG_SHR denotes their corresponding population shares obtained from the World Factbook of CIA (data available at https://www.cia.gov/library/publications/the-world-factbook/fields/402.html) and the home page of each country on Wikipedia, respectively. EMG, DEV, and WRD denote emerging, developed, and global markets, respectively.

Panel A. Google Search Volume over the World

Year	Annual (10^{12})	Monthly (10^{11})	Daily (10^{9})	Second (10^{4})
2004	0.086	0.072	0.236	0.273
2005	0.141	0.118	0.386	0.447
2006	0.231	0.192	0.633	0.732
2007	0.438	0.365	1.200	1.389
2008	0.637	0.531	1.745	2.021
2009	0.954	0.795	2.610	3.024
2010	1.325	1.104	3.627	4.201
2011	1.722	1.435	4.717	5.461
2012	1.874	1.562	5.134	5.942
2013	2.162	1.801	5.922	6.854
2014	2.095	1.746	5.740	6.644

Panel B. Internet Users per 100 People and Google Market Share

Country	EMG/DEV	IT_USER	G_SHR	LANGUAGE	LANG_SHR
Argentina	EMG	39.83%	95.63%	Spanish	Official
Chile	EMG	47.20%	97.80%	Spanish	Official (99.5%)
China	EMG	29.30%	23.80%	Chinese	Official
Colombia	EMG	33.30%	96.30%	Spanish	Official (99.2%)
Egypt	EMG	30.05%	97.00%	Egyptian Arabic	Official (69.8%)
Hungary	EMG	61.10%	97.40%	Hungarian	Official (99.6%)
India	EMG	7.90%	97.00%	English	Subsidiary official
Indonesia	EMG	9.70%	96.00%	Indonesian	Official
Israel	EMG	57.20%	97.70%	Hebrew	Official
Malaysia	EMG	58.50%	82.10%	Malay	Official
Mexico	EMG	30.00%	92.60%	Spanish	Official (92.7%)
Pakistan	EMG	8.29%	94.87%	English	Official
Peru	EMG	31.20%	98.00%	Spanish	Official (84.1%)
Philippines	EMG	19.70%	81.70%	English	Official
Poland	EMG	55.60%	97.40%	Polish	Official (98.2%)
Portugal	EMG	50.10%	97.10%	Portuguese	Official
Russia	EMG	39.20%	23.30%	Russian	Official (96.3%)
South Africa	EMG	23.80%	94.80%	English	Official
South Korea	EMG	81.40%	36.90%	Korean	Official
Taiwan	EMG	65.50%	47.60%	Chinese	Official
Thailand	EMG	22.10%	99.00%	Thai	Official (90.7%)
Australia	DEV	74.50%	93.00%	English	Official (76.8%)
Austria	DEV	73.20%	92.80%	German	Official (88.6%)
Canada	DEV	79.20%	92.70%	English	Official (58.7%)
Denmark	DEV	88.60%	95.00%	Danish	Official
France	DEV	69.80%	95.00%	French	Official (100.0%)
Germany	DEV	78.70%	91.70%	German	Official
Hong Kong	DEV	68.40%	68.00%	English	Official
Italy	DEV	48.80%	86.00%	Italian	Official (99.0%)
Japan	DEV	77.90%	75.30%	Japanese	Official (99.0%)
Netherlands	DEV	89.00%	94.70%	Dutch	Official
New Zealand	DEV	76.20%	92.00%	English	English (89.8%)
Singapore	DEV	68.80%	84.40%	English	Official (29.8%)
Spain	DEV	62.20%	92.90%	Spanish	Official (74.0%)
Sweden	DEV	90.10%	96.80%	Swedish	Official
Switzerland	DEV	81.10%	92.00%	German	Official (64.9%)
United Kingdom	DEV	81.30%	89.90%	English	Official (95.0%)
United States	DEV	74.60%	80.60%	English	Official (79.2%)
Total	WRD	57.74%	86.33%	NA	NA

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

We start with a description of construction procedures and then highlight the differences between our method and the Da et al. (2015) approach.

The main data source, Google Trends (https://www.google.com/trends), provides a Search Volume Index (SVI) of search terms across various countries in different languages since 2004. To determine the sentiment toward economic conditions, we first follow Da et al. (2015) and download the weekly SVI of search terms related to economics and finance. We ensure that we search these exact terms in each country based on the country's main language.[8] We construct the set of economics- and finance-related search terms by using words from the *Harvard IV-4 Dictionary* and the *Lasswell Value Dictionary*. These two dictionaries are widely used in the text analytics literature (e.g., Tetlock (2007)).

However, one may be concerned that these economics- and finance-related search terms capture only investor sentiment triggered by fundamental movements. To mitigate this concern, we also use an approach different from that of Da et al. (2015) and consider search terms unrelated to economic fundamentals to construct the second component of the sentiment index. Specifically, we choose keywords on weather, disasters, pollution, terrors/terrorists, holidays, and sports because of extensive evidence in the psychology and finance literature showing that market sentiment can be significantly influenced by events in these categories.[9]

In the following, we describe the index construction for each country in detail. We first discuss the method based on economics- and finance-related terms and then explain the construction based on search terms unrelated to economics and finance.

1. Economics- and Finance-Related Search Terms

a) In the set of 743 words with markers "Econ@" or "ECON" (from http://www.wjh.harvard.edu/~inquirer/spreadsheet_guide.htm), we focus on words that are likely to be associated with positive or negative sentiment (those that are labeled with the "positive" or "negative" tag). This method

[8]Panel B of Table 1 presents the languages we employ to construct our sentiment indices and the corresponding population shares of these languages. We choose the predominant or official language for each country. If a country has multiple official or predominant languages, we use the one that provides us with the most weekly observations of search terms.

[9]Specifically, weather affects investor sentiment. For example, Hirshleifer and Shumway (2003) find that sunshine positively correlates with stock returns through its positive effects on investor sentiment. Disasters such as airline crashes lead to negative mood and anxiety, which negatively affect stock prices (Kaplanski and Levy (2010)). Pollution can also foster negative mood among investors. For example, evidence in the psychology literature suggests that air pollution increases depression, anxiety, and helplessness (e.g., Evans, Jacobs, Dooley, and Catalano (1987)) and thus leads to negative effects on the stock market (Levy and Yagil (2011)). Terrorism attacks increase fear and anxiety. For example, Silver, Holman, McIntosh, Poulin, and Gil-Rivas (2002) find that 17% of the U.S. population outside of New York City reported post-traumatic stress 2 months after the attacks on Sept. 11, 2001. An upcoming holiday is likely to create positive investor mood, which may significantly influence the stock market. For example, Ariel (1990) finds that preholiday stock returns are significantly higher than normal daily returns. Hirshleifer, Jiang, and Meng (2018) also provide evidence that positive investor mood swings are a possible cause of preholiday stock return seasonality. Finally, important sports outcomes also influence investor sentiment. Edmans et al. (2007) show that a loss in an important soccer match leads to negative stock market returns the next day.

provides us with 149 words, such as "bankruptcy," "cost," "gold," "jobless," and "profit."

b) Because we are interested in local households' search activities, we translate these 149 English words into each country's corresponding language by using Google Translate (see https://translate.google.com).[10] For example, when we input "gold," Google Translate returns "金" in Chinese, "or" in French, and "oro" in Spanish.

c) We input these 149 translated words into Google Trends and identify the top 10 terms associated with each word. For example, in the United States, the top queries related to "gold" include "gold price," "price of gold," and "buy gold."

d) We keep only search terms that have at least 100 weekly SVI observations and remove those that are not clearly related to economics or finance. For example, in the terms related to "gold" searched by Australian households, "Gold Coast" is a metropolitan region south of Brisbane on Australia's east coast, which is famous for its long sandy beaches and surfing; thus, we exclude this term from our list for Australia. Another concern is that Google Translate may not provide the most accurate translation for search terms with multiple meanings in different contexts. To ensure that the translation is accurate, we employ 2 certified native speakers on UPWORK, a global freelancing platform, for each of the 21 languages. These 42 native speakers help us filter out search terms that seem to be related to finance or economics but are seldom used by native speakers. They also help us identify search terms that appear to be meaningless to Google Translate but are in fact frequently used by native speakers.

e) We download the weekly SVI (covering the search volume from Sunday to Saturday) of search terms from Jan. 2004 to Dec. 2014.

f) We calculate the weekly change in SVI (ΔSVI) for each search term, winsorize extreme observations, eliminate seasonality,[11] and standardize time series to make them comparable.[12] We eventually obtain the adjusted weekly change in search volume (ΔASVI).

g) To label search terms with positive or negative sentiment and identify how these search terms are relevant to market returns, we allow the market data to speak for themselves. Specifically, we employ the expanding backward

[10] Google Translate returns only one word that is the most common translation of each input.

[11] The seasonality might be relevant given the well-known seasonal patterns of stock returns (e.g., the January effect) and holidays in constructing our index. We thus use an alternative sentiment measure without adjusting seasonality and apply it to our main prediction test in Section III. The results are significant, although the magnitude is smaller. Please refer to Table IA6 in the Supplementary Material.

[12] We winsorize ΔSVI of each query at the 5% level. We regress ΔSVI on monthly dummies to control for seasonality and keep the residuals. We then normalize these residuals by using their standard deviations.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

rolling regressions of ΔASVI on the country's market returns from the beginning of the sample to the most recent June or December. Then, we identify the historical relation between the weekly change in search volume and returns.[13] Based on a U.S. daily sample, Da et al. (2015) find no terms with significant positive *t*-statistics and thus focus on the top 30 negative terms in constructing their FEARS index. However, our weekly data provide a large number of search terms with significant positive (and negative) *t*-statistics in most countries. One explanation for this difference is that weekly frequency potentially reduces noise in daily observations. Positive sentiment plays a more important role than negative sentiment in the capital market when pessimistic investors stay on the sidelines because of short-selling constraints (e.g., Miller (1977)). We therefore account for both positive and negative terms in forming our sentiment indices.

h) We construct our sentiment indices by averaging ΔASVI for the top 30 positive and top 30 negative search terms for every week and calculate the difference as the measure of sentiment based on economics- and finance-related search terms:[14]

$$(1) \quad \text{SENTIMENT}_{\text{ECON},t} = \sum_{i=1}^{30} R_{+}^{i}(\Delta\text{ASVI}_i) - \sum_{i=1}^{30} R_{-}^{i}(\Delta\text{ASVI}_i),$$

where $\sum_{i=1}^{30} R_{\pm}^{i}(\Delta\text{ASVI}_i)$ is the simple average of the top 30 positive (negative) search terms. Given the dispersion of investors' beliefs (e.g., Diether et al. (2002)), our index measures the net effect of sentiment on the market.

2. Search Terms Unrelated to Economics and Finance

a) We rely on the University of South Florida Free Association Norms (USF-FAN) to obtain the associated words of each of the six categories. USF-FAN is "the largest database of free association ever collected in the U.S. available to interested researchers and scholars" and is widely used in academic research (e.g., Palla, Derényi, Farkas, and Vicsek (2005), Steyvers and Tenenbaum (2005)); based on the data collected from more than 6,000 participants, the database provides a set of 5,019 stimulus words with a total of 72,000 associated word pairs.[15] We use 6 stimulus words, "weather," "disaster," "pollution," "terror/terrorist," "holiday," and "sport," to obtain 206 associated words. For example, we have "rain," "snow," and "sunny" for "weather" and "Christmas," "Thanksgiving," and "Halloween" for "holiday."

[13] During every June and December, we regress ΔASVI on the contemporaneous market returns by using historical data from Jan. 2004. We keep the top 30 search terms with the largest positive (negative) *t*-statistics as our positive (negative) sentiment portfolio to construct our sentiment measure for the following 6 months.

[14] Some countries have fewer than 30 positive (negative) search terms.

[15] A detailed description of the database can be found at http://w3.usf.edu/FreeAssociation/Intro.html.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

b) Following Hu and Liu (2004) and Wiebe, Wilson, and Cardie (2005), we use the Opinion Lexicon and the Multi-Perspective Question Answering (MPQA) Subjectivity Lexicon to identify the sentiment of these 206 words.[16] We keep only words identified with positive or negative sentiment, and we remove neutral words. A word is defined as neutral if its sentiment tags from the two lexicons are both neutral or are exactly opposite one another (e.g., a positive sentiment tag in the Opinion Lexicon but a negative sentiment tag in MPQA). For words in the "sport" category, we add "win" and "lose" to each word when identifying positive or negative sentiment. For example, instead of using "soccer," we use "soccer win" and "soccer lose." We end up with 91 words.

c) We then translate these 91 English words into each country's corresponding language by using Google Translate.

d) We input these 91 translated words into Google Trends and obtain the top 10 terms associated with each word. We keep only search terms that have at least 100 weekly SVI observations and remove terms that are clearly not related to one of the six categories. For example, we obtain both "cloud computing" and "cloudy day" associated with "cloud." Then, we delete "cloud computing" and keep "cloudy day" for the weather category.

e) After selecting these search terms, we follow the same procedure in steps v) to viii) for constructing economics- and finance-related terms to create our non-economics- and non-finance-related sentiment indices, $\text{SENTIMENT}_{\text{NON_ECON},t}$, for each country in our sample.

Interestingly, the average correlation between economics- and non-economics-related indices across our sample countries is only 0.065.[17] The low correlation indicates that the two sentiment measures capture market sentiment from different dimensions. We thus incorporate both dimensions by constructing a composite sentiment index. Specifically, we use the simple average of economics- and non-economics-related indices as our key sentiment measure, $\text{SENTIMENT}_t = (\text{SENTIMENT}_{\text{ECON},t} + \text{SENTIMENT}_{\text{NON_ECON},t})/2$.[18] Although we use this composite sentiment index to conduct most of the following

[16] The two lexicons assign three types of sentiment tags to words in their database: positive, negative, and neutral. Their database covers a large number of words (Bing Liu's Lexicon covers nearly 7,000 words, and MPQA covers over 8,000 words/phrases), and they are widely used in the sentiment analysis literature by computer science scholars. Please refer to https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html for a detailed description of Bing Liu's Lexicon. Sentiment tags provided by the two lexicons can be obtained from http://sentiment.christopherpotts.net/lexicons.html#opinionlexicon.

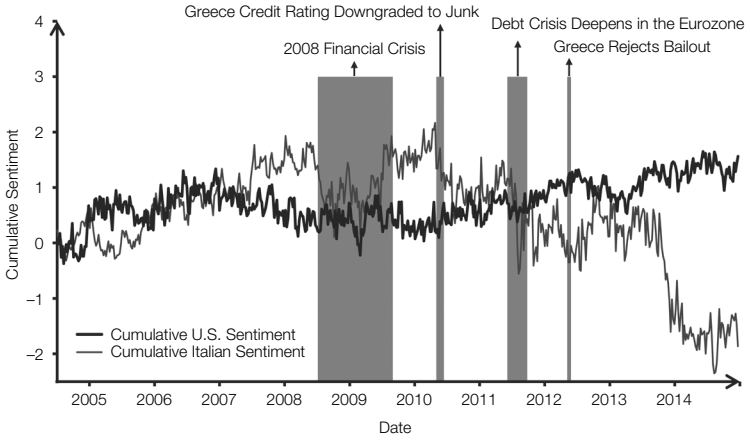
[17] We also compute and report the correlations for each country in Table IA1 in the Supplementary Material (we define the tables in the Supplementary Material with the prefix "IA").

[18] It is transparent and straightforward to construct our composite index based on the simple average of the two indices. We acknowledge that the weights on economics and noneconomics dimensions may vary across countries or economic environments. However, it is difficult to determine weights ex ante. Several ex post tests further suggest that it is reasonable to use the simple average: i) As shown later in Table 4, the contributions of the two indices to the return prediction are independent and similar in magnitude. ii) We also use an alternative way to construct the composite index: For each country, we first run time-series regressions of non-economics-related (economics-related) sentiment

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

FIGURE 1
Cumulative U.S. and Italian Sentiment over Time

Figure 1 shows the cumulative U.S. and Italian sentiment from July 2004 to Dec. 2014. The 2008 financial crisis period and periods with influential events during the European debt crisis are highlighted.



empirical analyses, the two sentiment indices independently yield a similar and consistent result, as shown later in Table 4. We calculate the weekly sentiment proxy for 38 countries from July 2004 to Dec. 2014, in which the first 6-month period from Jan. 2004 to June 2004 is used as the initial rolling regression window.[19]

Figure 1 plots the cumulative U.S. and Italian sentiment from July 2004 to Dec. 2014 as an example. Because our sentiment indices measure the changes in investor sentiment rather than the levels of investor sentiment, we present the cumulative value of the sentiment indices for illustration.[20] Apparently, the two cumulative sentiment indices vary substantially over time. As highlighted, the financial crisis and influential events during the European debt crisis caused investor sentiment to plunge immensely. Intuitively, the Italian sentiment seems to be more vulnerable to European shocks than is the U.S. sentiment. Nevertheless, this figure is only illustrative by characterizing the fluctuation in our sentiment measure. As suggested by the time-series pattern described earlier, the shift in

on market returns during the full sample period (July 2004–Dec. 2014). We then use the *t*-statistics as the weights for the non-economics- and economics-related sentiment measures and construct a *t*-statistic-weighted sentiment index for each country. We apply this index to repeat the analyses in Table 3 (panel regression) and in Table 5 (by-country regression). The results are presented in Table IA7 in the Supplementary Material. We find results similar to those found using the simple average index (as shown in Tables 3 and 5) in both magnitude and statistical significance. Note that we use the weights extracted from the full sample, which are inappropriate for constructing an index for return prediction. However, the similarity between the two indices suggests that the simple average index is a reasonable one for incorporating useful information from both economics and noneconomics environments.

[19]Table IA18 in the Supplementary Material tabulates search terms used to construct the sentiment indices over our full sample period for each country.

[20]The cumulative sentiment indices are calculated as the sum of the historical sentiment indices since the first week of July 2004.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

investor sentiment (especially economics-related sentiment) correlates with fundamental movements, because mispricing can stem from investors' underreaction or overreaction to fundamental news. More rigorous empirical investigations such as validation, return prediction, and sentiment channels are therefore warranted.

Table IA2 in the Supplementary Material presents summary statistics for our sentiment indices. The mean and median of the sentiment indices are close to 0 across countries. The explanation for this observation is that our sentiment indices measure the changes in investor sentiment rather than the level of investor sentiment. The quartiles of the indices show that our sentiment measures are largely symmetrically distributed around zero. Moreover, our sentiment indices are similarly distributed, with average standard deviations of 0.286 for developed countries and 0.298 for emerging countries.

C. Other Data

Our market return data come from Datastream. We download the country-level daily and weekly total return index (RI) in U.S. dollars. We use daily returns to calculate the market volatility (VOLATILITY) for each week and use weekly returns to test the return prediction of sentiment. We also download the market index in the country's local currency to test the robustness of our results.

For our validation test on the relation between sports events and our sentiment indices, we obtain the results of knockout stage matches from the 4 most influential soccer tournaments from 2004 to 2014 (i.e., World Cup in 2006, 2010, and 2014; European Championship in 2008 and 2012; Copa América in 2007 and 2011; and Asian Cup in 2007 and 2011). Specifically, we collect information on the winner, loser, and event date for each match. Table IA3 in the Supplementary Material presents summary statistics for the match results. We report the total number of wins and losses for each country in each tournament. In total, we have 72 matches played by 18 of our 38 sample countries. Among these 18 countries, Italy, Spain, and Germany, for instance, won the World Cup championship in 2006, 2010, and 2014, respectively.

We obtain identifiers and firm names of American Depositary Receipts (ADRs) from the 4 primary depository institutions: Citibank, The Bank of New York, J.P. Morgan, and Deutsche Bank. We then collect daily price information for these ADRs and their parent stocks from Datastream. Table IA9 in the Supplementary Material reports descriptive statistics for ADRs and their parent stocks used in the regression analysis. Our sample includes 1,731 pairs of ADRs and their parent stocks from 38 countries.

Moreover, we include a set of U.S. and local stock market and macroeconomic variables as controls. The market volatility index (VIX_{US}) is the implied volatility of the U.S. Standard & Poor's (S&P) 500 index options, and this measure reflects investors' expectations about the volatility of the U.S. stock market over the subsequent 30 days. We obtain VIX_{US} from the Chicago Board Options Exchange (CBOE). We also add market volatility (VOLATILITY), which is calculated as the standard deviation of daily returns in a week, to control for the local stock market risk. We further include a variable to measure U.S. macroeconomic activities ($ECONOMY_{US}$), which is constructed by Aruoba, Diebold, and Scotti (2009). $ECONOMY_{US}$ consists of several seasonally adjusted

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

macroeconomic factors, such as weekly initial jobless claims, monthly payroll employment, industrial production, personal income less transfer payments, manufacturing and trade sales, and quarterly real gross domestic product (GDP).[21] In addition, we control for the index of U.S. economic policy uncertainty (EPU_{US}), which is constructed by Baker, Bloom, and Davis (2015). EPU_{US} counts the number of U.S. newspaper articles with at least 1 term from the following three categories: i) "economic" or "economy," ii) "uncertain" or "uncertainty," and iii) "congress," "deficit," "Federal Reserve," "legislation," "regulation," or "White House." Furthermore, we include lagged stock market returns as controls, especially in the return prediction analysis.

Finally, to further address the concern about the correlation between the sentiment index and business-cycle and risk factors, we follow Sibley et al. (2016) and construct business-cycle and risk variables. Because of the lack of high-frequency macroeconomic information in non-U.S. countries, we can construct only U.S. variables as a proxy for global fundamental and risk factors. For variables available only on a monthly frequency, we use the monthly change as weekly proxies. Specifically, we obtain the U.S. monthly unemployment rate ($UNEMP_{US}$) and the consumer price index (CPI_{US}) from the U.S. Department of Labor Web site (https://www.bls.gov/cpi). The NBER recession dummy (NBER_RECESSION) is obtained from the NBER Web site (https://www.nber.org/cycles.html). Monthly data on consumption ($CONSUMPTION_{US}$) and disposable income ($INCOME_{US}$) are taken from the U.S. Department of Commerce Web site (https://www.commerce.gov/data-and-reports/economic-indicators). We obtain the monthly level of industrial production ($INDPROD_{US}$), weekly 3-month Treasury Bill rate ($TBILL_{US}$), weekly term spread ($TERM_{US}$), and weekly default spread ($DEFAULT_{US}$) from the Board of Governors of the Federal Reserve System Web site (https://fred.stlouisfed.org/categories). $TERM_{US}$ is the difference in yields between 10-year Treasury bonds and 3-month Treasury bills, and $DEFAULT_{US}$ is the difference in yields between BAA corporate bonds and AAA corporate bonds. We follow Lesmond, Ogden, and Trzcinka (1999) in constructing the weekly liquidity risk measure ($ILLIQ_{US}$), which is the percentage of stocks with zero returns in a week.

Following Ferreira and Matos (2008), we construct a measure of total institutional ownership (IO_{TOTAL}), which is the sum of a country's total institution holdings scaled by its total stock market capitalization at the end of each calendar year. The institutional holding data are from FactSet.[22] We thus use 1 minus IO_{TOTAL} as our proxy for retail investor trading activity (RETAIL_TOTAL). We report summary statistics for the retail holdings in Table IA4 in the Supplementary Material.

[21] Aruoba et al. (2009) have updated the data at https://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index.

[22] Note that FactSet does not provide institutional holding data for Argentina or Pakistan, so our sample consists of 36 countries in the following tests of the effect of retail investor trading activity.

III. Results

A. Validation of the Search-Based Sentiment Index

Given the intrinsic nature of sentiment and the complexity of measurement, it is important to validate our sentiment indices externally before we conduct any investigation into the return prediction of sentiment indices. In this section, we employ international sports events as the validation test for our sentiment measures.

Edmans et al. (2007) use international soccer results as an indicator of investor sentiment and argue that international soccer results drive investor sentiment in a substantial and unambiguous way and that this impact is correlated across the majority of individuals within a country. To test this conjecture, they examine the impact of international soccer results on stock prices and find that losses in soccer matches have a significant negative effect on the losing country's stock market. For example, a loss in the World Cup knockout stage leads to a next-day abnormal stock return of −49 bps. Their finding is consistent with previous psychological evidence that sports results have a significant effect on sentiment (e.g., Hirt, Zillmann, Erickson, and Kennedy (1992)).

Because sports sentiment is independent of a country's economic performance and fundamentals, it is straightforward and easy to examine the impact of influential soccer matches on our sentiment measure. Specifically, we connect our indices to the knockout stage match outcomes of the World Cup in 2006, 2010, and 2014; European Championship in 2008 and 2012; Copa América in 2007 and 2011; and Asian Cup in 2007 and 2011.[23] Our test includes 18 countries that played knockout stage matches in our sample of soccer tournaments from 2006 to 2014.

We then run the following panel regression:

$$(2)\quad \text{SENTIMENT}_{\text{NO_SPORT},i,t} = a + b\text{LOSS}_{i,t} + c\text{WIN}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},$$

where $\text{SENTIMENT}_{\text{NO_SPORT},i,t}$ is country i's sentiment in week t, constructed without sports-related search items. $\text{LOSS}_{i,t}$ equals 1 if country i loses a match in week t, and 0 otherwise. $\text{WIN}_{i,t}$ equals 1 if country i wins a match in week t, and 0 otherwise. $\text{CONTROLS}_{i,t}$ constitute a set of control variables that include stock market volatility ($\text{VOLATILITY}_{i,t}$) and lagged market returns in the prior 4 weeks ($\text{RETURN}_{i,t-1}$, $\text{RETURN}_{i,t-2}$, $\text{RETURN}_{i,t-3}$, and $\text{RETURN}_{i,t-4}$). We include country fixed effects and year-week fixed effects in the regression and cluster standard errors at both the country and year-week levels.

Panel A of Table 2 presents the results. Interestingly, our sentiment index is negatively correlated with the loss dummy at the 5% significance level. In terms of economic significance, a loss in these soccer matches reduces the sentiment index by 0.095 (or 0.098 without controls) in absolute magnitude and by 30.9% (or 31.9% without controls) relative to the standard deviation of the sentiment index. Although the sentiment index is positively associated with the win dummy, the

[23] To avoid a potential mechanical correlation between sports events and our sentiment indices, we remove the sports category and reconstruct the non-economics-related sentiment measure and the composite sentiment indices for the validation test in this section.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

results are not statistically significant. This asymmetric finding between wins and losses is highly consistent with the findings of Edmans et al. (2007), who provide two reasons for this result: First, whereas an increase in heart attacks, crimes, and suicides is shown after soccer losses, there is no evidence of improvements in mood of a similar magnitude after wins. Second, a win at the knockout stage advances a country only to the next stage, but a loss immediately removes the country from the competition.[24]

TABLE 2
Sentiment and Soccer Match Results

Table 2 reports how different sentiment indices separately respond to important soccer matches. We collect the results of the knockout stage games played by our sample countries in the World Cup 2006, 2010, and 2014; European Championship 2008 and 2012; Copa América 2007 and 2011; and Asian Cup 2007 and 2011. We run the following regression using different sentiment indices as the dependent variables:

$$\text{SENTIMENT}_{\text{NO_SPORT},i,t}/\text{SENTIMENT}_{\text{NON_ECON_NO_SPORT},i,t}/\text{SENTIMENT}_{\text{ECON},i,t} = a + b\text{LOSS}_{i,t} + c\text{WIN}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},$$

where $\text{SENTIMENT}_{\text{NO_SPORT},i,t}$ (Panel A) is country i's composite sentiment (excluding the sports category), $\text{SENTIMENT}_{\text{NON_ECON_NO_SPORT},i,t}$ (Panel B) is the non-economics-related sentiment (excluding the sports category), and $\text{SENTIMENT}_{\text{ECON},i,t}$ (Panel C) is the economics-related sentiment. $\text{LOSS}_{i,t}$ equals 1 if country i loses a match in week t, and 0 otherwise. $\text{WIN}_{i,t}$ equals 1 if country i wins a match in week t, and 0 otherwise. $\text{CONTROLS}_{i,t}$ include $\text{VOLATILITY}_{i,t}$, which is country i's stock market volatility based on total market index in week t, and $\text{RETURN}_{i,t-1}$ to $\text{RETURN}_{i,t-4}$, which are weekly returns lagged 1 to 4 weeks for country i. Models 1 and 2 use all data from 2006 to 2014; models 3 and 4 use the event-window periods of 6 months before and 6 months after each soccer tournament period (usually June to July); models 5 and 6 use the event-window periods of 3 months before and 3 months after each soccer tournament period. We include country (C) and year-week (T) fixed effects in the regressions. Standard errors are clustered at the country and year-week levels. t-statistics are reported in parentheses.

Panel A. Composite Sentiment Excluding Sports around Soccer Matches

	Dependent Variable: $\text{SENTIMENT}_{\text{NO_SPORT},i,t}$					
	Full Sample		[–6 Month, 6 Month]		[–3 Month, 3 Month]	
	Model					
Variable	1	2	3	4	5	6
LOSS_t	−0.098	−0.095	−0.097	−0.095	−0.097	−0.096
	(−2.03)	(−2.01)	(−2.01)	(−2.00)	(−1.97)	(−1.96)
WIN_t	−0.014	−0.010	−0.014	−0.009	−0.014	−0.011
	(−0.38)	(−0.26)	(−0.36)	(−0.24)	(−0.36)	(−0.29)
VOLATILITY_t		0.011		0.007		0.063
		(0.19)		(0.11)		(0.86)
RETURN_{t-1}		−0.009		−0.008		−0.004
		(−2.44)		(−1.88)		(−0.72)
RETURN_{t-2}		0.005		0.003		0.003
		(1.70)		(0.86)		(0.66)
RETURN_{t-3}		−0.006		−0.007		−0.006
		(−1.92)		(−2.19)		(−1.33)
RETURN_{t-4}		0.000		−0.003		−0.002
		(0.09)		(−0.70)		(−0.39)
Fixed effects	CT	CT	CT	CT	CT	CT
Clustering	CT	CT	CT	CT	CT	CT
No. of obs.	8,261	8,261	6,596	6,596	4,405	4,405
R^2	11.6%	11.9%	12.0%	12.2%	11.8%	11.9%

(continued on next page)

[24] For robustness, we also limit our sample to a 14-month event window that includes the 6 months before and 6 months after a 2-month soccer tournament period in models 3–4 and to an 8-month event window that includes the 3 months before and 3 months after a 2-month soccer tournament period in models 5–6. All of these specifications yield similar evidence.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 2 (continued)
Sentiment and Soccer Match Results

Panel B. Non-Economics-Related Sentiment Excluding Sports around Soccer Matches

	Dependent Variable: $SENTIMENT_{NON_ECON_NO_SPORT,i,t}$					
	Full Sample		[–6 Month, 6 Month]		[–3 Month, 3 Month]	
	Model					
Variable	1	2	3	4	5	6
$LOSS_t$	−0.112	−0.109	−0.112	−0.108	−0.112	−0.109
	(−2.62)	(−2.54)	(−2.61)	(−2.54)	(−2.59)	(−2.50)
WIN_t	−0.034	−0.028	−0.033	−0.027	−0.036	−0.031
	(−0.82)	(−0.68)	(−0.81)	(−0.64)	(−0.87)	(−0.74)
$VOLATILITY_t$		−0.027		−0.025		−0.013
		(−0.74)		(−0.56)		(−0.22)
$RETURN_{t-1}$		−0.012		−0.013		−0.009
		(−3.19)		(−3.00)		(−1.57)
$RETURN_{t-2}$		0.002		0.002		0.005
		(0.60)		(0.43)		(0.72)
$RETURN_{t-3}$		−0.006		−0.008		−0.004
		(−2.02)		(−2.51)		(−0.71)
$RETURN_{t-4}$		0.001		−0.002		−0.009
		(0.38)		(−0.47)		(−1.36)
Fixed effects	CT	CT	CT	CT	CT	CT
Clustering	CT	CT	CT	CT	CT	CT
No. of obs.	8,261	8,261	6,596	6,596	4,405	4,405
R^2	9.5%	9.6%	9.5%	9.7%	9.4%	9.5%

Panel C. Economics-Related Sentiment around Soccer Matches

	Dependent Variable: $SENTIMENT_{ECON,i,t}$					
	Full Sample		[–6 Month, 6 Month]		[–3 Month, 3 Month]	
	Model					
Variable	1	2	3	4	5	6
$LOSS_t$	−0.066	−0.064	−0.062	−0.060	−0.060	−0.061
	(−1.19)	(−1.18)	(−1.07)	(−1.07)	(−1.02)	(−1.04)
WIN_t	−0.000	0.003	0.003	0.006	0.006	0.006
	(−0.01)	(0.05)	(0.05)	(0.10)	(0.09)	(0.11)
$VOLATILITY_t$		0.030		0.008		0.110
		(0.35)		(0.08)		(0.92)
$RETURN_{t-1}$		−0.007		−0.004		0.000
		(−0.97)		(−0.61)		(0.01)
$RETURN_{t-2}$		0.006		0.002		−0.001
		(1.59)		(0.52)		(−0.16)
$RETURN_{t-3}$		−0.005		−0.005		−0.007
		(−0.99)		(−0.95)		(−0.97)
$RETURN_{t-4}$		−0.002		−0.005		0.004
		(−0.36)		(−0.82)		(0.63)
Fixed effects	CT	CT	CT	CT	CT	CT
Clustering	CT	CT	CT	CT	CT	CT
No. of obs.	8,261	8,261	6,596	6,596	4,405	4,405
R^2	9.9%	10.0%	10.1%	10.1%	10.1%	10.2%

We also examine how non-economics-related and economics-related sentiment indices respond to soccer events separately in Panels B and C of Table 2. Interestingly, although the coefficient of the non-economics-related sentiment index is negative and significant, the effect on the economics-related sentiment measure is statistically insignificant. This finding suggests that a shock by noneconomics factors may not have much of an effect on households' search behaviors for economics-related terms, which is also consistent with the low correlation

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

between the two measures. This international test thus validates the non-economics-based portion of our measure as a proxy for sentiment and provides support for its ability to capture sentiment.

B. Return Prediction

In this subsection, we study the relation between our sentiment indices and market returns across countries. According to sentiment theory, sentiment drives market prices away from economic fundamentals in the presence of limits to arbitrage. If high or low sentiment is not persistent, market prices will revert to the fundamentals after sentiment retreats. We form our basic empirical predictions based on this theory: Market returns should present a reversal pattern following sentiment; that is, high (low) sentiment should predict low (high) returns in the following week.

We construct a panel for the 38-country sample from July 2004 to Dec. 2014. In this panel, we pool our weekly sentiment indices and weekly market returns and run the following regressions:

$$\text{RETURN}_{i,t+1} = a + b\text{SENTIMENT}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t}, \quad (3)$$

where $\text{CONTROLS}_{i,t}$ constitute an array of control variables that include the weekly average of economic policy uncertainty in the United States ($\text{EPU}_{\text{US},t}$), weekly average of the CBOE daily market volatility index ($\text{VIX}_{\text{US},t}$), weekly average of daily macroeconomic activities ($\text{ECONOMY}_{\text{US},t}$), stock market volatility ($\text{VOLATILITY}_{i,t}$), and lagged market returns in the prior 4 weeks (RETURN_{t-1}, RETURN_{t-2}, RETURN_{t-3}, and RETURN_{t-4}). In addition, all of these regressions include country fixed effects, and standard errors are clustered at the country and year-week levels.[25]

Models 1 and 2 of Table 3 show the return prediction of sentiment for all countries in our sample. As shown, a 1-standard-deviation increase in market sentiment predicts a decline in weekly returns of 26 bps (annual returns of 13%). We further use the return prediction of our sentiment measure to create a profitable trading strategy. Specifically, for every week, we group countries into 2 equally weighted portfolios: a high-sentiment portfolio, which contains countries with positive sentiment indices, and a low-sentiment portfolio, which contains countries with negative sentiment indices. We then long the low-sentiment portfolio and short the high-sentiment portfolio for 1 week. This trading strategy generates, on average, a weekly return of 22 bps (annual return of 11% and Sharpe ratio of 2.0) with a *t*-statistic of 6.61. Applying this trading strategy solely to the U.S. stock market index, we obtain a weekly return of 34 bps (annual return of 18% and Sharpe ratio of 1.0) with a *t*-statistic of 3.22. Apparently, the trading strategy based on our full sample of countries can diversify risk, thus yielding a higher Sharpe ratio.

In addition, we run panel regressions separately for subsamples of emerging and developed countries. Models 3 and 4 report the results for developed markets and models 5 and 6 for emerging markets. Comparing the coefficients, we find

[25]Note that we do not control for year-week fixed effects in subsequent tests because we find that global sentiment (as defined in Section III) is crucial to sentiment across countries and prediction tests. The time fixed effects would absorb this important factor.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 3
Sentiment and Stock Returns

Table 3 presents the relation between weekly stock returns and our sentiment measure. We run the following regression:

$$\text{RETURN}_{i,t+1} = a + b\text{SENTIMENT}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t}.$$

In models 1 and 2, we use all 38 countries to run the regression; in models 3 and 4, we use only developed countries; and in models 5 and 6, we use only developing countries. $\text{CONTROLS}_{i,t}$ constitute the following control variables: $\text{EPU}_{\text{US},t}$ is the average of daily economic policy uncertainty measure from Baker et al. (2015) in week t. $\text{VIX}_{\text{US},t}$ is the average of the Chicago Board Options Exchange daily market volatility index in week t. $\text{ECONOMY}_{\text{US},t}$ is the average of a daily macroeconomic activities measure obtained from Aruoba et al. (2009) in week t. $\text{VOLATILITY}_{i,t}$ is country i's stock market volatility based on total market index in week t. $\text{RETURN}_{i,t-1}$ to $\text{RETURN}_{i,t-4}$ represent the lagged 1 to lagged 4 weekly returns for country i, respectively. EMG, DEV, and WRD denote emerging, developed, and global markets, respectively. We use country (C) and year-week (T) fixed effects in all regressions. Standard errors are clustered at the country and year-week levels. t-statistics are reported in parentheses.

	Dependent Variable: RETURN_{t+1}					
	WRD		DEV		EMG	
	Model					
Variable	1	2	3	4	5	6
SENTIMENT_t	−0.458	−0.450	−0.540	−0.511	−0.394	−0.395
	(−6.64)	(−7.66)	(−4.94)	(−5.71)	(−5.75)	(−5.83)
$\text{EPU}_{\text{US},t}$		−3.080		−3.401		−2.775
		(−1.69)		(−1.67)		(−1.67)
$\text{VIX}_{\text{US},t}$		3.195		3.909		2.697
		(1.97)		(1.97)		(1.82)
$\text{ECONOMY}_{\text{US},t}$		0.320		0.355		0.292
		(2.46)		(2.37)		(2.45)
VOLATILITY_t		−0.709		−0.977		−0.601
		(−1.39)		(−1.30)		(−1.28)
RETURN_t		0.007		−0.011		0.019
		(0.16)		(−0.20)		(0.51)
RETURN_{t-1}		0.048		0.038		0.053
		(1.07)		(0.68)		(1.35)
RETURN_{t-2}		−0.005		−0.004		−0.007
		(−0.11)		(−0.06)		(−0.17)
RETURN_{t-3}		0.010		−0.002		0.017
		(0.24)		(−0.03)		(0.46)
RETURN_{t-4}		0.025		0.024		0.025
		(0.62)		(0.43)		(0.75)
Fixed effects	C	C	C	C	C	C
Clustering	CT	CT	CT	CT	CT	CT
No. of obs.	20,748	19,836	9,282	8,874	11,466	10,962
R^2	0.6%	2.7%	0.9%	3.7%	0.4%	2.3%

that sentiment indices provide a stronger prediction for future returns in developed markets than in emerging markets: As shown in model 6, the coefficient of the sentiment indices is −0.395 for emerging markets and −0.511 for developed markets. The economic difference is also notable: A 1-standard-deviation increase in sentiment leads to a decrease in the market return in the following week of 23 bps in emerging markets. The magnitude is smaller than the 28 bps decrease in developed markets.

Table 4 provides additional regression results for our sentiment indices based on either economics- and finance-related search terms or search terms unrelated to economics and finance. As shown, both indices demonstrate significant predictive power for future stock market returns. A 1-standard-deviation increase in $\text{SENTIMENT}_{\text{ECON}}$ ($\text{SENTIMENT}_{\text{NON_ECON}}$) leads to a decrease in the market return in the following week of 16 (19) bps, which suggests that the non-economics-related sentiment index provides stronger predictive power.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 4

Economics- and Non-Economics-Related Sentiment and Stock Returns

Table 4 presents the relation between weekly stock returns and our economics- and non-economics-related sentiment measures. We run the following regression:

$$\text{RETURN}_{i,t+1} = a + b\text{SENTIMENT}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t}.$$

In models 1 and 2, we use our economics-related sentiment measure as the key independent variable; in models 3 and 4, we use our non-economics-related sentiment measure, and in models 5 and 6, we include both of the sentiment measures in our regression. $\text{CONTROLS}_{i,t}$ constitute the same control variables as defined in Table 3. We use country and year-week fixed effects in all regressions. Standard errors are clustered at the country (C) and year-week (T) levels. *t*-statistics are reported in parentheses.

	Dependent Variable: RETURN_{t+1}					
	Model					
Variable	1	2	3	4	5	6
$\text{SENTIMENT}_{\text{ECON},t}$	−0.206 (−5.28)	−0.201 (−5.76)			−0.194 (−5.14)	−0.195 (−5.66)
$\text{SENTIMENT}_{\text{NON_ECON},t}$			−0.287 (−5.71)	−0.269 (−5.89)	−0.275 (−5.57)	−0.263 (−5.74)
$\text{EPU}_{\text{US},t}$		−3.085 (−1.69)		−3.086 (−1.69)		−3.080 (−1.69)
$\text{VIX}_{\text{US},t}$		3.194 (1.97)		3.187 (1.96)		3.194 (1.97)
$\text{ECONOMY}_{\text{US},t}$		0.321 (2.46)		0.321 (2.46)		0.320 (2.46)
VOLATILITY_t		−0.714 (−1.40)		−0.699 (−1.36)		0.007 (0.17)
RETURN_t		−0.002 (−0.04)		0.000 (0.00)		−0.707 (−1.39)
RETURN_{t-1}		0.051 (1.13)		0.049 (1.11)		0.048 (1.07)
RETURN_{t-2}		−0.005 (−0.11)		−0.005 (−0.10)		−0.005 (−0.11)
RETURN_{t-3}		0.010 (0.25)		0.010 (0.25)		0.010 (0.24)
RETURN_{t-4}		0.025 (0.62)		0.026 (0.63)		0.025 (0.62)
Fixed effects	C	C	C	C	C	C
Clustering	CT	CT	CT	CT	CT	CT
No. of obs.	20,748	19,836	20,748	19,836	20,748	19,836
R^2	0.3%	2.5%	0.4%	2.5%	0.6%	2.7%

Given the low correlation between the two sentiment measures, their predictability remains almost unchanged when we include both sentiment indices in the regressions. This finding further suggests that both indices capture market sentiment from different dimensions and can independently predict future returns. Our method of adding two sentiment indices into one measure enables us to incorporate both into our prediction exercise.

In Table 5, we perform the same analysis by country. Panel A reports the results for each country, and Panel B summarizes the results for developed countries, emerging countries, and all sample countries. The results indicate that all countries except Israel and Mexico exhibit a negative relation between SENTIMENT_t and RETURN_{t+1}. Twenty of the 38 countries display a pattern that is significant at the 5% level. Panel A ranks the sample countries based on their economic significance. Among these countries, Spain has the strongest reversal pattern, where a 1-standard-deviation increase in sentiment predicts a decrease in weekly market returns of 73 bps.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 5
Sentiment and Stock Returns by Country

Table 5 reports the results of the following regression for each of our 38 countries from July 2004 to Dec. 2014:

$$\text{RETURN}_{t+1} = a + b\text{SENTIMENT}_t + \text{CONTROLS} + \varepsilon_t.$$

RETURN_t is the weekly market returns. CONTROLS are the control variables as defined in Table 3. Panel A presents the regression results of each country and ranks the countires based on their economic significance, which is the changes in weekly market returns with a 1-standard-deviation increase in sentiment measure. Panel B presents the summary statistics of the regression results in Panel A. We report the average of coefficients on SENTIMENT_t, the average of corresponding t-statistics, and the average of R^2s over the 38 countries in Panel B. NEGATIVE_SIG is the number of coefficients on SENTIMENT_t that are negative and significant at the 5% level. EMG, DEV, and WRD denote emerging, developed, and global markets, respectively.

Panel A. Effects of Sentiment by Country

Country	EMG/DEV	b	t-Statistic	R^2	Economic Sig.
Spain	DEV	−1.409	−4.05	5.3%	−0.73%
Italy	DEV	−0.953	−2.73	4.4%	−0.50%
Canada	DEV	−1.110	−2.84	5.1%	−0.45%
Australia	DEV	−1.151	−2.82	5.4%	−0.48%
Singapore	DEV	−0.609	−2.67	7.4%	−0.33%
Japan	DEV	−0.612	−2.48	3.3%	−0.28%
Indonesia	EMG	−0.927	−2.33	4.4%	−0.44%
China	EMG	−0.712	−2.41	2.6%	−0.39%
Austria	DEV	−0.527	−2.37	7.1%	−0.39%
Netherlands	DEV	−0.707	−2.28	5.9%	−0.36%
New Zealand	DEV	−0.465	−2.19	3.1%	−0.28%
Russia	EMG	−1.063	−2.16	6.1%	−0.50%
United States	DEV	−0.676	−2.21	8.8%	−0.25%
Chile	EMG	−0.571	−2.13	7.3%	−0.30%
Philippines	EMG	−0.557	−2.04	2.1%	−0.29%
Switzerland	DEV	−0.279	−2.07	5.6%	−0.25%
South Africa	EMG	−0.677	−2.13	8.0%	−0.38%
Pakistan	EMG	−0.495	−2.04	6.6%	−0.32%
United Kingdom	DEV	−0.742	−1.96	5.2%	−0.28%
Denmark	DEV	−0.455	−1.97	5.9%	−0.30%
Thailand	EMG	−0.575	−1.93	4.7%	−0.30%
Peru	EMG	−0.355	−1.76	5.2%	−0.21%
Taiwan	EMG	−0.337	−1.69	3.8%	−0.22%
Poland	EMG	−0.607	−1.67	3.7%	−0.32%
Germany	DEV	−0.347	−1.63	5.0%	−0.24%
Egypt	EMG	−0.397	−1.61	5.6%	−0.27%
Malaysia	EMG	−0.222	−1.54	3.4%	−0.15%
India	EMG	−0.663	−1.51	5.6%	−0.28%
France	DEV	−0.489	−1.32	3.4%	−0.20%
South Korea	EMG	−0.402	−1.33	4.1%	−0.24%
Hungary	EMG	−0.345	−0.84	4.7%	−0.19%
Argentina	EMG	−0.243	−0.79	2.2%	−0.14%
Hong Kong	DEV	−0.148	−0.63	3.0%	−0.09%
Portugal	EMG	−0.135	−0.60	2.5%	−0.09%
Colombia	EMG	−0.143	−0.48	6.6%	−0.08%
Sweden	DEV	−0.033	−0.12	3.7%	−0.02%
Israel	EMG	0.005	0.03	2.7%	0.00%
Mexico	EMG	0.159	0.51	7.3%	0.08%

Panel B. Summary Statistics of Coefficients on Sentiment in Panel A

	b	t-Statistic	NEGATIVE_SIG	R^2
EMG	−0.441	−1.45	7 out of 21	4.7%
DEV	−0.630	−2.14	13 out of 17	5.1%
WRD	−0.526	−1.76	20 out of 38	4.9%

As SVI mainly tracks households' search behavior, our sentiment measure is expected to have stronger predictive power in countries with a larger fraction of retail investors. Table 6 confirms this hypothesis. Specifically, we first compute the time-series average of retail ownership for each country and then define countries whose average retail ownership ranks in the top one-third in our sample as high-retail-ownership countries and countries in the bottom one-third as low-retail-ownership countries. We then conduct sentiment prediction analyses

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 6
Effect of Retail Holdings

Table 6 presents the effect of retail holdings on the return prediction of investor sentiment. We conduct the analyses from Table 3 separately for high- and low-retail-ownership countries. In models 1 and 2, we define countries whose time-series averages of retail holdings ($RETAIL_TOTAL_{i,t}$) are in the top one-third of all sample countries as high-retail-ownership countries and those in the bottom one-third as low-retail-ownership countries. $RETAIL_TOTAL_{i,t}$ is 1 minus the sum of country i's total institution holdings scaled by its total stock market capitalization at the end of year t. In models 3 and 4, we choose 80% as the cutoff and classify countries whose the time-series averages of $RETAIL_TOTAL_{i,t}$ are above 80% as high-retail-ownership countries and those below 80% as low-retail-ownership countries. Variables are defined in Table 3. We use country (C) and year-week (T) fixed effects in all regressions. Standard errors are clustered at the country and year-week levels. t-statistics are reported in parentheses. p-values are reported in square brackets.

	RETAIL_TOTAL Top Third		RETAIL_TOTAL Bottom Third		RETAIL_TOTAL > 0.8		RETAIL_TOTAL < 0.8
	Model						
Variable	1		2		3		4
$SENTIMENT_t$	−0.570		−0.317		−0.525		−0.340
	(−6.92)		(−4.15)		(−9.09)		(−4.50)
Diff. in SENTIMENT		0.253				0.185	
[p-value]		[0.020]				[0.048]	
$EPU_{US,t}$	−2.881		−3.711		−3.001		−3.726
	(−5.84)		(−7.74)		(−8.61)		(−7.91)
$VIX_{US,t}$	2.806		4.647		2.991		4.487
	(7.03)		(10.44)		(10.38)		(10.28)
$ECONOMY_{US,t}$	0.277		0.379		0.297		0.381
	(7.36)		(10.14)		(11.09)		(10.39)
$VOLATILITY_t$	−0.591		−1.520		−0.510		−1.411
	(−2.95)		(−6.46)		(−3.48)		(−6.20)
$RETURN_t$	0.006		−0.032		0.020		−0.023
	(0.45)		(−2.42)		(2.16)		(−1.88)
$RETURN_{t-1}$	0.046		0.031		0.056		0.032
	(3.54)		(2.39)		(5.91)		(2.50)
$RETURN_{t-2}$	0.012		−0.023		0.006		−0.026
	(0.98)		(−1.81)		(0.67)		(−2.10)
$RETURN_{t-3}$	0.021		0.010		0.006		0.016
	(1.63)		(0.74)		(0.69)		(1.28)
$RETURN_{t-4}$	0.033		0.026		0.031		0.019
	(2.61)		(2.05)		(3.41)		(1.56)
Fixed effects	C		C		C		C
Clustering	CT		CT		CT		CT
No. of obs.	6,264		6,264		12,006		6,786
R^2	2.7%		3.8%		2.8%		3.5%

for these 2 groups. Our sentiment index has significantly stronger power to predict returns in high- versus low-retail-ownership countries.[26]

To ensure the robustness of the aforementioned prediction, we conduct several additional tests and summarize the main findings in Table IA5 of the Supplementary Material. First, we calculate market returns again in the country's own currency to avoid the concern that sentiment affects the movement of the local currency's value and that the sentiment effect on exchange rates contaminates market returns. Models 1 and 2 show that the tests remain robust. Second, we perform the analysis while excluding the global financial crisis period from Sept. 2008 to Sept. 2009, as one may be concerned that this unusual period drives our results. All our tests remain significant in the noncrisis period, as shown in

[26]For robustness, we choose 80% as a cutoff (the mean of retail ownerships across our sample countries) to classify our sample into high- versus low-retail-ownership countries. We find similar results.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

models 3 and 4. This result confirms that our sentiment indices still predict market returns in the following week after we exclude the crisis period.

C. Fundamental Factors

Sibley et al. (2016) cast doubt on the Baker and Wurgler (2006) sentiment index and argue that business-cycle and risk factors could simply explain its predictive power regarding cross-sectional stock returns. We confirm that the return prediction of sentiment is not driven by economic fundamentals. Specifically, we show that our sentiment indices have little correlation with these fundamental factors and that the residuals from regressing sentiment indices on fundamental factors have similarly strong predictability with respect to market returns, as shown in the previous subsection.

We first regress our weekly sentiment measure on the changes in all 8 business-cycle and risk variables from Sibley et al. (2016) and report the results in Panel A of Table IA8 in the Supplementary Material. Both the international sample and the U.S. sample indicate that business-cycle and risk factors have little explanatory power with respect to our sentiment indices, with R^2 values of only 0.3% and 1.8%, respectively. By contrast, Sibley et al. find that approximately 63% of the variations in the Baker and Wurgler (2006) index can be explained by these business-cycle and risk factors. We then perform our return prediction tests again with sentiment residuals. Panel B reports the results. We find that these residuals consistently lead to mispricing and portend future returns. However, the limitation of this test is the lack of controls of economic fundamentals in non-U.S. countries. With this caveat, it remains plausible that the return prediction of our sentiment measure is influenced by a country's fundamental factors.

To further address the previous concern, we follow Baker et al. (2012) and focus on dual-listed companies because the shares of dual-listed companies have the same cash flows but trade in different markets. We find a positive relation between the sentiment index and the price difference that firms' fundamentals fail to explain, which confirms that our sentiment index does indeed capture the mispricing. Specifically, by using 1,731 pairs of ADRs and their parent stocks, we test how the difference in sentiment between ADRs' parent countries and the United States affect the difference in price between ADRs' parent stocks and ADRs. As shown in Table IA10 in the Supplementary Material, the sentiment differences are positively associated with the stock price differences in other countries away from the benchmark in the United States.[27]

D. Testing Sentiment Channels

In this subsection, we conduct difference-in-differences analyses around plausibly exogenous shocks to two sentiment channels proposed by Baker and Wurgler (2006): i) difficult to value and ii) limits to arbitrage. We study the implementation of MiFID as a shock to valuation difficulty and use the short-selling ban as evidence of limits to arbitrage. These empirical designs help mitigate potential endogeneity issues and confirm that our measure captures investor sentiment across countries.

[27] Again, a limitation of this test is that country-specific fundamentals could drive the results because of the lack of controls of economic fundamentals in non-U.S. countries.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

1. Difficult-to-Value Channel and MiFID

MiFID is an influential European Union legislation for investment services and regulated markets. It became effective for all European Union members on Nov. 1, 2007. MiFID regulates investment intermediaries (e.g., investment firms) that provide services in financial instruments and trading venues (e.g., stock exchanges) where financial instruments are traded. The European Commission states that MiFID has two major objectives: i) "protecting investors and safeguarding market integrity by establishing harmonized requirements governing the activities of authorized intermediaries" and ii) "promoting fair, transparent, efficient and integrated financial markets."[28]

Specifically, MiFID includes several provisions to ensure investor protection and market transparency (e.g., conflicts of interest, provisions on best execution, client order handling, transaction reporting obligations, and pre- and posttrade transparency requirements for regulated markets). Essentially, these provisions protect investors and enhance transparency through mitigating insider trading, limiting market manipulation, and limiting broker–agent conflicts (Cumming, Johan, and Li (2011)). For example, pretrade transparency provisions require regulated markets to make the information on current bid and offer prices available to the public, and posttrade transparency requires regulated markets to make the detailed information (e.g., price, volume, and time) on the executed transactions available to the public. Because investors could rely on only 1 or 2 exchanges to collect information before the implementation of MiFID, these requirements have significantly increased the sources of information on prices and transactions and hence have improved the transparency of financial markets. We use the implementation of MiFID to test whether the change in a country's information environments affect the relation between future stock returns and current sentiment in that country.

To test the impact of information environments on the relation between market returns and sentiment, we run the following panel regressions with country fixed effects and standard errors clustered at the country and year-week levels from Dec. 2006 to Oct. 2008:[29]

$$\begin{aligned}(4)\quad \text{RETURN}_{i,t+1} = {} & \alpha + \beta_1 \text{SENTIMENT}_{i,t} + \beta_2 \text{SENTIMENT}_{i,t} \times \text{INF_ENV}_{i,t} \\ & + \beta_3 \text{INF_ENV}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},\end{aligned}$$

where $\text{INF_ENV}_{i,t}$ equals 1 if country i is exposed to MiFID at time t, and 0 otherwise. If the information environment of stock markets improves, the stock price responds less to the change in market sentiment. We thus expect β_2 to be positive and significant. Panel A of Table 7 confirms our prediction with evidence that β_2 is positive and significant in model 2. This finding indicates that the implementation of MiFID has substantially weakened the sentiment effect on market returns.

[28] See https://eur-lex.europa.eu/legal-content/EN/LSU/?uri=celex:32004L0039.

[29] The sample period in Cumming et al. (2011) is Feb. 2006 to Oct. 2008. We use a more balanced sample here. As shown in the following robustness test, our results do not change if we use their sample period.

TABLE 7

Effects of Information Environments and Short-Selling Bans

Table 7 reports the effects of Markets in Financial Instruments Directive (MiFID) and short-selling bans on the return predictability of investor sentiment. Panel A reports the results of the following panel regression:

$$\text{RETURN}_{i,t+1} = \alpha + \beta_1 \text{SENTIMENT}_{i,t} + \beta_2 \text{SENTIMENT}_{i,t} \times \text{INF_ENV}_{i,t} + \beta_3 \text{INF_ENV}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},$$

where $\text{RETURN}_{i,t+1}$ is country i's market returns in week $t+1$ and $\text{SENTIMENT}_{i,t}$ is country i's market sentiment in week t. In model 2, $\text{INF_ENV}_{i,t}$ equals 1 if country i is exposed to MiFID at time t, and 0 otherwise. The sample period is from Dec. 2006 to Oct. 2008.

Panel B reports the results of the following panel regression:

$$\text{RETURN}_{i,t+1} = \alpha + \beta_1 \text{SENTIMENT}_{i,t} + \beta_2 \text{SENTIMENT}_{i,t} \times \text{BAN}_{i,t} + \beta_3 \text{BAN}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},$$

where $\text{RETURN}_{i,t+1}$ is country i's market returns in week $t+1$ and $\text{SENTIMENT}_{i,t}$ is country i's market sentiment in week t. $\text{BAN}_{i,t}$ equals 1 if country i has imposed a short-selling ban at week t, and 0 otherwise. The sample period is Sept. 2007 (1 year before the first initiation date of bans) to Oct. 2009 (1 year after the last initiation date of bans). $\text{CONTROLS}_{i,t}$ are the control variables defined in Table 3. We include country fixed effects in all regressions. Standard errors are clustered at the country (C) and year-week (T) levels. t-statistics are reported in parentheses.

Panel A. Sentiment and Information Environments

	Dependent Variable: RETURN_{t+1}	
	Model	
Variable	1	MiFID_t 2
SENTIMENT_t	−0.335	−0.430
	(−3.65)	(−3.99)
$\text{SENTIMENT}_t \times \text{INF_ENV}_t$		0.431
		(2.22)
INF_ENV_t		−0.041
		(−0.21)
Controls	Yes	Yes
Fixed effects	C	C
Clustering	CT	CT
No. of obs.	3,610	3,610
R^2	8.8%	8.9%

Panel B. Sentiment and Short-Selling Bans

	Dependent Variable: RETURN_{t+1}	
	Model	
Variable	1	2
SENTIMENT_t	−0.758	−0.556
	(−4.19)	(−2.70)
$\text{SENTIMENT}_t \times \text{BAN}_t$		−0.642
		(−2.33)
BAN_t		0.209
		(0.88)
Controls	Yes	Yes
Fixed effects	C	C
Clustering	CT	CT
No. of obs.	4,104	4,104
R^2	8.5%	8.6%

2. Limits-to-Arbitrage Channel and Short-Selling Bans

Short-selling constraints are considered one of the most important limits to arbitrage (e.g., Jones and Lamont (2002), Nagel (2005), Gromb and Vayanos (2010), and Chu, Hirshleifer, and Ma (2017)). In practice, there are significant cross-country variations in the scope and types of short-selling constraints, which are called short-selling bans (Jain, Jain, McInish, and McKenzie (2013)). Even within a country, short-selling bans may vary over time. For example, during

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

the financial crisis from 2007 to 2009, many countries that had not previously forbidden short sales imposed short-selling bans on either financial stocks or all stocks. An interesting characteristic of these short-selling bans is that different countries imposed and then lifted these bans on different dates during the financial crisis.[30]

To study the limits-to-arbitrage channel, we collect information on short-selling bans from Beber and Pagano (2013) and Jain et al. (2013) for our 38 countries. As shown in Table IA11 of the Supplementary Material, 23 countries that imposed short-selling bans comprise the treatment group, and 15 countries that never imposed bans comprise the control group. The inception dates of short-selling bans range from Sept. 19, 2008 (Canada, United Kingdom, and United States) to Oct. 30, 2008 (Japan). The durations of these bans also vary across countries, with the longest being over 2 years (e.g., Japan) and the shortest being less than 1 month (e.g., United States). We choose 1 year before the inception date of the first ban and 1 year after the inception date of the last ban as the sample period for our main analysis. There are also countries that never imposed bans during our sample period and countries that have always banned short sales.

To test the impact of short-selling bans on the relation between stock market returns and investor sentiment, we run the following panel regressions with country fixed effects and standard errors clustered at the country and year-week levels:

$$\begin{aligned}(5)\quad \text{RETURN}_{i,t+1} = {} & \alpha + \beta_1 \text{SENTIMENT}_{i,t} + \beta_2 \text{SENTIMENT}_{i,t} \times \text{BAN}_{i,t} \\ & + \beta_3 \text{BAN}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},\end{aligned}$$

where $\text{BAN}_{i,t}$ equals 1 if country i has imposed a short-selling ban on either financial stocks or all stocks at week t, and 0 otherwise. Our main interest is the coefficient of the interaction term β_2. We conjecture that short-selling bans would make it more difficult for investors to correct the stock mispricing caused by market sentiment, which would lead to a larger impact of the current week's sentiment on the next week's market returns. Given the negative relation between future returns and current sentiment documented in the previous section, we expect β_2 to be negative and significant.

The results in Panel B of Table 7 are consistent with our prediction. A negative β_2 at the 5% significance level indicates that the return prediction of market sentiment is greater for countries with short-selling bans than for countries without short-selling bans on stocks. Furthermore, we conduct the same analysis from Panel B of Table 7 for high- and low-sentiment periods separately. A week is defined as a high- (low-) sentiment week if market sentiment in that week is larger (smaller) than zero. The results are shown in Table IA12 of the Supplementary Material, where columns 1 and 2 present the results for high- and

[30] Two types of bans were imposed on short sales during the financial crisis period: bans on naked short sales, where the seller does not first borrow the security, and bans on covered short sales, where the seller manages to borrow the security. Some countries imposed a ban only on naked short sales, whereas other countries may have used a combination of these two types of bans. According to Beber and Pagano (2013), covered short-selling bans were used more frequently than naked short-selling bans at the beginning of the financial crisis period, whereas only naked short-selling bans were in effect at the end of the financial crisis.

low-sentiment periods, respectively. Our focus is the coefficients on interaction terms. As presented in the table, sentiment has stronger predictive power during high-sentiment periods, which supports the notion that mispricing is more likely to stem from the short leg of the arbitrage when short-selling bans are implemented.

E. Global Sentiment and Sentiment Comovement

Does sentiment comove across countries? Is there global sentiment that affects sentiment in specific countries? In this section, we study the commonality of sentiment across countries. Moreover, we examine whether capital market integration affects sentiment comovement. To test this hypothesis, we use an event study: the implementations of IFRS across countries and over time. Finally, we investigate whether global sentiment predicts future market returns. We construct a global sentiment index and examine the extent to which global sentiment affects local sentiment. We simply average the sentiment indices of the 38 countries in our sample to form the global sentiment index.[31]

We next examine how a country's sentiment is explained by global sentiment in the following regression:

$$\text{SENTIMENT}_{i,t} = a + b\text{SENTIMENT_G}_t + \text{CONTROLS}_{i,t} + \varepsilon_{i,t}, \quad (6)$$

where $\text{SENTIMENT}_{i,t}$ is country i's sentiment in week t, and SENTIMENT_G_t is the simple average of $\text{SENTIMENT}_{i,t}$ for our 38 sample countries, except country i, in week t.

Table 8 reports the results of regressing a country's sentiment on global sentiment. First, models 1 and 2 show that on average, 2.7% of the variation in a country's sentiment is driven by global sentiment. Second, we document a significant difference between developed and emerging countries in models 3 and 4: Investor sentiment has a higher correlation with global sentiment in developed countries than in emerging countries, as developed countries are more integrated into the global market. In terms of economic significance, a 1-standard-deviation change in global sentiment is associated with a 17.2%-standard-deviation change in sentiment in developed markets and a 12.3%-standard-deviation change in sentiment in emerging markets.[32]

In Table IA13 of the Supplementary Material, we also examine whether non-economics- and economics-related sentiment contribute differently to the comovement of investor sentiment. We find that the comovement of economics-related sentiment is stronger than that of non-economics-related sentiment. It is plausible that large commonality in economics-related sentiment is driven by

[31] To avoid the mechanical correlation between the global sentiment index and country-specific indices, in our following comovement test, we use the simple average of the sentiment across all countries but the focal country. We also consider different weighted averages, including weighted population, weighted GDP per capita, weighted Internet usage, and weighted sentiment averages of Google's search market share; all of these measures yield similar results. To make our construction more transparent, we use the simple average method to construct the global sentiment index when reporting the following analysis.

[32] In addition, we perform the regression for each country and report the results on the left-hand side of Panel A of Table IA17 in the Supplementary Material. Apparently, sentiment in all countries except Pakistan has a positive relation with global sentiment.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

TABLE 8
Comovement of Sentiment

Table 8 reports the relation between market sentiment and global sentiment. We run the following regression:

$$\text{SENTIMENT}_{i,t} = a + b\text{SENTIMENT_G}_t + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},$$

where $\text{SENTIMENT}_{i,t}$ is country i's market sentiment in week t and SENTIMENT_G_t is the simple average of $\text{SENTIMENT}_{i,t}$ of our 38 sample countries, except country i, in week t. The sample period is from July 2004 to Dec. 2014. Models 1 and 2 use all 38 countries (WRD), and models 3 and 4 focus on developed countries (DEV) and emerging countries (EMG), respectively. We also report the difference in coefficients on SENTIMENT_G_t between developed countries and emerging countries. Variables are defined in Table 3. We include country fixed effects in all regressions. Standard errors are clustered at the country (C) and year-week (T) levels. t-statistics are reported in parentheses. p-values are reported in square brackets.

	Dependent Variable: SENTIMENT_t			
	WRD		DEV	EMG
	Model			
Variable	1	2	3	4
SENTIMENT_G_t	0.596 (13.85)	0.572 (12.99)	0.672 (10.78)	0.488 (8.64)
Diff. in SENTIMENT_G_t [p-value]			0.184 [0.015]	
$\text{EPU}_{\text{US},t}$		0.018 (1.35)	−0.017 (−0.42)	0.051 (1.66)
$\text{VIX}_{\text{US},t}$		−0.053 (−2.51)	−0.104 (−2.42)	−0.027 (−0.73)
$\text{ECONOMY}_{\text{US},t}$		−0.004 (−2.36)	−0.007 (−2.47)	−0.001 (−0.40)
VOLATILITY_t		−0.024 (−1.25)	−0.023 (−0.67)	−0.021 (−0.86)
RETURN_{t-1}		−0.009 (−6.89)	−0.012 (−4.31)	−0.007 (−4.99)
RETURN_{t-2}		−0.000 (−0.07)	−0.004 (−1.30)	0.002 (1.36)
RETURN_{t-3}		−0.002 (−1.65)	−0.000 (−0.22)	−0.004 (−2.25)
RETURN_{t-4}		−0.000 (−0.19)	−0.002 (−0.93)	0.001 (0.64)
Fixed effects Clustering	C CT	C CT	C CT	C CT
No. of obs. R^2	20,786 2.3%	19,836 2.7%	8,874 3.9%	10,962 2.0%

globally linked economic activities. That is, the economics-related sentiment component is likely triggered by global news about fundamentals, whereas the non-economics-related portion is more country specific and is unlikely to spill over to other countries. Next, we use an event study to explore the role of economic linkage, especially market integration, in sentiment comovement.

There are two mechanisms through which global sentiment travels around the world. First, the trading activities of foreign sentiment investors can influence the sentiment of domestic investors, which is similar to the correlated trading behavior of international and institutional investors (e.g., Kamara, Lou, and Sadka (2008), Karolyi, Lee, and Van Dijk (2012)).[33] Second, investor sentiment can be spread across countries through channels such as word-of-mouth,

[33] If foreign investors represent a group of marginal investors in the domestic market, the sentiment of foreign investors will drive asset prices to deviate from economic fundamentals in the domestic market.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

the Internet, and the media (e.g., Hong, Kubik, and Stein (2004), Tetlock (2007), and Baker et al. (2012)). Both of these mechanisms can be characterized as the effect of market integration (e.g., Bekaert, Harvey, Lundblad, and Siegel (2013), Carrieri, Chaieb, and Errunza (2013), and Ng, Solnik, Wu, and Zhang (2013)). We use IFRS as the basis of an event to examine the impact of market integration on sentiment comovement. Regulators advocate IFRS to enhance the comparability of financial statements (Daske, Hail, Leuz, and Verdi (2008)), and the increased comparability of financial statements across countries consequently benefits investors. For example, the literature shows that IFRS can improve information sharing and capital flows across countries, which contributes to a higher degree of capital market integration (DeFond et al. (2011), De George et al. (2016)). We thus use the adoption of IFRS to test how market integration affects comovement between a country's market sentiment and global sentiment.[34]

Specifically, we run the following regression:

$$\begin{aligned} (7) \qquad \text{SENTIMENT}_{i,t} \;=\; & \alpha + \beta_1 \text{SENTIMENT_G}_t \\ & + \beta_2 \text{SENTIMENT_G}_t \times \text{IFRS}_{i,t} \\ & + \beta_3 \text{IFRS}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t}, \end{aligned}$$

where $\text{SENTIMENT}_{i,t}$ is country i's market sentiment in week t, and SENTIMENT_G_t is global sentiment in week t. $\text{IFRS}_{i,t}$ equals 1 when country i adopts IFRS in year $t-1$, and 0 otherwise. Table IA15 in the Supplementary Material presents our main results, where model 1 is the baseline model without the IFRS dummy, and model 2 is our main regression with the IFRS dummy. Our focus is β_2, the interaction between global sentiment and the IFRS adoption dummy. As shown in Table IA15, β_2 is positive and significant, which suggests that global sentiment has a greater positive impact on sentiment in countries that adopted IFRS than in those that did not. The effect is also economically significant. For countries that did not adopt IFRS, a 1-standard-deviation increase in global sentiment is related to an 11.7%-standard-deviation increase in countries' market sentiment, whereas for countries that adopted IFRS, the corresponding increase is 17.8%. This result is consistent with our prediction that market integration increases comovement between countries' market sentiment and global sentiment.

Having documented the strong international commonality of sentiment, we investigate the return prediction of global sentiment across countries. We decompose the sentiment of a country into a global sentiment component and a local sentiment component, where local sentiment is defined as the residual of regressing the country's sentiment on global sentiment. We then regress future

[34]Table IA14 in the Supplementary Material presents the years of mandatory IFRS adoption for each country in our sample. We designate the 30 countries that adopted IFRS as the treatment group and the 8 countries that did not as the control group. European countries, Hong Kong, and South Africa adopted IFRS in 2005. Several countries, such as Israel and New Zealand, adopted IFRS between 2007 and 2013, whereas the remaining countries, such as Japan and India, never adopted IFRS during our sample period. Given the relatively wide distribution of the IFRS adoption years, we use the full sample period from 2004 to 2014 for our main analysis. In robustness tests, we also follow the literature in using event windows circa 2005.

market returns on both current global sentiment and current local sentiment.[35] The results are presented in Table IA16. Global sentiment, rather than local sentiment, plays a major role in predicting future returns for both developed and emerging countries; however, the return prediction of global sentiment is more pronounced in developed countries, whereas local sentiment has similar predictive power in both emerging and developed countries. These findings again imply that developed countries are well integrated into the global market and that their markets are thus more likely to be influenced by global sentiment.

IV. Conclusion

In this article, we use households' Google search behavior to construct weekly sentiment indices for 38 markets from 2004 to 2014. We take advantage of international settings to validate the non-economics-related component of our measure as a proxy for sentiment based on an event study of sports outcomes. We show that our sentiment measure is a contrarian predictor of country-level market returns. We further mitigate potential endogeneity issues by conducting two difference-in-differences analyses around plausibly exogenous shocks to difficult-to-value and limits-to-arbitrage channels. Finally, we document an important role of global sentiment in driving sentiment and predicting market returns across countries.

There is still much to be done for future research. With international sentiment indices, researchers could improve our understanding of international markets from a behavioral perspective. Moreover, heterogeneity across countries, capital market integration, and the implementation of specific policies and regulations provide rich opportunities to explore the sentiment effect in a broader context.

Supplementary Material

Supplementary Material for this article is available at https://doi.org/10.1017/S0022109019000061.

References

Ariel, R. A. "High Stock Returns before Holidays: Existence and Evidence on Possible Causes." *Journal of Finance*, 45 (1990), 1611–1626.

Aruoba, S. B.; F. X. Diebold; and C. Scotti. "Real-Time Measurement of Business Conditions." *Journal of Business and Economic Statistics*, 27 (2009), 417–427.

Baker, M., and J. Wurgler. "The Equity Share in New Issues and Aggregate Stock Returns." *Journal of Finance*, 55 (2000), 2219–2257.

Baker, M., and J. Wurgler. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance*, 61 (2006), 1645–1680.

[35] The specification is:

$$\text{RETURN}_{i,t+1} = a + b\text{SENTIMENT_G}_t + c\text{SENTIMENT_L}_{i,t} + \text{CONTROLS}_{i,t} + \varepsilon_{i,t},$$

where $\text{RETURN}_{i,t+1}$ is country i's market returns in week $t+1$, SENTIMENT_G_t is global sentiment in week t, and $\text{SENTIMENT_L}_{i,t}$ is country i's local sentiment in week t. We also perform the same analysis by country and report the results on the right-hand side of Panel A of Table IA17 in the Supplementary Material.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

Baker, M., and J. Wurgler. "Investor Sentiment in the Stock Market." *Journal of Economic Perspectives*, 21 (2007), 129–151.
Baker, M.; J. Wurgler; and Y. Yuan. "Global, Local, and Contagious Investor Sentiment." *Journal of Financial Economics*, 104 (2012), 272–287.
Baker, S. R.; N. Bloom; and S. J. Davis. "Measuring Economic Policy Uncertainty." Research Paper, Chicago Booth (2015).
Barber, B. M.; T. Odean; and N. Zhu. "Do Retail Trades Move Markets?" *Review of Financial Studies*, 22 (2009), 151–186.
Beber, A., and M. Pagano. "Short-Selling Bans around the World: Evidence from the 2007–09 Crisis." *Journal of Finance*, 68 (2013), 343–381.
Bekaert, G.; C. R. Harvey; C. T. Lundblad; and S. Siegel. "The European Union, the Euro, and Equity Market Integration." *Journal of Financial Economics*, 109 (2013), 583–603.
Brown, G. W., and M. T. Cliff. "Investor Sentiment and the Near-Term Stock Market." *Journal of Empirical Finance*, 11 (2004), 1–27.
Carrieri, F.; I. Chaieb; and V. Errunza. "Do Implicit Barriers Matter for Globalization?" *Review of Financial Studies*, 26 (2013), 1694–1739.
Choi, H., and H. Varian. "Predicting the Present with Google Trends." *Economic Record*, 88 (2012), 2–9.
Chu, Y.; D. Hirshleifer; and L. Ma. "The Causal Effect of Limits to Arbitrage on Asset Pricing Anomalies." Working Paper No. w24144, National Bureau of Economic Research (2017).
Cumming, D.; S. Johan; and D. Li. "Exchange Trading Rules and Stock Market Liquidity." *Journal of Financial Economics*, 99 (2011), 651–671.
Da, Z.; J. Engelberg; and P. Gao. "In Search of Attention." *Journal of Finance*, 66 (2011), 1461–1499.
Da, Z.; J. Engelberg; and P. Gao. "The Sum of All FEARS: Investor Sentiment and Asset Prices." *Review of Financial Studies*, 28 (2015), 1–32.
Das, S. R., and M. Y. Chen. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web." *Management Science*, 53 (2007), 1375–1388.
Daske, H.; L. Hail; C. Leuz; and R. Verdi. "Mandatory IFRS Reporting around the World: Early Evidence on the Economic Consequences." *Journal of Accounting Research*, 46 (2008), 1085–1142.
DeFond, M.; X. Hu; M. Hung; and S. Li. "The Impact of Mandatory IFRS Adoption on Foreign Mutual Fund Ownership: The Role of Comparability." *Journal of Accounting and Economics*, 51 (2011), 240–258.
De George, E. T.; X. Li; and L. Shivakumar. "A Review of the IFRS-Adoption Literature." *Review of Accounting Studies*, 21 (2016), 898–1004.
De Long, J. B.; A. Shleifer; L. H. Summers; and R. J. Waldmann. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98 (1990), 703–738.
Diether, K. B.; C. J. Malloy; and A. Scherbina. "Differences of Opinion and the Cross Section of Stock Returns." *Journal of Finance*, 57 (2002), 2113–2141.
Drake, M. S.; D. T. Roulstone; and J. R. Thornock. "Investor Information Demand: Evidence from Google Searches around Earnings Announcements." *Journal of Accounting Research*, 50 (2012), 1001–1040.
Edmans, A.; D. Garcia; and Ø. Norli. "Sports Sentiment and Stock Returns." *Journal of Finance*, 62 (2007), 1967–1998.
Evans, G. W.; S. V. Jacobs; D. Dooley; and R. Catalano. "The Interaction of Stressful Life Events and Chronic Strains on Community Mental Health." *American Journal of Community Psychology*, 15 (1987), 23–34.
Ferreira, M. A., and P. Matos. "The Colors of Investors' Money: The Role of Institutional Investors around the World." *Journal of Financial Economics*, 88 (2008), 499–533.
Ginsberg, J.; M. H. Mohebbi; R. S. Patel; L. Brammer; M. S. Smolinski; and L. Brilliant. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature*, 457 (2009), 1012–1014.
Gromb, D., and D. Vayanos. "Limits of Arbitrage: The State of the Theory." *Annual Review of Financial Economics*, 2 (2010), 251–275.
Hirshleifer, D. A.; D. Jiang; and Y. Meng. "Mood Beta and Seasonalities in Stock Returns." Working Paper No. w24676, National Bureau of Economic Research (2018).
Hirshleifer, D. A., and T. Shumway. "Good Day Sunshine: Stock Returns and the Weather." *Journal of Finance*, 58 (2003), 1009–1032.
Hirt, E. R.; D. Zillmann; G. A. Erickson; and C. Kennedy. "Costs and Benefits of Allegiance: Changes in Fans' Self-Ascribed Competencies after Team Victory versus Defeat." *Journal of Personality and Social Psychology*, 63 (1992), 724–738.
Hong, H.; J. D. Kubik; and J. C. Stein. "Social Interaction and Stock-Market Participation." *Journal of Finance*, 59 (2004), 137–163.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

Hong, H., and J. C. Stein. "Disagreement and the Stock Market." *Journal of Economic Perspectives*, 21 (2007), 109–128.
Hu, M., and B. Liu. "Mining and Summarizing Customer Reviews." *Proceedings of the Tenth Association for Computing Machinery SIGKDD International Conference on Knowledge Discovery and Data Mining*. Seattle, Washington, USA: Association for Computing Machinery (2004), 168–177.
Huang, D.; F. Jiang; J. Tu; and G. Zhou. "Investor Sentiment Aligned: A Powerful Predictor of Stock Returns." *Review of Financial Studies*, 28 (2015), 791–837.
Hwang, B. H. "Country-Specific Sentiment and Security Prices." *Journal of Financial Economics*, 100 (2011), 382–401.
Jain, A.; P. K. Jain; T. H. McInish; and M. McKenzie. "Worldwide Reach of Short Selling Regulations." *Journal of Financial Economics*, 109 (2013), 177–197.
Jones, C. M., and O. A. Lamont. "Short-Sale Constraints and Stock Returns." *Journal of Financial Economics*, 66 (2002), 207–239.
Kamara, A.; X. Lou; and R. Sadka. "The Divergence of Liquidity Commonality in the Cross-Section of Stocks." *Journal of Financial Economics*, 89 (2008), 444–466.
Kamstra, M. J.; L. A. Kramer; and M. D. Levi. "Winter Blues: A Sad Stock Market Cycle." *American Economic Review*, 93 (2003), 324–343.
Kaplanski, G., and H. Levy. "Sentiment and Stock Prices: The Case of Aviation Disasters." *Journal of Financial Economics*, 95 (2010), 174–201.
Karolyi, G. A.; K. H. Lee; and M. A. Van Dijk. "Understanding Commonality in Liquidity around the World." *Journal of Financial Economics*, 105 (2012), 82–112.
Keynes, J. M. *The General Theory of Interest, Employment and Money*. New York, NY: Macmillan (1936).
Kogan, L.; S. A. Ross; J. Wang; and M. M. Westerfield. "The Price Impact and Survival of Irrational Traders." *Journal of Finance*, 61 (2006), 195–229.
Kogan, L.; S. A. Ross; J. Wang; and M. M. Westerfield. "Market Selection." *Journal of Economic Theory*, 168 (2017), 209–236.
Lee, C. M.; P. Ma; and C. C. Wang. "Search-Based Peer Firms: Aggregating Investor Perceptions through Internet Co-Searches." *Journal of Financial Economics*, 116 (2015), 410–431.
Lemmon, M., and E. Portniaguina. "Consumer Confidence and Asset Prices: Some Empirical Evidence." *Review of Financial Studies*, 19 (2006), 1499–1529.
Lesmond, D. A.; J. P. Ogden; and C. A. Trzcinka. "A New Estimate of Transaction Costs." *Review of Financial Studies*, 12 (1999), 1113–1141.
Levy, T., and J. Yagil. "Air Pollution and Stock Returns in the US." *Journal of Economic Psychology*, 32 (2011), 374–383.
Miller, E. M. "Risk, Uncertainty, and Divergence of Opinion." *Journal of Finance*, 32 (1977), 1151–1168.
Mondria, J.; T. Wu; and Y. Zhang. "The Determinants of International Investment and Attention Allocation: Using Internet Search Query Data." *Journal of International Economics*, 82 (2010), 85–95.
Nagel, S. "Short Sales, Institutional Investors and the Cross-Section of Stock Returns." *Journal of Financial Economics*, 78 (2005), 277–309.
Ng, L.; B. Solnik; E. Wu; and B. Zhang. "Characterizing Global Financial and Economic Integration Using Cash Flow Expectations." Working Paper, University of Wisconsin–Milwaukee (2013).
Palla, G.; I. Derényi; I. Farkas; and T. Vicsek. "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society." *Nature*, 435 (2005), 814–818.
Pástor, L., and P. Veronesi. "Stock Valuation and Learning about Profitability." *Journal of Finance*, 58 (2003), 1749–1790.
Pástor, L., and P. Veronesi. "Rational IPO Waves." *Journal of Finance*, 60 (2005), 1713–1757.
Pástor, L., and P. Veronesi. "Was There a Nasdaq Bubble in the Late 1990s?" *Journal of Financial Economics*, 81 (2006), 61–100.
Qiu, L., and I. Welch. "Investor Sentiment Measures." Working Paper No. w10794, National Bureau of Economic Research (2006).
Shiller, R. J. "Speculative Prices and Popular Models." *Journal of Economic Perspectives*, 4 (1990), 55–65.
Shleifer, A., and R. W. Vishny. "The Limits of Arbitrage." *Journal of Finance*, 52 (1997), 35–55.
Sibley, S. E.; Y. Wang; Y. Xing; and X. Zhang. "The Information Content of the Sentiment Index." *Journal of Banking and Finance*, 62 (2016), 164–179.
Silver, R. C.; E. A. Holman; D. N. McIntosh; M. Poulin; and V. Gil-Rivas. "Nationwide Longitudinal Study of Psychological Responses to September 11." *Journal of the American Medical Association*, 288 (2002), 1235–1244.
Soo, C. K. "Quantifying Sentiment with News Media across Local Housing Markets." *Review of Financial Studies*, 31 (2018), 3689–3719.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press

Stambaugh, R. F.; J. Yu; and Y. Yuan. "The Short of It: Investor Sentiment and Anomalies." *Journal of Financial Economics*, 104 (2012), 288–302.

Stambaugh, R. F.; J. Yu; and Y. Yuan. "The Long of It: Odds that Investor Sentiment Spuriously Predicts Anomaly Returns." *Journal of Financial Economics*, 114 (2014), 613–619.

Steyvers, M., and J. B. Tenenbaum. "The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth." *Cognitive Science*, 29 (2005), 41–78.

Tetlock, P. C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance*, 62 (2007), 1139–1168.

Wiebe, J.; T. Wilson; and C. Cardie. "Annotating Expressions of Opinions and Emotions in Language." *Language Resources and Evaluation*, 39 (2005), 165–210.

Yu, J., and Y. Yuan. "Investor Sentiment and the Mean-Variance Relation." *Journal of Financial Economics*, 100 (2011), 367–381.

https://doi.org/10.1017/S0022109019000061 Published online by Cambridge University Press