DOI: 10.1111/fire.12276

PAPER SUBMITTED FOR REVIEW

The Financial Review WILEY

Investor awareness or information asymmetry? Wikipedia and IPO underpricing

Thomas Boulton[1] | **Bill B. Francis**[2] | **Thomas Shohfi**[2] | **Daqi Xin**[3]

[1] Farmer School of Business, Miami University, Oxford, Ohio, USA

[2] Lally School of Management, Rensselaer Polytechnic Institute, Troy, New York, USA

[3] Nankai Business School, Nankai University, Tianjin, China

Correspondence
Thomas Shohfi, Lally School of Management, Rensselaer Polytechnic Institute, Pittsburgh Bldg, 110 8th St, Troy, NY 12180, USA.
Email: shohft@rpi.edu

Abstract

We use the presence of a Wikipedia article for initial public offering (IPO) firms to test theories of information asymmetry and investor awareness. Although we find limited support for the former, our results provide strong support for theories of investor awareness. Specifically, IPO firms with a Wikipedia article exhibit significantly higher underpricing and offer price revisions than do IPO firms without a Wikipedia article. Investor awareness has positive long-term effects, including greater analyst following and institutional ownership for up to 3 years after the offering. The effect is robust to firm-specific Google search volume, news coverage, retail trading intensity, social media activity, propensity score matching, and an instrumental variable approach.

KEYWORDS
initial public offerings, investor attention, underpricing, Wikipedia

JEL CLASSIFICATION
D82, G14, G24, G32

1 | INTRODUCTION

The manner in which the world consumes information has changed dramatically since researchers began studying initial public offering (IPO) underpricing (Ibbotson, 1975; Logue, 1973). Possibly, no invention since the television in the 1920s has done more to democratize the availability of information than the Internet. In recent years, the Internet has evolved from a medium to consume information passively to a place where users collaborate to create content. There is perhaps no better example of this collaborative effort than Wikipedia, the leading free online encyclopedia where anyone can create and edit content.

As private companies or subsidiaries of public companies, information about IPO firms is often limited. For many potential investors, the issuer's carefully crafted registration statement is the primary source of information used to

 © 2021 The Eastern Finance Association

evaluate the IPO as an investment opportunity. Federal securities laws limit the information that issuers and their representatives can share with the public between the time the registration statement is filed and declared effective by the U.S. Securities and Exchange Commission (SEC).[1] However, these limitations do not apply to the collaborative efforts of the Wikipedia community, which makes Wikipedia a potentially valuable source of information for IPO investors.

Wikipedia ranks among the 10 most popular Web sites in the United States and is the world's leading online reference source.[2] Despite Wikipedia's potential importance as a source of information for investors, its impact on IPO firms is an unexplored issue. However, evidence suggests that potential investors reference a company's Wikipedia article around its IPO. To illustrate, we report mean traffic for Wikipedia articles for 30 days around IPO filing and issuance dates in Figure 1. Panel a indicates a 45.4% increase in article views at the IPO filing date relative to the 30-day mean. Panel b shows a much larger traffic increase of 143.2% on the IPO issuance date that continues into the following day (70.46% increase). These results motivate us to examine the impact of Wikipedia on IPO pricing and long-run outcomes.

We contend that Wikipedia could impact interactions between IPO issuers and the investment banks they employ (Baron, 1982), IPO issuers and potential investors (Welch, 1989), and different investor groups (Rock, 1986). Information disparities make it difficult to precisely price a firm's IPO (Bradley et al., 2004) and are believed to contribute to underpricing that results in the large first-day gains exhibited by many IPOs (Ljungqvist, 2007). The economic consequences of these information effects are significant. For instance, Ritter (1987) and others find that underpricing is the largest single cost of going public for the majority of IPO issuers.

Wikipedia also has the potential to increase investor attention to IPO firms. Prior research finds that stock prices do not fully reflect value-related information until the information grabs investor attention (Frederickson & Zolotoy, 2016; Hirshleifer et al., 2004; Hong et al., 2007). Recent studies show that the media plays an important role in the information environment of capital markets (Bhattacharya et al., 2009; Engelberg & Parsons, 2011; Fang & Peress, 2009; Tetlock, 2007). The manner in which investors obtain information from Wikipedia, however, differs from traditional and social media. The latter are "pushed" to investors through newspapers and online platforms, whereas information from Wikipedia is typically "pulled" (i.e., it is specifically sought out). For example, according to SimilarWeb, 85.89% of Wikipedia's traffic is from active searching.[3] Consistent with the idea that Wikipedia is related to investor attention, we find that the presence of a Wikipedia article is positively correlated with measures of investor attention previously considered in the literature, including news coverage (Liu et al., 2014) and StockTwits activity (Cookson & Niessner, 2020). However, multivariate analysis indicates that much of the variation in the presence of a Wikipedia article is not explained by other measures of investor attention. Thus, we posit that the presence and content of a Wikipedia article could have an incremental effect on IPO outcomes due to Wikipedia's impact on investor attention.

We find that firms with a Wikipedia article when they go public (hereafter Wikipedia firms) experience significantly higher underpricing than firms without a Wikipedia article (21.0% vs. 12.7%). The association between the existence of and number of visits to a Wikipedia article and IPO underpricing is also evident in multivariate regressions that control for firm- and issue-related factors that have been shown to affect underpricing, and to a variety of robustness checks including controlling for abnormal firm-specific Google search volume (Da et al., 2011), social media activity, and news outlet intensity around the IPO. Additionally, we employ instrumental variable regression and propensity score matching methods. Our results consistently point to a statistically significant and economically large positive relation between the presence of a pre-IPO Wikipedia article and IPO underpricing. Because this is inconsistent with the notion that a Wikipedia article reduces information asymmetry, we draw on prior research on investor attention to explain why Wikipedia firms tend to experience greater IPO underpricing.

[1] https://www.sec.gov/fast-answers/answersquiethtm.html

[2] http://www.alexa.com/siteinfo/wikipedia.org

[3] https://www.similarweb.com/website/wikipedia.org#search (as of July 2, 2019)

Da, Engelberg, & Gao (2011) and Liu, Sherman, & Zhang (2014) provide evidence that investor attention is associated with IPO outcomes. An important difference between these studies is that they differ on whether transient retail investors or longer horizon institutional investors drive the impact of investor attention on IPO outcomes. Da, Engelberg, & Gao (2011) find that high initial returns are followed by long-run underperformance for IPOs that receive high investor attention. This is consistent with Barber & Odean (2008), who find that individual investors tend to buy "attention-grabbing" stocks, which generates temporary price pressure that leads to higher stock prices and lower future returns. However, using the trade classification method of Boehmer et al. (2021), we show that Wikipedia's association with IPO underpricing is not principally driven by retail investors.

Liu, Sherman, & Zhang (2014) provide evidence consistent with the Merton (1987) investor recognition model that predicts that increased investor attention has positive long-term effects for firms. Consistent with their results, we find that the presence of a pre-IPO Wikipedia article is associated with long-term attention measures, including greater analyst following and more institutional investors. Combined with the fact that Wikipedia firms are more likely to revise their offer price upward during the bookbuilding period, which is dominated by institutional investors (Benveniste & Spindt, 1989), our results indicate that institutional investor attention plays an important role in the positive relation between Wikipedia and IPO underpricing.

To summarize, we provide evidence that Wikipedia's user-generated content captures the attention of primary capital market participants. The fact that the presence of a Wikipedia article is associated with significantly higher IPO underpricing demonstrates the importance of the collaborative efforts of the Wikipedia community. Our results complement recent studies of Internet stock message boards (Antweiler & Frank, 2004), financial blogs (Saxton & Anker, 2013), crowdsourced financial research Web sites (Chen et al., 2014), Facebook (Zhou et al., 2015), and Twitter (Blankespoor et al., 2014). Similar to Wikipedia, these platforms use decentralized channels to aggregate information. This differs from corporate disclosure and media coverage where information is diffused unidirectionally, which raises concerns about objectivity and validity (Brown & Hillegeist, 2007; Gao & Ritter, 2010; Verrecchia, 1983). Our results also have implications for the emerging field of social economics and finance, which posits that investor sentiment is a social phenomenon influenced by factors including changes in the way that investors communicate with one another (Hirshleifer, 2020).

To the best of our knowledge, we provide the first in-depth analysis of the relation between the most prominent user-generated reference source on the Internet and IPO underpricing. Our finding that Wikipedia firms have significantly higher underpricing contrasts with the notion that underpricing is compensation to investors for limited information (Rock, 1986) and is consistent with the notion that Wikipedia articles increase investor attention (Da, Engelberg, & Gao, 2011; Liu, Sherman, & Zhang, 2014). One possible concern is the reliability of information reported on Wikipedia. Greenstein & Zhu (2012) identify three tenets that Wikipedia articles seek to attain: a neutral point of view, verifiability, and the absence of original research. Although there is some evidence of bias and slant in Wikipedia articles on controversial topics involving subjective information (Greenstein & Zhu, 2012; Greenstein et al., 2016), information provided on Wikipedia tends to be accurate in the areas of history (Holman Rector, 2008), medicine (Devgan et al., 2007), pharmacology (Clauson et al., 2008), philosophy (Bragues, 2007), politics (Brown, 2011), and science (Giles, 2005). Given that information in articles about companies with a forthcoming IPO is unlikely to be controversial or subjective, we contend that our findings are not driven by bias or slant in the Wikipedia articles.

Our study contributes to the burgeoning literature on the role of media in financial markets. For example, studies find that local media coverage is associated with local trading activities (Engelberg & Parsons, 2011), that media sentiment predicts stock returns and trading behavior (Tetlock, 2007), and that traditional media coverage predicts lower subsequent stock volatility and turnover (Jiao et al., 2020). In the case of IPOs, studies show that more media coverage during the quiet period is associated with more attention-driven retail purchases (Bushee et al., 2020), media sentiment and first-day returns are positively correlated (Bajo & Raimondo, 2017), and long-run returns are lower for IPOs with more pre-IPO newspaper articles (You et al., 2018). However, the unique information structure of Wikipedia distinguishes it from traditional media and most social media platforms. Specifically, Wikipedia is organized by topic and information accumulates over time due to contributions of the Wikipedia user community. Comparatively, information on traditional and social media is more dispersed. Given that individuals have limited information processing

ability (Hirshleifer & Teoh, 2003; Hong & Stein, 1999), Wikipedia is likely to draw attention from investors due to significantly reduced information acquisition and processing costs (Wu et al., 2014).

2 | WIKIPEDIA AND IPO UNDERPRICING

2.1 | Wikipedia

Launched on January 15, 2001, Wikipedia is one of the most popular Web sites in the world with nearly 870 million unique visitors each month.[4] As of December 2020, Wikipedia has 40.7 million registered users and 129,000 active contributors.[5] The English Wikipedia, which is one of 317 international editions, includes more than 6.2 million articles and typically receives 3–4 billion page views per month.[6] Wikipedia is so ubiquitous that *Time* magazine recently named it the third most influential Web site of all time.[7]

The basic unit of information on Wikipedia is an article, which distinguishes it from social media platforms such as Facebook and Twitter. An important characteristic of Wikipedia is that articles evolve over time from the collaborative effort of the Wikipedia community. Wikipedia applies several mechanisms to improve the authenticity of content. For example, the "pending changes" system requires an established Wikipedia editor to review edits made by new users.[8] Kumar et al. (2016) report that 90% of hoaxes submitted to Wikipedia are caught in under an hour, suggesting that the editorial process is effective in policing user contributions.

Articles describing private and public firms are an influential component of Wikipedia. Such articles typically start with a general description of the company, followed by sections that detail company history, events, products, organizations, strategies, and competitors. Important events are usually reported in a standardized format such as "On [Day Month, Year], [the company] [did something]." When a company's name is used in a search query, its Wikipedia article is generally prioritized. A study conducted in 2012 finds that Wikipedia articles appear on the first page of results for 99% of Google searches and as the very first result for 56% of searches.[9]

Previous studies provide evidence that market participants seek company-related information from Wikipedia. For instance, Xu & Zhang (2013) find that management disclosure and investor reaction are influenced by the presence of a Wikipedia article about the firm. A survey of business journalists, analysts, and investors who were asked about preferred sources of information other than firms' corporate Web sites concluded, "Wikipedia is the most popular social media site for individuals looking for such information, used by more than three quarters of respondents" (Bradshaw, 2008). Another survey reports that Wikipedia is the top social media Web site used by business journalists, sell-side analysts, and buy-side analysts to find general firm information (Comprend, 2015).

2.2 | Initial public offerings

2.2.1 | Information asymmetry

IPO firms are private companies or subsidiaries of public companies prior to the offering, meaning that information about them is limited. Prior research suggests that limited information about IPO firms contributes to information

[4] Wikipedia 2020 unique visitor statistics.

[5] https://en.wikipedia.org/wiki/Special:Statistics

[6] https://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm

[7] http://time.com/4960202/most-influential-websites/

[8] "Wikipedia unlocks divisive pages for editing," BBC News, June 5, 2010.

[9] https://www.searchenginewatch.com/2012/02/13/wikipedia-appears-on-page-1-of-google-for-99-of-searches-study

disparities between issuers and underwriters (Baron, 1982), issuers and IPO investors (Welch, 1989), and different investor groups (Rock, 1986). These information disparities are thought to drive the large first-day returns observed for many IPO firms (e.g., Ljungqvist, 2007). Because underpricing reduces the proceeds that an IPO firm raises, it represents a substantial portion of the cost of going public for many firms (Ritter, 1987).

Evidence indicates that IPO firms and their representatives take actions to reduce information asymmetry and improve IPO outcomes. Some firms attempt to improve the information environment by providing more timely and informative disclosures to investors (Jog & McConomy, 2003; Leone et al., 2007). The creation and cultivation of a Wikipedia article could have a similar effect on the information environment of IPO firms. If Wikipedia is associated with better information dissemination and therefore less information asymmetry, then we should observe a negative relation between a pre-IPO Wikipedia article and IPO underpricing. Thus, our first hypothesis is as follows:

H1: *The presence of a pre-IPO Wikipedia article is negatively correlated with initial returns.*

2.2.2 | Investor attention

Investors have limited information processing ability, which makes attention a valuable resource. Merton (1987) points out that investors who are unfamiliar with a firm are unlikely to include it in their portfolio and, due to "setup" costs, are less likely to respond to firm-specific announcements. In his model, an increase in investor awareness could have positive long-run effects for a firm. For example, a larger investor base is associated with a lower cost of capital and higher market valuation. This suggests that firms have incentives to promote investor awareness.

Barber & Odean (2008) find that individual investors tend to buy "attention-grabbing" stocks, which results in an increase in stock prices. Because investors have many choices of stocks to buy, attention helps to narrow investors' choice set (Odean, 1999). Attention-induced price increases should be short-lived if they result from temporary price pressure and not information about firm fundamentals. IPO firms also need to attract attention to sell their shares to investors and create a liquid secondary market. Prior studies on the effect of investor attention on IPO outcomes differ concerning the long-term impact of investor attention on IPO firms. Da, Engelberg, & Gao (2011) find that short-term investor attention measured by Google search volume predicts long-run underperformance of IPO stocks, while Liu, Sherman, & Zhang (2014) find that investor attention, measured by pre-IPO media coverage, has a positive effect on IPO firms' long-term value, liquidity, analyst coverage, and institutional ownership. Liu, Sherman, & Zhang (2014) suggest that the difference between the two studies may result from the type of investor attention that is captured (i.e., retail vs. institutional).

A Wikipedia article indicates that a collaborative effort is underway to gather and report information on the company when other available information sources are limited. Moreover, given the high reliability of Wikipedia, Wikipedia could also act as an accreditor for the legitimacy of IPO firms. For instance, according to a survey conducted in 2014, 64% of the British public trust Wikipedia more than news media including the BBC.[10] In the context of IPOs, sparsity of firm information could lead to an increased reliance on Wikipedia content by potential investors. If firms with a Wikipedia article are more likely to grab investor attention, which Da, Engelberg, & Gao (2011) find is positively correlated with first-day returns, we predict the following:

H2: *The presence of a pre-IPO Wikipedia article is positively correlated with initial returns.*

[10] https://www.cnbc.com/2014/08/11/brits-trust-wikipedia-more-than-the-news-survey.html

3 | DATA AND METHODOLOGY

3.1 | Sample selection

We begin by collecting completed US IPOs with an offer price of at least $5 per share issued between 2006 and 2016 from the Thomson Reuters Securities Data Company (SDC) New Issues database. Although Wikipedia was launched in 2001, we begin our IPO sample in 2006 for several reasons. First, there are no IPO firms with a Wikipedia article at the time of their offering before 2004. Second, we find that the informativeness of Wikipedia articles before 2006 is limited. For example, the average number of words in an IPO firm's Wikipedia article increases from 100 in 2005 to 400 in 2006. Third, during its early years, Wikipedia had low awareness among Internet users. According to alexa.com, Wikipedia traffic ranked in the top 500 Web sites in October 2004, top 100 in April 2005, and top 30 in January 2006.[11]

Following prior IPO literature, we exclude foreign issuers, American Depositary Receipts, closed-end funds, natural resource limited partnerships, real estate investment trusts, unit offers, small best efforts offerings, financial firms, and stocks not covered by the CRSP database. After imposing these filters, we are left with a sample of 974 IPOs. We retrieve stock price and return data from CRSP and accounting data from COMPUSTAT.

3.2 | Variables

We use a web crawler to search for an IPO firm's Wikipedia article and manually check its accuracy. We assign a Wikipedia article to an IPO if the article is titled with the name of (1) the IPO firm, (2) the IPO firm's parent company if it is the IPO firm's parent before the first trading date, (3) the IPO firm's major subsidiary, (4) a company from which the IPO firm separates, (5) the firm's predecessor, or (6) the core product or service and primarily contains information about the firm.[12] We provide further details of our Wikipedia article identification process in Exhibit A of the Appendix in the Supporting Information.

To help identify the presence and content of a firm's Wikipedia article at the time of the public offering, we use the "date of page creation" provided for each Wikipedia article. Exhibit B of the Appendix in the Supporting Information contains LinkedIn's "article information" that reports basic information including article length, page ID, number of article watchers, article creator, date of article creation, and total number of edits. At the bottom of the information page, there are "external tools" links to revision history, page view statistics, and other information. For IPO firms with a Wikipedia article prior to the first trading day, we set the indicator variable *Wikipedia* equal to 1, and 0 otherwise. We identify 330 firms that have a Wikipedia article at the time of their IPO.

The relevant Wikipedia article for our analysis is the last historical version prior to the first trading day. We use the "view history" tag to access historical versions. Information on the "revision history" page includes time of modification, editor, IP address, flag of minor edit,[13] and net article change size. In Exhibit C of the Appendix in the Supporting Information, for purpose of illustration, we provide LinkedIn's Wikipedia article as of its IPO date (May 19, 2011). Its revision history indicates that the Wikipedia community quickly impounds IPO information. In the calendar week prior to the IPO date, the article averages 0.9 revisions per day. During the 3-day window centered on LinkedIn's IPO date, there are 19 revisions. Consistent with Wikipedia's verifiability tenet (Greenstein & Zhu, 2012), references used to compose the article are listed at the end.

[11] https://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than…

[12] Of 330 IPOs, 306 have a Wikipedia article titled with its own company name. Results are qualitatively unaltered if we exclude the other 24 IPO companies.

[13] An editor can mark a modification as a "minor edit" if he or she believes differences between the new version and the previous version do not require additional review. Typographical corrections and reformatting are common examples.

Typically, information from a company's S-1 filing is rapidly integrated into its Wikipedia article.[14] For example, LinkedIn filed its S-1 with the SEC on January 27, 2011. On the next day, the following was added to the "Company background" section: "LinkedIn filed for IPO on 27 January 2011. The listing could raise $175 million. According to the prospectus, the company's revenue doubled for the first 9 months of 2010." On May 19, information about the initial pricing and trading of LinkedIn's IPO was added. Given Wikipedia's detailed editing history, we can compare any two versions of a Wikipedia article and identify differences. We demonstrate this in Exhibit D.2 of the Appendix in the Supporting Information, where we show a comparison of two historical articles for LinkedIn. To capture the information aggregation and quality of a Wikipedia article, we construct three variables based on the latest historical version of an IPO firm's Wikipedia article. First, *wiki_revisions* is equal to the total number of Wikipedia article revisions from the S-1 filing date to the day before the issue date. Second, *wiki_references* is the number of references in a Wikipedia article. Third, *wiki_words* is the number of words in the main text of a Wikipedia article.

Following prior literature, we construct a number of measures related to the IPO event. These variables include offer price revision, venture capital backing, top-tier underwriter, share overhang, IPO proceeds, and other IPO firm and event characteristics. Detailed definitions of all variables are provided in the Appendix.

4 | MAIN RESULTS

4.1 | Descriptive statistics

In Figure 2, we report the total number of IPOs with and without a Wikipedia article for each year during our sample period. The number of IPOs without a Wikipedia article drops from 105 in 2006 to 11 in 2008 during the financial crisis. As IPO activity resumes following the crisis, the number of IPOs without a Wikipedia article reaches 105 in 2014 before falling to 61 in 2015 and 51 in 2016. Comparatively, the number of IPOs with a Wikipedia article exhibits a similar pattern but with lower volatility. After reaching a trough in 2008 with six IPOs, the number of IPOs with a Wikipedia article increases gradually to 62 in 2014 before falling to 19 in 2016.

We report summary statistics for IPOs with and without a pre-IPO Wikipedia article in Table 1, Panel A. We winsorize all continuous variables at the top and bottom 1% to limit the influence of extreme values. The difference in average underpricing for IPOs with and without a Wikipedia article is striking (21.0% and 12.7%, respectively). This difference provides preliminary support for the investor attention hypothesis. The average offer price of IPOs with a Wikipedia article is adjusted upward from the midpoint of the initial filing range by 1.76%. This compares to a downward adjustment of 8.22% for IPOs without a Wikipedia article. Wikipedia firms are less likely to be backed by a venture capital investor and more likely to hire a top-tier investment bank as their underwriter. *Overhang* indicates that 4.24 (3.16) shares are retained for each share sold by firms with (without) a pre-IPO Wikipedia article. Wikipedia firms are more likely to have positive earnings and tend to report greater sales and total assets than IPO firms without a Wikipedia article. Thirteen percent more IPO firms are classified as high tech in the Wikipedia sample.[15] Wikipedia firms also have a longer history, more news coverage, and higher levels of post-IPO institutional ownership. On their IPO issue date, Wikipedia firms have greater overall share and retail trading volume (Boehmer, Jones, Zhang, & Zhang, 2021), similar Google *ASVI* (Da, Engelberg, & Gao, 2011), and 12.6 more *stock_twits* (Cookson & Niessner, 2020). In sum, these results indicate that there are significant differences between IPOs with and without a Wikipedia article.

In Table 1, Panel B, we report summary statistics for four Wikipedia-specific variables: *wiki_revisions*, *wiki_references*, *wiki_words*, and *wiki_traffic*. These variables suggest that there is substantial heterogeneity with respect to the amount of information aggregated and attention received for IPO firms with Wikipedia articles.

[14] We provide examples of revisions to LinkedIn's Wikipedia article in Exhibit D.1 of the Appendix in the Supporting Information.

[15] The industry distribution for our IPO sample is reported in Figure A.1 of the Appendix in the Supporting Information.

Panel A. IPO filing date

Panel B. IPO issuance date

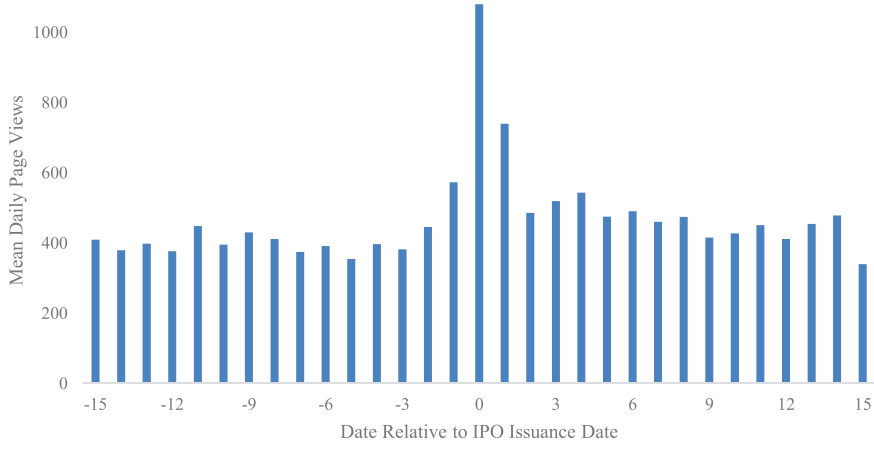

FIGURE 1 Wikipedia article traffic. This figure displays Wikipedia traffic relative to the IPO filing date (Panel (a)) and IPO issuance date (Panel (b)) using Wikipedia article-level page view data available at https://dumps.wikimedia.org/other/analytics/ and aggregated by http://wikishark.com

4.2 | Determinants of a Wikipedia article

Due to the differences between IPOs with and without a Wikipedia article reported in Table 1, we examine the determinants of the existence of a Wikipedia article when firms go public. Table 2 reports the results of a probit model with the dependent variable set equal to 1 for IPOs with a Wikipedia article, and 0 otherwise. Independent variables include relevant IPO and firm characteristics defined in Table 1, including an important proxy for investor attention, the number of news articles about the IPO firm between the S-1 filing and IPO date (*log_news*). We find that IPO firms with top-tier underwriters, greater share overhang, more sales, longer history, and more news coverage are more likely to have a Wikipedia article when they go public. In sum, the evidence indicates that more established firms are more likely to have a Wikipedia article. To the extent that a Wikipedia article is merely a proxy for firm visibility, we would expect to observe lower underpricing for IPOs with a Wikipedia article because more established firms are less risky, on average (Loughran & Ritter, 2004).

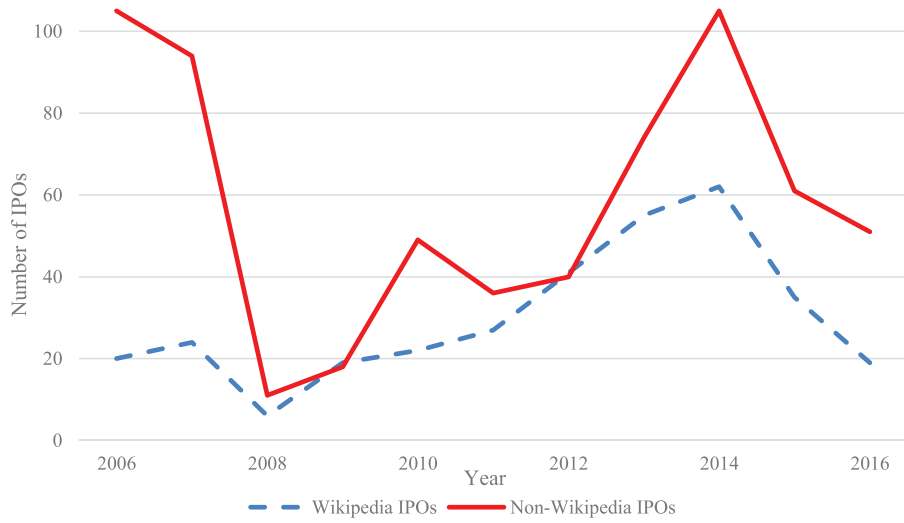

FIGURE 2 Number of IPOs by year with and without a Wikipedia article. This figure depicts the number of IPO firms with and without a pre-IPO Wikipedia article by year. The horizontal axis is the IPO year, and the vertical axis is the number of IPOs

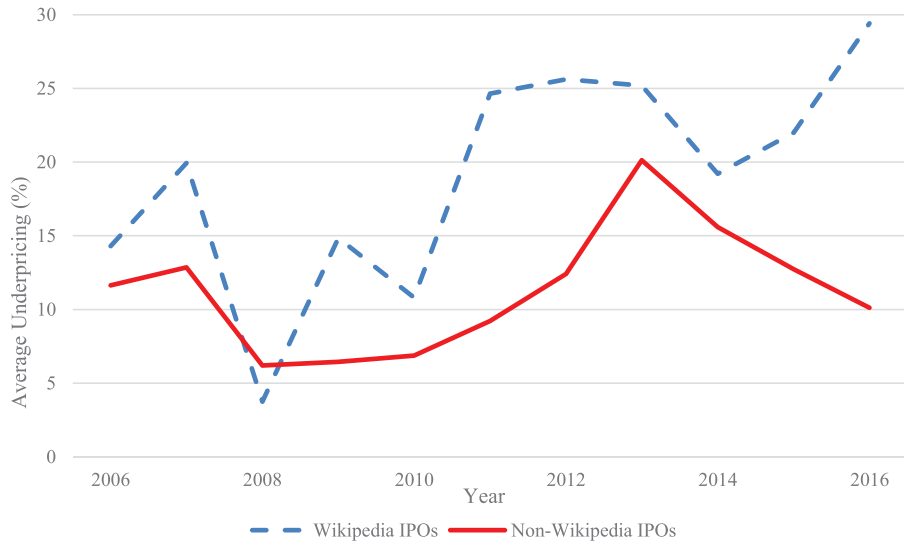

FIGURE 3 Average underpricing by year for IPOs with and without a Wikipedia article. This figure reports average underpricing for IPOs with and without a pre-IPO Wikipedia article by year. The horizontal axis is IPO year, and the vertical axis is average underpricing

4.3 | IPO underpricing

In Table 1, we report that average underpricing is 8.27 percentage points higher for IPO firms with a Wikipedia article. In Figure 3, we display the average underpricing for IPO firms with and without a Wikipedia article over our sample period. In every year but 2008, average underpricing is higher for IPO firms with a Wikipedia article than for firms without one. Although underpricing of Wikipedia IPOs is more volatile, the two samples exhibit similar underpricing

TABLE 1 Descriptive statistics

Panel A: Comparison of Wikipedia and non-Wikipedia IPO characteristics						
	Wikipedia IPOs (*N* = 330)		Non-Wikipedia IPOs (*N* = 644)			
	Mean	SD	Mean	SD	Difference	*t*-Stat
underpricing	20.99	28.68	12.72	22.23	8.27	4.58***
offer_revision	1.76	19.35	−8.22	20.45	9.96	7.47***
VC	0.45	0.50	0.56	0.49	−0.11	−3.10**
top_tier	0.94	0.23	0.76	0.42	0.18	8.64***
overhang	4.24	2.54	3.16	2.12	1.08	6.61***
pos_EPS	0.48	0.50	0.34	0.47	0.14	4.07***
sales	1210.42	2369.09	274.48	759.11	935.94	7.00***
tech	0.42	0.49	0.29	0.45	0.13	3.97***
age	24.82	27.24	15.73	19.29	9.09	5.40***
news	7.24	11.84	2.86	4.89	4.38	6.44***
nasdaq15	0.69	3.30	0.99	3.14	−0.30	−1.36
offering_size	402.41	1293.14	142.97	204.66	259.44	3.62***
adexptosales	0.03	0.08	0.01	0.04	0.02	4.962***
analyst_preIPO	0.16	0.71	0.03	0.27	0.13	3.18**
ASVI	2.31	7.42	3.03	8.43	−0.72	−1.31
stock_twits	16.01	29.32	6.07	12.61	9.94	7.38***
assets	3557.96	18,316.16	419.79	1613.58	3138.17	3.11**
offer_width	14.76	4.38	15.43	4.95	−0.67	2.152*
shares_filed	19.32	38.90	9.13	9.33	10.19	4.69***
instown_pct_post	0.39	0.29	0.34	24.89	0.05	2.70**
proceeds	402.41	1293.14	142.97	204.66	259.44	3.62***
vol_retail_pct	3.32	2.56	2.76	2.72	0.56	3.09***

Panel B: Wikipedia-specific variables						
	n	Mean	SD	25th	Median	75th
wiki_revisions	330	18.01	35.28	2.00	7.00	17.00
wiki_references	330	15.87	26.12	3.00	8.00	20.00
wiki_words	330	539.68	621.41	191.00	343.00	626.00
wiki_traffic	286	1539.73	9997.81	57.00	222.00	582.00

Note: This table reports summary statistics of Wikipedia IPOs and non-Wikipedia IPOs from 2006 to 2016. Variables are winsorized at the 1st and 99th percentiles. Panel A compares IPO and firm characteristics of Wikipedia IPOs and non-Wikipedia IPOs. Panel B reports Wikipedia-specific variables. The Appendix provides variable definitions.
***, **, and * indicate that the means are significantly different at the 1%, 5%, and 10% levels, respectively.

patterns. The exception is 2015 and 2016, when underpricing increases substantially for firms with a Wikipedia article and falls for firms without a Wikipedia article.

In Table 3, we report ordinary least squares (OLS) regressions that control for other factors shown in prior studies to be correlated with IPO underpricing. All models also include year fixed effects and robust standard errors clustered

TABLE 2 Likelihood of having a Wikipedia article at IPO

Variables	(1) Wikipedia
VC	0.0754
	(0.190)
top_tier	0.472***
	(0.143)
overhang	0.0621**
	(0.0314)
pos_EPS	−0.0255
	(0.161)
log_sales	0.117***
	(0.0444)
tech	0.301
	(0.210)
log_age	0.202***
	(0.0602)
log_news	0.193***
	(0.0595)
Constant	−2.511***
	(0.343)
N	974
Year FE	Yes
Pseudo-R^2	0.207

Note: This table reports probit regression results with a Wikipedia indicator as the dependent variable for IPOs from 2006 to 2016. The Appendix provides variable definitions. Standard errors reported in parentheses are robust and adjusted for clustering by Fama–French (1997) 48-industry and year.
*** and ** indicate significance at the 1% and 5% levels, respectively.

on IPO year and industry. In Model 1, we report the results of our baseline regression of underpricing on an indicator variable (*Wikipedia*) that is set equal to 1 for IPOs with a Wikipedia article, and 0 otherwise. Consistent with our univariate results, we find that the presence of a Wikipedia article is associated with underpricing that is 7.7 percentage points higher than for IPOs without a Wikipedia article. In Model 2, we include firm and IPO control variables and find that IPO firms with a Wikipedia article experience underpricing that is 5.9 percentage points higher. To illustrate the economic significance of this result, consider the average proceeds raised for our IPO sample ($231 million). An additional 5.9 percentage points of underpricing corresponds to a $13 million decrease in total IPO proceeds.

Because Wikipedia is an information aggregation platform, we include *log_news* to control for other information sources that capture investor attention (Table 3, Model 3). When *log_news* is added to the model, the magnitude and significance of the coefficient on *Wikipedia* decreases but remains statistically significant.[16] This is consistent with the notion that *Wikipedia* captures investor attention beyond what is included in traditional news coverage.

[16] The results in Table 3 are robust to the endogenous treatment regression of Heckman (1978) using the inverse Mills ratio from the probit model in Table 2. The results also remain significant when adding controls for Ravenpack news sentiment and the interaction between *log_news* and news sentiment.

TABLE 3 IPO underpricing

Variables	(1) Underpricing	(2) Underpricing	(3) Underpricing	(4) Underpricing
Wikipedia	7.686***	5.878**	4.950*	
	(1.788)	(2.162)	(2.325)	
log_wiki_traffic				1.603***
				(0.395)
VC		8.479***	8.466***	2.073***
		(2.517)	(2.564)	(0.650)
top_tier		6.464***	6.650***	8.285***
		(1.162)	(1.260)	(2.529)
overhang		1.573***	1.430***	6.315***
		(0.295)	(0.269)	(1.279)
pos_EPS		4.414*	4.229	1.238***
		(2.432)	(2.472)	(0.219)
log_sales		−0.0736	−0.201	4.044
		(0.602)	(0.597)	(2.324)
nasdaq15		0.437	0.392	−0.288
		(0.318)	(0.309)	(0.600)
tech		2.348	2.573	0.432
		(2.966)	(2.862)	(0.332)
log_age		−1.340	−1.344	2.190
		(1.419)	(1.371)	(2.659)
log_news			2.480**	−1.299
			(0.797)	(1.402)
Constant	12.92***	−0.932	−2.108	−1.011
	(1.494)	(3.857)	(3.771)	(3.881)
N	974	974	974	930
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.035	0.112	0.121	0.134

Note: This table displays OLS regressions that include 974 US IPOs from 2006 to 2016. Underpricing is the percent change from offer price to the first closing price. The Appendix provides variable definitions. Standard errors in parentheses are robust and adjusted for clustering by Fama–French (1997) 48-industry and year.
***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

To the extent that the existence of a Wikipedia article proxies for investor attention, an underlying assumption is that investors refer to Wikipedia to gather information about IPO firms. To provide additional evidence that Wikipedia proxies for investor attention, we extract the number of Wikipedia page views on the IPO issue date for each firm from the Wikimedia Foundation.[17] Because page-view data are only available after December 2007, we restrict our traffic analysis to Wikipedia IPOs issued after this month (286 IPOs) and all non-Wikipedia IPOs (644 IPOs) and set

[17] Page-view data back to July 1, 2015 can be accessed from https://tools.wmflabs.org/pageviews. Prior to 2015, data are available via https://dumps.wikimedia.org/other/pagecounts-raw/. Unfortunately, information about the source of Wikipedia traffic (e.g., geography, institutions vs. retail investors) is not provided by Wikipedia.

log_wiki_traffic to zero for non-Wikipedia IPOs. The results in Table 3, Model 4, indicate that Wikipedia article traffic has incremental explanatory power for underpricing.[18]

Due to significant differences between IPOs with and without a Wikipedia article, we also use propensity score matching (Rosenbaum & Rubin, 1983) to construct a matched sample of IPOs. Matches are based on the IPO firm characteristics used in Table 3. For each IPO firm with a Wikipedia article, we use the propensity score to identify the nearest match with replacement among IPO firms without a Wikipedia article.[19] This procedure ensures that matched firms have similar characteristics. Table 4, Panel A, compares the matched samples. We find that the samples are similar except for differences in *tech* ($p < 0.1$).

We report the results of the underpricing regressions for the matched sample in Table 4, Panel B. The results are similar to those reported in Table 3 for the full sample—namely, IPO firms with a Wikipedia article exhibit greater underpricing than their matched counterparts. For example, Model 3 indicates that the presence of a Wikipedia article is associated with first-day returns that are 7.9 percentage points higher.

4.4 | Investor attention

4.4.1 | Wikipedia and investor attention

In this section, we compare Wikipedia to other measures of investor attention used in the literature. Although distinct in its own manner, our approach is inspired by the analysis of the Google Search Volume Index (SVI) as a measure of investor attention reported in Da, Engelberg, & Gao (2011). We include a number of potential proxies for investor attention in this analysis. *Log_offering_size* captures the size of the IPO, where larger IPOs are expected to garner more investor attention. We also consider the ratio of advertising expenditures to sales (*adexptosales*) because firms can draw attention through their marketing efforts. *Log_analyst_preIPO* controls for the number of analysts who issue an earnings forecast for the IPO firm between the S-1 filing and IPO date. Only 48 of 974 IPOs have one or more analyst forecasts during this period.

We follow Da, Engelberg, & Gao (2011) to construct abnormal SVI (*ASVI*), which is the natural logarithm of SVI during the IPO week minus the median SVI during the prior 8 weeks:

$$ASVI_t = \log(SVI_t) - \log(Med(SVI_{t-1}, \ldots, SVI_{t-8})). \quad (1)$$

ASVI captures the attention jump around the IPO event reflected in Google searches. To determine the search term for an IPO company, we start with the company name reported in SDC and then match to search terms based on how investors might search for the company in Google. Notably, to capture the attention of investors rather than consumers, we keep suffixes like "Inc" or "Corp" for companies in retail or service industries. For example, the search term "Tumi Inc" is presumed to be for information about the company, whereas a search for "Tumi" is assumed to come from people who are interested in suitcases or bags. In addition, retail and service companies may have strong seasonality that could obscure the change in investors' search volume around an IPO if we use product name or brand name. We obtain *ASVI* for 912 IPO companies.[20] We also follow Cookson & Niessner (2020) and manually collect StockTwits "cashtag" (i.e., $TICKER) counts on the issue and first trading dates for all IPO firms in our sample.

[18] Results are similar if we only use the 286 IPOs with traffic data.

[19] Matching without replacement leads to significant differences in *overhang*, *log_sales*, and *log_news* between treated and control group, but the regression results are similar.

[20] Fifteen IPOs do not have Google Trends data available, while 47 IPOs have median SVI values equal to zero. *ASVI* cannot be calculated for these IPOs, which reduces the sample size in Table 5, Model 4.

TABLE 4 IPO underpricing (matched sample)

Panel A: Two-sample mean comparison after propensity score matching			
Variable	**Wikipedia**	**Non-Wikipedia**	**Diff *t*-stat**
VC	0.45	0.44	0.39
top_tier	0.94	0.93	0.48
overhang	4.24	4.45	−0.92
pos_EPS	0.48	0.51	−0.86
log_sales	5.41	5.35	0.31
nasdaq15	0.69	0.74	−0.19
tech	0.42	0.36	1.68*
log_age	2.83	2.95	−1.61
log_news	1.30	1.21	0.89

Panel B: OLS regression of underpricing for matched sample				
	(1)	**(2)**	**(3)**	**(4)**
Variables	**Underpricing**	**Underpricing**	**Underpricing**	**Underpricing**
Wikipedia	8.135**	8.302**	7.923**	
	(2.679)	(3.048)	(2.922)	
log_wiki_traffic				2.425***
				(0.377)
VC		8.773**	8.580**	8.158**
		(3.307)	(3.437)	(3.566)
top_tier		8.038	8.268	7.214
		(4.899)	(5.193)	(5.070)
overhang		0.779	0.682	0.625
		(0.627)	(0.626)	(0.542)
pos_EPS		4.438*	4.249	3.748
		(2.427)	(2.493)	(2.475)
log_sales		−0.514	−0.715	−0.944
		(0.664)	(0.772)	(0.705)
nasdaq15		0.541	0.499	0.613
		(0.357)	(0.371)	(0.355)
tech		−1.608	−1.289	−2.159
		(3.793)	(3.728)	(3.476)
log_age		−3.385*	−3.346*	−2.959
		(1.623)	(1.650)	(1.681)
log_news			1.720	0.991
			(1.097)	(0.902)
Constant	12.70***	8.367	7.653	9.240
	(1.250)	(7.099)	(7.594)	(7.363)

(Continues)

TABLE 4 (Continued)

Panel B: OLS regression of underpricing for matched sample				
	(1)	(2)	(3)	(4)
Variables	Underpricing	Underpricing	Underpricing	Underpricing
Observations	660	660	660	572
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.045	0.125	0.129	0.167

Note: This table displays OLS regressions including 330 Wikipedia and 330 matched US IPOs from 2006 to 2016. Matches are based on IPO firm characteristics and the number of news articles about the IPO firm between the S-1 filing and IPO date. Underpricing is the percent change from offer price to the first closing price. The Appendix provides variable definitions. Standard errors in parentheses are robust and adjusted for clustering by Fama–French (1997) 48-industry and year.
***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

In Table 5, Panel A, we report simple correlations between Wikipedia and other variables of interest measured at the IPO level. In general, the correlations between Wikipedia and the other attention measures are low (−0.04 to 0.34). For example, the correlation between Wikipedia and *log_news* is 0.22, while its correlation with *stock_twits* is 0.18. Interestingly, the correlation between Wikipedia and *ASVI* is slightly negative but not statistically significant.

In Table 5, Panel B, we use the Wikipedia indicator as the dependent variable in probit analyses to examine the relation between Wikipedia and other measures of attention. We find that Wikipedia is positively correlated with several of the other measures of investor attention, including offer size and the advertising expenditures–to–sales ratio. Interestingly, Wikipedia is positively correlated with *log_news* but is not significantly correlated with *ASVI* and *stock_twits* in these regressions. Pseudo-R^2 values range from 0.155 to 0.177, which suggests that a significant fraction of the variation in Wikipedia is not explained by other measures of attention. These results support the notion that Wikipedia is not only an indicator of investor attention but that it also captures unique aspects of investor attention compared to measures used in prior studies.

4.4.2 | Offer price revision

Hanley (1993, p. 241) notes that "the greater the participation of institutions, as measured by their post-issue holdings, the higher is the absolute change in offer price." Therefore, if an IPO firm's Wikipedia article aggregates information related to institutional investor attention, we expect it to be associated with the offer price revision process (Benveniste & Spindt, 1989). In Table 6, Model 1, we regress *offer_revision*, the percent change from the midpoint of initial filing range to the final offer price, on *Wikipedia* and control variables. We follow Hanley (1993) and Hanley & Hoberg (2010) and add the following control variables: *offer_width*, which is the difference between the high and low offer prices quoted in the preliminary prospectus divided by the low offer price; *shares_filed*, which is the expected number of shares offered; and *instown_pct_post*, which is the percentage of institutional ownership at the end of first quarter after IPO.

We find that a pre-IPO Wikipedia article is positively related to the magnitude of the price revision. Specifically, IPO offer prices for firms with a Wikipedia article are revised upward 6.38% more, on average, than IPO offer prices for firms without a Wikipedia article. In Model 2, we examine the effect of Wikipedia-specific variables. The variable *wiki_aggregate* is the union of negative, uncertain, and litigious words based on the Loughran & McDonald (2011) dictionary. More negative sentiment is expected to be negatively related to offer price revision because it reflects higher risk. We include *log_wiki_revisions* to control for information updates in an IPO firm's Wikipedia article during the

TABLE 5 Wikipedia and alternative measures of attention

Panel A: Correlations						
Variables	**(1)**	**(2)**	**3**	**4**	**5**	**6**
(1) Wikipedia	1.00					
(2) log_offering_size	0.34***	1.00				
(3) adexptosales	0.18***	0.09***	1.00			
(4) log_analyst_preIPO	0.16***	0.34***	0.10***	1.00		
(5) log_news	0.22***	0.29***	0.17***	0.19***	1.00	
(6) ASVI	−0.04	−0.05	0.00	0.03	−0.04	1.00
(7) stock_twits	0.18***	0.21***	0.02	0.33***	0.17***	0.01

Panel B: Wikipedia and alternative measures of attention				
	(1)	**(2)**	**(3)**	**(4)**
Variables	**Wikipedia**	**Wikipedia**	**Wikipedia**	**Wikipedia**
log_offering_size	0.479***	0.455***	0.413***	0.369***
	(0.115)	(0.113)	(0.110)	(0.111)
adexptosales	4.089***	4.102***	3.696**	3.604***
	(1.465)	(1.473)	(1.458)	(1.374)
log_analyst_preIPO		0.367	0.312	0.244
		(0.375)	(0.412)	(0.417)
log_news			0.184***	0.206***
			(0.065)	(0.071)
ASVI				−0.005
				(0.006)
stock_twits				0.006
				(0.004)
Constant	−2.877***	−2.773***	−2.689***	−2.591***
	(0.488)	(0.480)	(0.474)	(0.422)
Observations	974	974	974	912
Year FE	Yes	Yes	Yes	Yes
Pseudo-R^2	0.155	0.157	0.170	0.177

Note: Panel A presents a correlation matrix between the Wikipedia indicator variable and alternative measures of investor attention. Panel B displays probit regression results with *Wikipedia* as the dependent variable. The Appendix provides variable definitions. Standard errors reported in parentheses are robust and adjusted for clustering by Fama–French (1997) 48-industry and year.

*** and ** indicate significance at the 1% and 5% levels, respectively.

bookbuilding process. For non-Wikipedia firms, *wiki_aggregate* and *log_wiki_revisions* are set equal to zero.[21] We find that *log_wiki_revisions* is positively related to offer price revisions, which suggests that IPO firms with more developed Wikipedia articles experience larger offer price revisions. Presumably, this is due to increased institutional investor attention.

[21] Results are similar if non-Wikipedia IPOs are omitted.

TABLE 6 Offer price revision

	(1)	(2)	(3)	(4)
Variables	**Offer_revision**	**Offer_revision**	**Pos_revision**	**Pos_revision**
Wikipedia	6.369***		0.319**	
	(1.702)		(0.143)	
wiki_aggregate		−0.035		−0.029
		(0.515)		(0.042)
log_wiki_revisions		2.562**		0.156**
		(0.937)		(0.063)
VC	0.838	0.721	0.251***	0.245***
	(2.912)	(3.003)	(0.084)	(0.086)
top_tier	5.423*	5.823**	0.620***	0.632***
	(2.458)	(2.514)	(0.135)	(0.143)
overhang	0.998***	0.938***	0.057***	0.056***
	(0.113)	(0.121)	(0.013)	(0.014)
pos_EPS	4.058*	4.062*	0.241	0.232
	(2.020)	(2.104)	(0.193)	(0.197)
log_sales	0.605	0.630	0.028	0.029
	(0.420)	(0.423)	(0.033)	(0.034)
nasdaq15	0.298	0.272	0.017	0.016
	(0.315)	(0.331)	(0.016)	(0.017)
tech	4.086	4.084	0.134	0.126
	(2.376)	(2.470)	(0.128)	(0.135)
log_age	−3.651***	−3.568***	−0.170***	−0.166***
	(0.610)	(0.672)	(0.057)	(0.056)
log_news	0.239	0.214	0.055	0.045
	(0.446)	(0.426)	(0.047)	(0.048)
offer_width	−0.171	−0.167	−0.004	−0.004
	(0.161)	(0.161)	(0.014)	(0.013)
shares_filed	0.032	0.016	0.002	0.001
	(0.024)	(0.024)	(0.002)	(0.002)
instown_pct_post	0.027	0.026	0.003**	0.003**
	(0.033)	(0.033)	(0.001)	(0.001)
Constant	−10.343*	−10.085*	−1.253***	−1.226***
	(4.723)	(4.737)	(0.379)	(0.372)
Observations	974	974	974	974
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.141	0.141		
Pseudo-R^2			0.102	0.105

(Continues)

TABLE 6 (Continued)

Note: This table reports OLS (Models 1 and 2) and probit (Models 3 and 4) regression results for offer price revision–related variables. *offer_revision* is the percent change from the midpoint of initial filing price range to final offer price. *pos_revision* is an indicator variable that equals 1 if final *offer_revision* is positive, and 0 otherwise. The Appendix provides variable definitions. Standard errors displayed in parentheses are robust and adjusted for clustering within Fama–French (1997) 48-industry and year.

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

We further examine the relation between *Wikipedia* and the direction of offer price revisions. The dependent variable of interest here is *pos_revision*, which is an indicator variable set equal to 1 if *offer_revision* is positive, and 0 otherwise. In Table 6, Model 3, *Wikipedia* is associated with more upward offer price revisions. In Model 4, we report that *log_wiki_revisions* is positively related to upward offer price revisions. In sum, Table 6 provides evidence that the presence of a Wikipedia article helps to explain the magnitude and direction of an IPO firm's offer price revision. The results further suggest that offer price revisions are incomplete as evidenced by the positive and significant effect of the presence of a Wikipedia article on IPO underpricing. This is consistent with Bradley et al. (2003), who find that IPO prices only partially adjust to public information before the IPO date. Thus far, the evidence is consistent with our investor attention hypothesis (H2).[22] Because prior research differs on the long-run impact of investor attention on IPO firms (Da, Engelberg, & Gao, 2011; Liu, Sherman, & Zhang, 2014), the next section further explores this issue.

4.4.3 | The long-run effects of investor attention

To study the long-run impact of investor attention on IPO firms, we follow the approach of Liu, Sherman, & Zhang (2014), with the primary difference being our proxy for investor attention. They use media coverage as their measure of investor attention, which we control for by using *log_news*, thereby allowing us to isolate any incremental effect of the presence of a Wikipedia article. To the extent that investor attention from a firm's Wikipedia article has positive long-run effects, we would expect to observe a positive relation between *Wikipedia* and both post-IPO analyst coverage and institutional ownership.

The results in Table 7 are consistent with the notion that investor attention captured by Wikipedia articles has positive long-run effects for IPO firms. Specifically, we find that Wikipedia firms have greater analyst following and more institutional investors for at least 3 years after the IPO. Consistent with Liu, Sherman, & Zhang (2014), *log_news* is also positively correlated with analyst coverage and institutional ownership. The remaining variables are generally consistent with expectations. For example, sales and underwriter reputation (firm age and NASDAQ listing) are positively (negatively) associated with analyst coverage and institutional ownership.

The evidence to this point indicates that a pre-IPO Wikipedia article is positively correlated with IPO underpricing and long-run institutional attention. We also examine whether there is a price reversal for IPOs that receive investor attention, as reported by Da, Engelberg, & Gao (2011). To conserve space, we report the results in Table A.2 of the Appendix in the Supporting Information. In this analysis, the dependent variable is the cumulative IPO return from weeks 5 to 52 after the IPO event. The remaining variables follow Da, Engelberg, & Gao (2011). The evidence indicates that investor attention from Wikipedia is not correlated with post-IPO returns. This differs from Da, Engelberg, & Gao (2011), who report price reversals for IPOs with both high investor attention and high first-day returns. These results, combined with the results in Table 7, support prior research, which posits that pre-IPO investor attention has long-run benefits for IPO firms.

[22] Bradley, Cooney, Jordan, and Singh (2004) relate low offer price precision to information asymmetry. However, Table A.1 of the Appendix in the Supporting Information shows that *Wikipedia* is not a significant determinant of integer offer prices.

TABLE 7 Long-run attention

	Analyst following			Number of institutional investors		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Wikipedia	1.307**	1.850***	1.951***	8.886***	20.263*	13.930**
	(0.432)	(0.502)	(0.541)	(2.507)	(9.311)	(5.521)
log_news	0.466*	0.615**	0.981***	10.153***	17.567***	16.167***
	(0.231)	(0.230)	(0.223)	(1.846)	(3.691)	(2.296)
VC	−0.414	−0.229	0.109	0.423	4.114	1.606
	(0.422)	(0.628)	(0.638)	(3.346)	(7.844)	(5.585)
top_tier	1.349***	1.824***	2.016***	20.560***	34.523***	30.105***
	(0.362)	(0.451)	(0.513)	(2.893)	(3.567)	(4.584)
overhang	0.211	0.213	0.301	0.227	4.167***	2.460**
	(0.117)	(0.159)	(0.172)	(0.647)	(0.697)	(0.889)
pos_EPS	0.187	0.493	0.292	7.502**	8.497	10.933*
	(0.318)	(0.463)	(0.388)	(2.898)	(7.388)	(5.255)
log_sales	0.317***	0.310***	0.320***	2.631**	5.834*	4.271**
	(0.079)	(0.093)	(0.099)	(0.945)	(2.539)	(1.568)
nasdaq15	−0.052	−0.089**	−0.116**	−1.432**	−1.953	−2.008*
	(0.042)	(0.032)	(0.047)	(0.472)	(1.483)	(0.998)
tech	0.418	0.560	0.599	−1.182	−10.164	−5.279
	(0.427)	(0.603)	(0.561)	(2.818)	(7.694)	(5.084)
log_age	−0.602**	−0.851*	−0.764*	0.149	−0.780	−0.897
	(0.267)	(0.431)	(0.413)	(2.316)	(5.139)	(2.808)
Constant	3.158**	3.823**	3.188**	16.145**	13.162	18.249**
	(1.006)	(1.553)	(1.425)	(5.660)	(16.823)	(8.039)
N	967	860	740	820	732	630
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.304	0.282	0.290	0.279	0.303	0.303

Note: This table reports OLS results of the relation between Wikipedia and short-/long-run attention. Dependent variables of *analyst following* are the number of analyst estimates 1, 2, and 3 years after IPO event. Dependent variables of institutional investors are the *number of institutional investors* at the first, second, and third fiscal year–ends after IPO. Year- and Fama–French 48-industry-fixed effects are included. Reported in parentheses are standard errors, which are adjusted for clustering by year- and Fama–French (1997) 48-industry classification.

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

5 | ENDOGENEITY CONCERNS

Given the different characteristics of the Wikipedia and non-Wikipedia samples, the existence of a Wikipedia article may not be random. For example, IPO firms may choose to create and edit their own Wikipedia article, and this behavior may be correlated with IPO underpricing. Moreover, given the information aggregation function of Wikipedia, if

there exists another information channel that affects both the likelihood of an IPO firm having a Wikipedia article and IPO underpricing, then we cannot attribute the effect to Wikipedia.

Even though we cannot completely rule out these scenarios, we attempt to address endogeneity concerns using the average number of daily new English Wikipedia articles created during each month (*log_aggWiki_new*) as an instrument for the *Wikipedia* indicator variable. As company-related articles represent a small component of newly created Wikipedia articles, new article creation exhibits substantial time series variation, and Wikipedia editor capacity declines slightly over the sample period, we expect that fewer company-specific articles will be created when special events happen that grab the Wikipedia community's attention and draw down on said capacity.[23] However, the instrumental variable is unlikely to be correlated with IPO underpricing because the English Wikipedia encompasses 13 large article content categories with a daily average of 1105 (range of 570—2105 in our sample) new articles created each day.

The results in Table 8 confirm this expectation. In the first stage of the two-stage least squares analysis (Models 1 and 3), *log_aggWiki_new* is negatively correlated with both *Wikipedia* and *log_wiki_traffic*. The *F*-statistics for the Cragg–Donald weak instrument identification test are 42.942 and 146.451, respectively, which are well above the Stock et al. (2002) critical value of 8.96 at a 5% significance level. In the second stage (Models 2 and 4), the instrumented variables are both highly significant. Overall, the results support the hypothesis that Wikipedia captures investor attention of IPO firms that results in larger first-day returns.

6 | OTHER MEASURES OF INVESTOR ATTENTION

6.1 | Google search volume

Da, Engelberg, & Gao (2011) find that the Google SVI has low correlation with other attention measures including abnormal returns, turnover, and news coverage. In addition, they show that SVI mainly captures retail investor attention. Given that Wikipedia articles rank high in Google search results, one concern is that Wikipedia traffic is simply a derivative of Google search activity. As noted earlier, the correlation between *ASVI* and *Wikipedia* (−0.04) is not significant. There are several possible reasons for this. For instance, one might interpret *ASVI* as a measure of attention change and *Wikipedia* as a measure of the attention level. Second, many company searches in Google are for reasons unrelated to investing (e.g., product information). Third, Google and Wikipedia may represent attention from different categories of market participants—noise traders and sophisticated investors (Hervé et al., 2019). Finally, Wikipedia began receiving less referral traffic from Google search results following Wikipedia's adoption of hypertext transfer protocol secure (HTTPS) in 2011. Importantly, all of our multivariate specifications incorporate year fixed effects.

Table 9, Model 1, reports regression results after we add *ASVI* to the baseline model (Table 3, Model 3). The estimated coefficient and significance of *Wikipedia* are similar to those in the baseline regression.[24] However, *ASVI* does not predict underpricing in our model. These results suggest that while Wikipedia and Google search both allow investors to "pull" information about specific companies, the impact of Wikipedia on IPO underpricing is meaningfully incremental beyond search engine activity and remains economically important.

[23] See https://en.wikipedia.org/wiki/Wikipedia:Contents/Categories for more information on Wikipedia article categories, https://stats.wikimedia.org/EN/TablesArticlesNewPerDay.htm for articles created per day, and https://stats.wikimedia.org/#/en.wikipedia.org/contributing/editors/normal|line|2006-02-01~2017-02-01|activity_level~25..99-edits*100..-edits|monthly for editor capacity.

[24] Results in Table 9 are similar if we use *log_wiki_traffic* in place of *Wikipedia*.

TABLE 8 Endogeneity

	Underpricing		Underpricing	
	First stage	Second stage	First stage	Second stage
Variables	(1)	(2)	(1)	(2)
log_aggWiki_new	−0.326***		−3.130***	
	(0.050)		(0.259)	
Wikipedia		25.787***		
		(9.015)		
log_wiki_traffic				2.685***
				(0.878)
VC	0.011	8.362***	0.179	8.163***
	(0.034)	(1.981)	(0.174)	(1.855)
top_tier	0.121***	3.824	0.558***	5.437**
	(0.039)	(2.554)	(0.202)	(2.208)
overhang	0.021***	1.070**	0.184***	1.128***
	(0.006)	(0.456)	(0.033)	(0.415)
pos_EPS	−0.000	3.591*	0.146	3.196*
	(0.033)	(1.947)	(0.171)	(1.808)
log_sales	0.027***	−0.762*	0.134***	−0.436
	(0.006)	(0.415)	(0.030)	(0.337)
nasdaq15	−0.009**	0.584**	−0.063***	0.510**
	(0.004)	(0.257)	(0.022)	(0.236)
tech	0.092***	0.323	0.515***	1.308
	(0.032)	(1.948)	(0.165)	(1.752)
log_age	0.065***	−2.390*	0.184*	−1.204
	(0.020)	(1.326)	(0.104)	(1.122)
log_news	0.061***	0.669	0.442***	1.062
	(0.013)	(0.865)	(0.067)	(0.753)
Constant	2.073***	1.817	20.446***	0.371
	(0.352)	(4.217)	(1.828)	(3.776)
Cragg–Donald *F*-statistic	42.942		146.451	
Observations	974	974	930	930
R^2	0.215	0.105	0.302	0.124

Note: This table displays results of two-stage least square (2SLS) regression of IPO underpricing with *Wikipedia* and *log_wiki_traffic* as the endogenous variables. *Wikipedia* and *log_wiki_traffic* are instrumented with *log_aggWiki_new*, which is the natural log of the daily average of aggregate new articles created on English Wikipedia for each month. Underpricing is the percent change from offer price to the first closing price. The Appendix provides variable definitions.
***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

6.2 | Social media activity

Social media is an increasingly important source of information for investors (Blankespoor, Miller, & White, 2014; Zhou, Lei, Wang, Fan, & Wang, 2015). Like traditional media, social media can "push" firm-specific information to investors. Any association between Wikipedia and IPO underpricing may therefore be attributed to an increase in

TABLE 9 Robustness checks

Variables	(1) Underpricing	(2) Underpricing	(3) Vol_retail_pct
Wikipedia	4.600*	4.376*	0.210
	(2.315)	(2.287)	(0.223)
ASVI	0.093		
	(0.065)		
stock_twits		0.063**	
		(0.026)	
VC	8.991***	7.878**	−1.166**
	(2.540)	(2.557)	(0.418)
top_tier	7.472***	6.633***	0.260***
	(1.314)	(1.395)	(0.053)
overhang	1.438***	1.313***	−0.367
	(0.295)	(0.277)	(0.282)
pos_EPS	3.992	4.233	−0.090
	(2.468)	(2.525)	(0.079)
log_sales	−0.181	−0.267	0.006
	(0.619)	(0.598)	(0.027)
nasdaq15	0.492	0.430	0.255*
	(0.332)	(0.315)	(0.135)
tech	2.572	2.644	0.105
	(2.845)	(2.771)	(0.092)
log_age	−1.162	−1.129	0.255*
	(1.456)	(1.412)	(0.135)
log_news	2.899***	2.040**	−0.158
	(0.754)	(0.787)	(0.263)
Constant	−4.312	−1.789	3.014***
	(3.679)	(3.658)	(0.194)
Observations	912	974	974
Year FE	Yes	Yes	Yes
Adjusted/Pseudo-R^2	0.127	0.129	0.356

Note: This table reports regression results for various robustness checks. *ASVI* is abnormal Google search volume for the IPO company. *stock_twits* is the number of ticker references on Twitter on the IPO date according to Cookson & Niessner (2020). *vol_retail_pct* is the percentage of retail trading volume on the IPO date using the trade-level classification algorithm of Boehmer, Jones, Zhang, & Zhang (2021). The Appendix provides variable definitions. Standard errors in parentheses are robust and adjusted for clustering by Fama–French (1997) 48-industry and year.
***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

social media activity about an IPO firm. To examine this possibility, we follow Cookson & Niessner (2020) and manually collect StockTwits "cashtag" (i.e., $TICKER) counts on the issue and first trading dates for all IPO firms in our sample. In Model 2 of Table 9, we incorporate *stock_twits* into our main underpricing specification and find that it enters into the regression positively and is significant at the 5% level. *Wikipedia* remains positively correlated with underpricing in this augmented model.

6.3 | Retail versus institutional investor attention

Da, Engelberg, & Gao (2011) and Liu, Sherman, & Zhang (2014) differ in their attribution of IPO effects from investor attention to retail and institutional investors, respectively. Therefore, identifying IPO trading volume by investor type helps distinguish between these two lines of research. In Model 3 of Table 9, we examine the percentage of first-day trading volume initiated by retail investors (*vol_retail_pct*) estimated using the trade-level algorithm of Boehmer, Jones, Zhang, & Zhang (2021). When we separate trading volume by investor type, we find that the existence of a Wikipedia article prior to a firm's IPO is not associated with the proportion of trading attributed to retail investors. This result is consistent with the proposition that the increase in investor attention from a Wikipedia article is not driven exclusively by retail investors.[25]

7 | CONCLUSION

We investigate the impact of Wikipedia on IPOs. A firm's Wikipedia article has far more page visits on its IPO issuance date and is a potentially valuable external source of information beyond the carefully crafted S-1 regulatory filings that accompany IPOs. On the one hand, a Wikipedia article may reduce information disparities among IPO participants, which allows for more precise offer prices (Bradley, Cooney, Jordan, & Singh, 2004) and mitigates information effects that contribute to underpricing (Ljungqvist, 2007). On the other hand, Wikipedia has the potential to increase investor attention of IPO firms that prior research has associated with larger first-day returns (Da, Engelberg, & Gao, 2011).

We find that firms that have a Wikipedia article when they go public experience significantly higher underpricing than firms without a Wikipedia article. This effect is greater when the firm's Wikipedia article receives more visits. We draw on prior research on investor attention to explain the positive link between a Wikipedia article and underpricing. Consistent with Liu, Sherman, & Zhang (2014), we find that IPO firms with a Wikipedia article benefit from greater analyst following and attract more institutional investors for up to 3 years following the offering. Importantly, these results withstand a battery of robustness checks.[26] Overall, our results are consistent with the Merton (1987) investor attention model that predicts that higher investor attention shifts the demand curve and has positive long-run effects.

Jimmy Wales, his staff, and the contributor community of Wikipedia have fundamentally changed the information environment of the world. Future research may examine other effects of Wikipedia articles on capital markets, firm activities, and general economic activity. Wikipedia continues to provide detailed access to page histories, traffic, and other useful data for research applications. As Wikipedia continues to grow (adding 4000+ new English-language articles per week), its influence will continue to expand.

ACKNOWLEDGMENTS

The authors thank Richard Warr (editor), two anonymous reviewers, Jennifer Conrad, Paolo Fulghieri, Mustafa Gültekin, Alfred Liu, Zhe Shen, Weineng Xu, Xiaoyun Yu, and seminar participants at the 2017 Southwestern Finance Association Conference (Little Rock), 2017 China International Conference in Finance (Hangzhou, China), 2018 Financial Management Association Conference (San Diego), University of North Carolina at Chapel Hill, SUNY-Albany, and Rensselaer Polytechnic Institute for helpful comments and discussion. Research funding was provided by the Lindmor Professorship (Boulton) and computing resources by the Donald Shohfi Financial Research Fund. Early versions of this study were circulated under the title "The Information Environment of the Firm and IPO Underpricing." Any remaining errors or omissions remain the responsibility of the authors.

[25] We also examine the relation between *Wikipedia* and Bloomberg's "News Heat—Daily Max Readership" measure of institutional investor attention (Ben-Rephael, Da, & Israelsen, 2017). The correlation between these two variables (−0.016) is statistically insignificant.

[26] We conduct additional robustness checks in the Appendix in the Supporting Information.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance*, *59*, 1259–1294.

Bajo, E., & Raimondo, C. (2017). Media sentiment and IPO underpricing. *Journal of Corporate Finance*, *46*, 139–153.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*, 785–818.

Baron, D. P. (1982). A model of the demand for investment banking advising and distribution services for new issues. *Journal of Finance*, *37*, 955–976.

Benveniste, L. M., & Spindt, P. A. (1989). How investment bankers determine the offer price and allocation of new issues. *Journal of Financial Economics*, *24*, 343–361.

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies*, *30*, 3009–3047.

Bhattacharya, U., Galpin, N., Ray, R., & Yu, X. (2009). The role of the media in the Internet IPO bubble. *Journal of Financial and Quantitative Analysis*, *44*, 657–682.

Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter™. *Accounting Review*, *89*, 79–112.

Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2021). Tracking retail investor activity. *Journal of Finance*. Forthcoming.

Bradley, D. J., Cooney, J. W., Jordan, B. D., & Singh, A. K. (2004). Negotiation and the IPO offer price: A comparison of integer vs. non-integer IPOs. *Journal of Financial and Quantitative Analysis*, *39*, 517–540.

Bradley, D. J., Jordan, B. D., & Ritter, J. R. (2003). The quiet period goes out with a bang. *Journal of Finance*, *58*, 1–36.

Bradshaw, T. (2008, December 1). Companies woo investors via social websites. *Financial Times*, p. 18.

Bragues, G. (2007). Wiki-philosophizing in a marketplace of ideas: Evaluating Wikipedia's entries on seven great minds. Working paper, University of Guelph-Humber.

Brown, A. R. (2011). Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. *PS: Political Science & Politics*, *44*, 339–343.

Brown, S., & Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, *12*, 443–477.

Bushee, B. J., Cedergren, M., & Michels, J. (2020). Does the media help or hurt retail investors during the IPO quiet period? *Journal of Accounting and Economics*, 69. https://doi.org/10.1016/j.jacceco.2019.101261

Chen, H., De, P., Hu, Y., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*, 1367–1403.

Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *Annals of Pharmacotherapy*, *42*, 1814–1821.

Comprend. (2015). *Comprend's Capital Market Report 2015*. Stockholm: Comprend.

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *Journal of Finance*, *75*, 173–228.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*, 1461–1499.

Devgan, L., Powe, N., Blakey, B., & Makary, M. (2007). Wiki-surgery? Internal validity of Wikipedia as a medical and surgical reference. *Journal of the American College of Surgeons*, *205*, S76–S77.

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance*, *66*, 67–97.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*, 153–193.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance*, *64*, 2023–2052.

Frederickson, J. R., & Zolotoy, L. (2016). Competing earnings announcements: Which announcement do investors process first? *Accounting Review*, *91*, 441–462.

Gao, X., & Ritter, J. R. (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics*, *97*, 33–52.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*, 900–901.

Greenstein, S., Gu, Y., & Zhu, F. (2016). Ideological segregation among online collaborators: Evidence from Wikipedians. National Bureau of Economic Research Working Paper w22744, Cambridge, MA.

Greenstein, S., & Zhu, F. (2012). Collective intelligence and neutral point of view: The case of Wikipedia. National Bureau of Economic Research Working Paper w18167, Cambridge, MA.

Hanley, K. W. (1993). The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics*, *34*, 231–250.

Hanley, K. W., & Hoberg, G. (2010). The information content of IPO prospectuses. *Review of Financial Studies*, *23*, 2821–2864.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, *46*, 931–959.

Hervé, F., Zouaoui, M., & Belvaux, B. (2019). Noise traders and smart money: Evidence from online searches. *Economic Modelling*, *83*, 141–149.
Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *Journal of Finance*, *75*, 1779–1831.
Hirshleifer, D., Hou, K., Teoh, S. H., & Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*, 297–331.
Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*, 337–386.
Holman Rector, L. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, *36*, 7–22.
Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, *54*, 2143–2184.
Hong, H., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, *83*, 367–396.
Ibbotson, R. (1975). Price performance of common stock new issues. *Journal of Financial Economics*, *2*, 235–272.
Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization*, *176*, 63–90.
Jog, V., & McConomy, B. J. (2003). Voluntary disclosure of management earnings forecasts in IPO prospectuses. *Journal of Business Finance & Accounting*, *30*, 125–168.
Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. *Proceedings of the 25th International Conference on World Wide Web*. Geneva: International World Wide Web Conferences Steering Committee.
Leone, A. J., Rock, S., & Willenborg, M. (2007). Disclosure of intended use of proceeds and underpricing in initial public offerings. *Journal of Accounting Research*, *45*, 111–153.
Liu, L. X., Sherman, A. E., & Zhang, Y. (2014). The long-run role of the media: Evidence from initial public offerings. *Management Science*, *60*, 1945–1964.
Ljungqvist, A. (2007). IPO underpricing. In B. E. Eckbo (Ed.), *Handbook of empirical corporate finance* (pp. 375–422). Amsterdam: North Holland.
Logue, D. (1973). On the pricing of unseasoned equity issues: 1965–1969. *Journal of Financial and Quantitative Analysis*, *8*, 91–103.
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, *66*, 35–65.
Loughran, T., & Ritter, J. (2004). Why has IPO underpricing changed over time? *Financial Management*, *22*, 5–37.
Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance*, *42*, 483–510.
Odean, T. (1999). Do investors trade too much? *American Economic Review*, *89*, 1279–1298.
Ritter, J. R. (1987). The costs of going public. *Journal of Financial Economics*, *19*, 269–281.
Rock, K. (1986). Why new issues are underpriced. *Journal of Financial Economics*, *15*, 187–212.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.
Saxton, G. D., & Anker, A. E. (2013). The aggregate effects of decentralized knowledge production: Financial bloggers and information asymmetries in the stock market. *Journal of Communication*, *63*, 1054–1069.
Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*, 518–529.
Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, *62*, 1139–1168.
Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, *5*, 179–194.
Welch, I. (1989). Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *Journal of Finance*, *44*, 421–449.
Wu, W., Zhang, X., & Zheng, R. (2014). Crowd governance: The monitoring role of Wikipedia in the financial market. Working paper, Hong Kong University of Science and Technology.
Xu, S. X., & Zhang, X. (2013). Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly*, *37*, 1043–1068.
You, J., Coakley, J., Firth, M., Fuertes, A., & Shen, Z. (2018). *Driving the presence of investor sentiment: The role of media tone in IPOs*. Working paper, Xiamen University.
Zhou, M., Lei, L., Wang, J., Fan, W., & Wang, A. G. (2015). Social media adoption and corporate disclosure. *Journal of Information Systems*, *29*, 23–50.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Boulton T, Francis BB, Shohfi T, Xin D. Investor awareness or information asymmetry? Wikipedia and IPO underpricing. *Financial Review* 2021;*56:*535–561.
https://doi.org/10.1111/fire.12276

APPENDIX

VARIABLE DEFINITION

Variables	Definition
underpricing	Percent change from offer price to the first closing price
offer_revision	Percent change of offer price from the midpoint of initial offer price range*100
wikipedia	Indicator variable equal to 1 if the firm had a Wikipedia article when it went public
wiki_traffic	Number of Wikipedia article page views of IPO firms on the first trading date
adexptosales	Advertising expense divided by sales in the previous fiscal year
age	The number of years between IPO firm's founding year and IPO year
aggWiki_new	Daily average number of new English Wikipedia articles for each month
analyst_preIPO	Number of analysts who have issued an earnings forecast during the bookbuilding period
assets	Total assets (MM USD) of IPO firms in the latest fiscal year–end prior to IPO date
ASVI	Log of SVI (weekly Google Trends search index of the IPO company name) during the week minus the log of median SVI during the 8 weeks prior to the IPO date
instown_pct_post	Percentage of shares held by institutional investors at the end of first quarter after IPO
nasdaq15	Buy-and-hold return for the NASDAQ composite index over the 15 days prior to IPO
news	Number of news mentions of the IPO firm during bookbuilding with firm relevancy greater than 75 based on Ravenpack's proprietary algorithm
offer_width	The difference between the high and low offer prices quoted in the preliminary prospectus divided by the low offer price
offering_size	Number of shares offered, in millions
overhang	Retained shares divided by total number of shares offered in the IPO
pctnum_s1 (_wiki)	The total count of numbers in the IPO S-1 filing (Wikipedia article) divided by the total count of numbers and words
pos_EPS	Indicator variable equal to 1 if the IPO firm has positive earnings per share during the last 12-month period prior to IPO date
pos_revision	Indicator variable equal to 1 if *offer_revision* is positive
proceeds	Total amount (MM USD) of offering proceeds obtained by issuer
sales	Total amount of sales (MM USD) for the most recent 12 months prior to IPO date
shares_filed	The expected number of shares offered, in millions
stock_twits	The number of StockTwits mentions of the firm ticker (i.e., $TICKER) on the IPO issue and first trading dates

Variables	Definition
tech	Indicator variable equal to 1 if the main business of an IPO firm is classified in SDC as high tech
top_tier	Indicator variable equal to 1 if the lead underwriter(s) has a top ranking according to Loughran and Ritter (2004)
VC	Indicator variable equal to 1 if the IPO firm is backed by venture capital
vol_retail_pct	The proportion of retail trading volume estimated at the individual trade level using the Boehmer, Jones, Zhang, and Zhang (2021) methodology
wiki_aggregate	Percentage of words in an IPO firm's Wikipedia article using the union of negative, uncertain, and litigious words in the Loughran and McDonald (2011) dictionary
wiki_references	Number of references in a Wikipedia article
wiki_revisions	Total number of Wikipedia article revisions during the bookbuilding period