Journalists and the Stock Market *

Casey Dougal, Joseph Engelberg, Diego García, and Christopher A. Parsons†

February 3, 2011

Abstract

We find that a small set of financial columnists has a causal effect on short-term aggregate stock market prices. For some journalists ("bulls") the market reaction is consistently positive, whereas for others ("bears") it is negative. Because bulls and bears are rotated exogenously in our setting, we can make causal inferences about the media's impact on aggregate market returns. Journalist effects are much stronger after extreme returns, suggesting that *amplification* or *attenuation* of existing sentiment is the mechanism underlying the financial media's influence.

Keywords: *Media Bias, Market Efficiency, Financial Journalism.*

JEL: H53, I38, J31, J33.

---

*We thank Stephan Siegel, Paul Tetlock, Sheridan Titman, and seminar participants at Boston College, INSEAD, University of Arizona, UNC, and the University of Toronto (Economics) for helpful feedback and suggestions.The usual disclaimer applies.

†All authors are from the Kenan-Flagler Business School, University of North Carolina. Casey Dougal, (Email) casey_dougal@kenan-flagler.unc.edu, (Tel) 208-284-3117; Joseph Engelberg, (Email) joseph_engelberg@kenan-flagler.unc.edu, (Tel) 919-962-6889; Diego García, (Email) diego_garcia@unc.edu, (Tel) 919-962-8404; Christopher Parsons, (Email) chris_parsons@kenan-flagler.unc.edu, (Tel) 919-962-4132.

1 Introduction

The media is often modeled as a faceless institution, but its main output – news content – is generated by specific people. This is important because unlike, say, making tires or processing paper, writing is a fiercely individualistic craft that allows an author's style, persuasion, views, or bias to be injected into the finished product. In this paper, we present direct evidence that the writing of specific journalists has a *casual* effect on aggregate market outcomes.

This is surprising because, at any point in time, individual columnists are unlikely to possess information advantages relative to the market as a whole, let alone consistently over a period of several years. Thus, any persistent return predictability related to specific authors must arise from their "sentiment" or spin of public information. From 1970 to 2007, we find that the short-term returns on the Dow Jones Industrial Average (DJIA) can be predicted knowing only the author of a widely read market summary article, the *Wall Street Journal's* "Abreast of the Market" (AOTM) column.

Ordinarily we would be concerned about the endogenous nature of news coverage in an article summarizing market events. As Tetlock (2007, p. 1139) notes, "It is unclear whether the financial news media induces, amplifies, or simply reflects investors' interpretations of stock market performance." Making a distinction between a reflective and a causal role for financial media thus requires exogenous variation in news content, or reporting uncorrelated with underlying events.

Our setting is particularly useful in this regard. Over the nearly four decades we study, columnists rotate frequently – three different journalists write the AOTM column in the typical month – and often according to regular schedules. Moreover, journalists differ markedly in their writing styles such as sentence structure, complexity, article length, and even pessimism or optimism about market conditions. Our empirical strategy exploits rotation and these cross-journalist differences to identify a causal effect on investor behavior.

In our main tests, the dependent variable in a linear regression is the daily excess return on the DJIA Index. The control variables include several lags of returns, day of the week dummies, time effects, lagged volume, and lagged volatility. Our primary interest is the twenty-five vectors of journalist indicators, one for each financial columnist writing for the *WSJ* during our sample period. These fixed effects are statistical stand-ins for both observable and unobservable content differences that persist between financial columnists.

We find that journalist fixed effects are significant predictors of future DJIA returns. Specifying only the name of the financial columnist writing for the *WSJ* on a given day increases predictive power of the regression by 40-50% relative to the other control variables. In linear restriction tests, journalists are jointly significant at predicting returns on both the day the journalist's article is published, as well as the day immediately following.

Because we are examining signed rather than absolute market returns, we can interpret the columnist coefficients as capturing their average bullishness or bearishness. When a bullish (bearish) columnist writes, the market inches upward (downward) a few basis points. However, this specification masks what is potentially a more important question: can journalists exert a unilateral influence on investor behavior, or must certain conditions be met for them to have an effect?

Answering this speaks directly to the mechanism underlying any media effects we observe. If the role of financial journalists is ultimately to provide color and interpretation to market events, we would expect for their effects to be highest around news events and volatile returns. On the other hand, evidence on limited attention might suggest that investors are least persuadable during these busy periods, and therefore, that financial journalism might matter in "quieter" times.[1]

To address this issue, we augment our benchmark specification by interacting each of the columnist fixed effects with lagged stock returns. Positive interaction coefficients identify journalists that contribute to positive serial correlation, effectively amplifying whatever investor sentiment may exist. A negative coefficient suggests the opposite – a contrarian writer who tempers enthusiasm, dampening the market response. We find that richer specifications including these interactions increase the explanatory power of the regression by yet another 50%, and yield significant coefficients for a full eighty percent of the journalists in our sample. Together, a coherent story emerges: journalists influence market returns, but primarily via *amplification* or *attenuation.* In other words, journalists matter in financial markets, but not in isolation from underlying events.

The biggest challenge to a causal interpretation is that the selection of journalists may not be orthogonal to future market returns. For example, one might worry about an editor assigning a certain writer after steep declines – unless we can perfectly control for any continuation or reversal

[1]See DellaVigna and Pollet (2009) and Hirshleifer, Lim, and Teoh (2009).

effects,[2] future returns might be spuriously correlated with the presence of certain journalists. Fortunately, the fact that there is considerable predictability in columnist scheduling allows us to deal directly with this possibility.

Rather than explaining stock returns using which journalist's article was published that day, we instrument for the author using his past scheduling information. For example, a common arrangement is for one journalist to write on Monday to Thursday for a few weeks, and for a different one to spell him on Fridays. These and other scheduling patterns make past writing activity a valid predictor for future activity, but importantly, not in a way that can be plausibly related to future returns. Although somewhat weaker than the benchmark regressions, the instrumental variable specifications yield jointly significant coefficients for both the journalist indicators ($p = 0.02$) and their interactions with returns ($p < 0.001$). Because we are using only information exogenous to returns to predict journalist arrival, this specification represents perhaps the strongest evidence for a causal relation.

We conclude our analysis with a number of robustness exercises. While our main regressions use DJIA close-to-close prices, the results are similar if we use DJIA open-close prices, or if we analyze other series such as the CRSP value-weighted or S&P 500 Index. Our results do not appear to be driven by outliers, for either returns or journalists. If we use GARCH-adjusted or winsorized returns as our dependent variable, the results hold. Similarly, if we include only the ten most frequently credited authors, the relations we document in the main analysis remain.

While interpreting our results, it is important to note that we do not attempt to relate direct, daily content measures – e.g., word counts, sentence length, article tone – to daily stock returns. There are two reasons why. First, because article content largely reflects contemporaneous market conditions, one worries about reverse causation, as Tetlock (2007) discusses. This is not likely to be a concern when relating average returns to the day a particular journalist writes. Second, and perhaps more importantly from a methodological perspective, our measure of news content is not limited to those generated by computerized algorithms. Because we rely exclusively on columnist rotation to identify content changes, our method will pick up *any* persistent variation, whether or not we can measure it otherwise. In other words, our study is agnostic about whether the usual

[2] Note that we already include several days of lagged returns, but the underlying relationship may be more complicated than this linear specification.

content proxies (e.g., word counts) generate the return effects we observe, or whether more subtle stylistic elements are responsible.

A number of studies have documented the media's ability to shift public opinion, particularly with regard to voting patterns (DellaVigna and Kaplan (2007), and Gerber, Karlan, and Bergan (2009)). What makes the present results so surprising is the strong theoretical assumption that the media should not, apart from information effects, be able to influence prices, certainly at the aggregate level.[3] Whereas there are models to explain why consumers might be susceptible to biased reporting – and by extension, why media outlets may then have an incentive to misreport to them (Gentzkow and Shapiro (2006)) – prices of financial securities should not reflect bias due to informed traders. That they do is consistent with market frictions limiting such arbitrage. However, the impact of journalists on returns is strongest at the very end of our sample (2000 to 2007), when transaction costs were at historic lows and opportunistic hedge funds were most active.

Our study contributes to a growing literature exploring the connection between the media and the stock market. Tetlock (2007) is seminal in this regard, showing that the percentage of negative words in the AOTM predicts next-day market returns. One of our key aims is to shed light on the underlying mechanism – i.e., to distinguish whether the media reflects existing investor sentiment or causes it to change. Because journalist arrival is random, our experimental design allows us to claim a cause-and-effect relationship between the print media and financial market outcomes. While a number of similarly motivated studies present evidence that the media can influence the prices and trading of individual securities (Huberman and Regev (2001), Reuter and Zitzewitz (2006), Engelberg and Parsons (2010)), causal evidence at the market level is absent. This distinction is important, because although journalists could conceivably be informed about a particular stock, this is nearly impossible at the aggregate level.

Finally, our paper adds to a growing literature that focuses on the effects of individuals on economic outcomes. For example, Bertrand and Schoar (2003) demonstrate that specific managers influence firms' investment and capital structure decisions, while Chevalier and Ellison (1999) show that personal fund manager characteristics are related to investment choices and fund performance.

[3] An information story would require a few financial columnists to have persistent information advantages over the entire market, a claim that seems implausible in the short term, even more so over many years. Second, recalling that columnists are affiliated with return patterns of a particular sign, these information advantages would need to be both journalist- and sign-specific. For example, columnist John Smith would need to consistently receive private, *positive* signals about future returns.

Like these studies, we show that individualism matters – not only for the expression of media content, but also for aggregate economic outcomes based on it.

The paper is organized as follows. In Section 2, we describe the data and define our key variables. Section 3 presents evidence that specific journalists influence the aggregate market. Within this section, we also characterize whether the effects of financial journalism vary with market conditions. We deal with the potential endogeneity of journalist scheduling in Section 4, presenting the results when we instrument for journalist arrival using past scheduling information. Section 5 concludes.

2 Data

2.1 Market returns and news articles

Two main data sources are used in this paper: the "Abreast of the Market" (AOTM) column from the *Wall Street Journal*, and the Dow Jones Industrial Average (DJIA) price and dividend series. Our sample period spans January 1, 1970 to December 31, 2007. Data further back is available for both sources, but it was not until approximately 1970 that the AOTM column was consistently published with the accompanying author's name.

Our main dependent variable is the excess daily return on the DJIA Index.[4] From Yahoo! Finance, we extract a daily series of closing prices for the DJIA and then we add the price-weighted dividend yield for each of the index's components because the DJIA is a price-weighted index.[5] Defining r_t as the DJIA excess return and p_t as the level of the DJIA index at the close of day t, we have

$$r_{t+1} = \frac{p_{t+1} - p_t}{p_t} - r_{f,t+1} + dp_{t+1} \tag{1}$$

where $r_{f,t+1}$ is the one-month Treasury bill rate obtained from the Center for Research in Securities Prices (CRSP), and dp_{t+1} is the price-weighted average dividend yield for the stocks in the DJIA index defined as

$$dp_{t+1} = \frac{\sum_{i \in \mathrm{DJIA}} d_{i,t+1}}{\sum_{i \in \mathrm{DJIA}} p_{i,t}}. \tag{2}$$

[4]We use DJIA returns as our dependent variable following Tetlock (2007), who argues that the AOTM column tends to disproportionately cover the blue-chip stocks of the DJIA. However, our main results are nearly identical when we use other aggregate return series, e.g., the S&P 500 Index or the CRSP Value-Weighted Index (see Tables 8 and 9).

[5]Changes in the level of the DJIA ignore distributions to shareholders. See Sialm and Shoven (2000).

One-day lagged prices are used in the calculation of the DJIA aggregate dividend yield. This is to avoid the price adjustments that occur following a dividend issue or stock split. Over our sample period, the total excess return of the DJIA averaged 2.6 basis points, equating to an annualized excess return of 6.5 percent. Daily excess return volatility is approximately 101 basis points, implying an annualized Sharpe ratio of 0.41, nearly identical to that found in other diversified return indices.[6]

Additionally, in the regressions that follow other variables are included to control for known sources of return predictability such as day-of-the-week or liquidity effects and microstructure effects such as bid-ask bounce or non-synchronous trading. We construct the *Controls* vector which includes five lags of detrended daily log volume from the New York Stock Exchange (NYSE) obtained from CRSP, five lags of detrended squared DIJA residuals which proxy for volatility, day-of-the-week dummies, a dummy variable for the month of January, and year fixed-effects. Both log volume and squared DIJA residuals are detrended by subtracting their past 60-day moving average. DIJA residuals are demeaned DIJA returns. To control for heterskedasticity or auto-correlation in regression residuals all regression standard errors are calculated using Newey-West standard errors with five lags. Using these controls makes our regression specifications comparable to those in Tetlock (2007).

AOTM is one of the most widely read market summary columns in the United States. It provides analysis of prior market activity, describes some notable company-specific events, and sometimes offers predictions for the future.[7] Electronic text copies of AOTM columns dated after 1984 are available from different sources. Data prior to 1984 is obtained from Proquest's historical *Wall Street Journal* archive, which stores the articles as scanned images. To convert these images to text files, we use ABBYY OCR software.[8] Typically, this process yields a high quality of transcription. Any errors in this process are likely to be idiosyncratic, and will thus bias the coefficients of interest to zero.

During our sample period, the AOTM column was published Monday through Friday with a

[6] For example, over this same time period the CRSP Value-Weighted Index annualized Sharpe ratio was also 0.41.

[7] See Tetlock (2007) for more discussion.

[8] OCR, or optical character recognition, is the electronic translation of scanned images of handwritten, typewritten, or printed text into machine-encoded text. The ABBYY OCR software we use performs OCR using intelligent character recognition (ICR). This type of OCR works by searching the scanned image for common elements such as open spaces, closed forms, lines, diagonals intersecting and so on to identify letters. Typically, the accuracy rates using ICR are very high.

few exceptions on national holidays. Occasionally, there are two articles published before the next trading day. In this instance only the most recent article is used. Additionally, during this period there are 40 days when the stock market is open, but no AOTM column is available from our data sources. Overall, our sample includes 9,552 articles, over a period of 9,592 open market days. We restrict our attention to journalists which wrote at least fifty AOTM columns. For a small number of articles (76), we are either unable to identify the author, or the author wrote fewer than fifty articles. A similarly small number (516) were co-written, in which case authorship is credited to neither journalist.[9] Overall, this results in a set of twenty-five authors, which account for over 80 percent of the articles in our sample period.

2.2 Journalist scheduling

Table 1 presents a number of statistics related to each journalist's writing schedule. Moving across the table, we first list the journalist's last name, the years he was active (AOTM was not authored by any female columnists during our sample period), and the total number of articles he wrote. As seen, a few journalists are responsible for the majority of the articles, with Hillery (2,413), O'Brien (1,215), and Talley (915) being credited the most frequently. The median author, McLean, is associated with 103 articles.

A crucial feature of our identification strategy is that journalists tend to alternate or rotate with one another over the same time period. We show this graphically in Figure 1, which plots with X's the dates each journalist wrote, separately for each columnist by row. We note that five authors were responsible for the bulk of the writing during 1970–1984, whereas in the late 1980s and early 1990s there was significantly more turnover. However, the more important observation is that at any point in time, there are multiple active authors. For example, at the year 1980 mark, we see frequent activity from four different columnists: Hillery, Elia, Marcial, and Metz. Inspection of other dates reveals a similar pattern. Without this overlap we would not be able to separately identify any impact journalists might have on investor behavior from simple time trends.

Returning to Table 1, we see also that journalists tend to write articles in relatively brief spells. Shown in the fourth column is *Number of Rotations*, which identifies the number of instances where, for each time a journalist wrote, a different columnist wrote the following day. For example,

[9] Our results are nearly identical if dual authorship is credited.

Marcial penned 625 articles, but had only only 364 *Rotations.* This means that for $625 - 364 = 261$ days, Marcial directly followed one of his own articles the next business day, e.g., writing on a consecutive Wednesday and Thursday. Because our empirical tests will ultimately compare market returns between days when, for example, Marcial's articles were published to days when they were not, these transitions are important. For journalists that write less frequently (the bottom half on the table), the typical spell falls between 1 and 2 days, whereas for the more frequent authors, spells are two to three times longer on average. Garcia is a notable outlier, writing over 588 columns, but rotating only 70 times, for an average spell length of over 8 days.

The final five columns show the breakdown of each columnist's writing by day of the week. From this it is clear that there are two distinct types of writers – those assigned for all or most of the business week, and those slated for one particular day. As an example of the former type, Garcia's articles are distributed relatively equally across all days, with Monday (16%) being only slightly less common than the other four days. O'Brien, Raghavan, and Gonzalez are other examples. By contrast, Browning's articles are almost always published on Mondays (90%), with Sease (69% on Mondays), Rosenberg (70% on Fridays), Ip (76% on Mondays), and Levingston (51% on Mondays), exhibiting similar concentration on a particular day.

Such strong day-of-the-week patterns across journalists suggests that at least part of what we observe arises from pre-determined, semi-regular schedules. Figure 2 gives some graphical intuition for this claim, plotting detailed schedules over three sample sub-periods: July–December 1972, July–December 1994, and July–December 2000. Each row corresponds to a different journalist. For example, in the top panel, the writings of Hillery are shown in the bottom row, Rosenberg in the row above, Dorfman above him, and an indicator for "No Author" in the top row. The X's correspond to Mondays, circles to Fridays, and crosses to the other three weekdays.

Looking first at the top panel, note the remarkable regularity between the two dominant writers, Hillery and Rosenberg. Of the 26 Mondays, Hillery wrote the AOTM article for 21 of them, and of the 26 Fridays, Rosenberg was active for all but 4. The pattern for the other weekdays is even more pronounced. Beginning after Hillery's first article (the second week shown), he missed only 13 Tuesdays, Wednesdays, or Thursdays. Six of these days were missed consecutively in the last two weeks of August, and three more the week leading up to Christmas – all almost certainly corresponding to vacation time. For nearly half the sample (12 weeks), a *completely* deterministic alternation between Hillery and Rosenberg is observed, with Rosenberg writing only on, but on every, Friday.

The second panel plots journalist schedules for the second sub-period, and although not as predictable as the first, nevertheless indicates considerable regularity. Kansas writes most Mondays, as well as some other days from time to time, while Pettit is the mode weekday writer over the period. Mollison appears to spell Pettit from his regular duties for two weeks in late August and early September respectively, but was otherwise active only sporadically. The remaining journalists – Bauman, Granahan, Levingston, Frank, and Arvedlund – wrote only an article or two each, and at seemingly random times. The final graph shows a similar pattern, with O'Brien being the regular, weekday writer, and except on rare occasions (e.g., the likely vacation week seen yet again prior to Labor Day), being active every day except Mondays. Over this period, AOTM articles published on Monday were authored by five journalists, with no apparent pattern.

An important caveat is that although the evidence in Table 1, and Figure 2 seem to suggest some degree of scheduling predictability, this is not necessary (although it would be sufficient) for our later return regressions to be properly specified. Because we will be examining returns *after* a given columnist's article is published, the main concern is that certain writer selection somehow depends on future market returns. Clearly, this will not be the case if journalist rotations are completely deterministic, as some periods in our sample appear to be, or completely random.

Even in times when the journalist selection rule is less obvious, note that mis-specification requires that: 1) editors have private knowledge of short-term market returns, 2) editors have an incentive to make selections based on this information, and 3) any such relationship be sign-specific – e.g., O'Brien must be *consistently* selected prior to abnormally good days, or Rosenberg prior to bad ones. Whether or not one views these possibilities as jointly, or even individually plausible, the predictability indicated in Figure 2 allows us to address such concerns directly. In robustness tests (Table 7), we will instrument for each journalist's arrival using his recent schedule, e.g., using whether Rosenberg wrote last Friday to predict whether he writes this Friday. Here, identification comes purely through factors orthogonal to future returns, and allows us to be more confident in our claim of a causal relation between media content and stock returns.

2.3 Writing styles

Relative to other studies that explore the impact of written content on investor behavior, an important methodological distinction is how we measure content. Almost without exception, similar papers have characterized written articles using computerized algorithms that count, for example, "negative" or "positive" words using financial dictionaries (e.g., Loughran and McDonald (2009)). Such procedures have the advantage of being able to link one or more specific content metrics to investor behavior. However, a shortcoming is that automated programs may neither completely, nor accurately, summarize how a *human* audience interprets the written word. To the extent that subtlety, phrasing, irony, understatement, or the countless other stylistic tools available to writers impact how an article is perceived, reliance on impersonal algorithms may cause the econometrician to miss important variation in news content.

The key to our identification is that such stylistic differences likely differ across journalists, but may be difficult to directly measure. It is obvious that Charles Dickens and William Faulkner, for example, employ different themes and rhetorical techniques in their writings, and indeed, such differences have occupied the attention of literary critics for decades. But it seems equally obvious that what truly makes Dickens 'Dickens' (or Faulkner 'Faulkner') cannot be easily quantified, regardless of how sophisticated the analysis may be. Like performing a violin concerto or preparing a fine meal, summarizing the nuances of a written article may be impossible, relative to simply specifying its creator.

Ultimately, this argument highlights both the main strength and weakness of our empirical strategy. While focusing on cross-journalist differences in writing style, we will implicitly capture any persistent differences in stylistic or thematic choices, no matter how difficult they may be to directly measure. On the other hand, such a reduced form approach is relatively agnostic about the underlying mechanisms – the tools each author employs to distinguish his writing from that of his peers. We will not focus on whether O'Brien chooses simpler sentences, writes longer article, uses more positive words, etc. – not only because we lack theoretical guidance,[10] but also because we have little hope of summarizing cross-journalistic differences with a small number of observable

[10] There is a complete lack of theory relating quantifiable measures of media content to stock returns. Existing empirical studies focus almost exclusively on an article's "tone," but there are a number of other measures – e.g., simplicity of language, article length, etc. – that could either affect returns directly, or through interaction with other metrics.

characteristics.

Nonetheless, the analysis of these observable content metrics, shown in Table 2, is useful, if for no other reason than to illustrate the magnitude of cross-journalist differences in writing style. Panel A shows means, standard deviations, and 90/10 percentile breakpoints for *Pessimism*, *Syllables*, *Words Per Sentence (WPS)*, *% Complex*, and *Fog*. Using Loughran and McDonald (2009), we count the number of "positive" and "negative" words, as well as the total number of words used by each author daily.[11] *Pessimism*, shown in the first column, is equal to the difference between the percentage of negative and positive words, but scaled across all authors to have zero mean and unit variance. The average number of syllables per word (*Syllables*), sentences per AOTM column (*Sentences*), and words per sentence *(WPS).* Complex words (*Complex*) are defined as having three or more syllables and are tabulated as a fraction of total words. The final column, *Fog*, reports the Fog readability score which indicates the number of years of formal education a reader of average intelligence would need to read and understand the article in one sitting. For example, a *Fog* score of 18 would be considered unreadable, a score of 12 appropriate for a high school graduate, and so on. As indicated by a mean of 11.1, the typical AOTM article would be appropriate for a high school senior.

What is most important for our purposes is the analysis in Panel B, which regresses each of these five content metrics on the vector of columnist fixed effects. Beginning with the first column, we see that journalists are, in aggregate, important predictors of the positive-negative word balance in AOTM articles. Regardless of how standard errors are calculated, the p-value for their joint significance is consistently less than 0.01%. The most persistently optimistic columnists are Marcial (-0.29) and Hillery (-0.17), while the most pessimistic are Elia (0.56) and Pettit (0.48). In interpreting these differences, it is important to note that these represent marginal effects, *after controlling for past returns, volume, and volatility.* Consequently, the coefficients in Table 2 capture the incremental impact of a given columnist writing, given whatever prevailing market conditions may be.

Moving across the table, we find that columnist fixed effects are even more important for the number of number of syllables, words per sentence, percentage of complex words, and the Fog in-

[11] See http://www.nd.edu/˜mcdonald/Word_Lists.html. We use these dictionaries in particular because they account for a number of the nuances related to financial language.

dex. The increase in explanatory power ranges from a full ten percent (for *Syllables*) to five percent (for *Fog*), and in most cases, well over half the journalists are individually significant at conventional levels. Note that differences in writing style do not appear correlated with writing activity; journalists at the top of the list (the most frequent authors) do not appear to be systematically more positive, more brief, or more complex than their sporadic counterparts. Overall, the data suggests there are significant differences in the writing style of the twenty-five journalists under consideration.

3 Journalists and market returns

3.1 Univariate patterns

We begin our main analysis with some simple univariate comparisons. In Table 3, for each columnist we list: 1) $\bar{r}_{wrote}$, the average excess return on the days his articles are published, 2) $\bar{r}_{day\ after}$, the average excess return the day after his article is published, and 3) $\bar{r}_{other}$, the average excess return on all other days over his writing tenure. For example, Table 1 indicates that Hillery authored 2413 AOTM articles from 1970 to 1984. The average DJIA return on the days these articles were published was slightly less than 1 bp, and on the day afterward, 3 bp. By contrast, the average DJIA return from 1995 to 2002 on the complement set of days when another journalist wrote the AOTM column was -3.4 bp. This comparison thus holds constant the average returns over each journalist's tenure, so that the fact that one journalist wrote mostly in the 1970s (when average returns were low), while another wrote in the late 1990s (when they were not), is not a concern.

We see that of the twenty-five *WSJ* columnists, over one-third are associated with significant abnormal returns, either the day their articles are published or the day afterward. These are split fairly evenly between positive and negative abnormal returns. In an absolute sense, a few journalists (e.g., Pasha and McGee) are associated with particularly striking abnormal returns, in the range of 40 bp per day. However, the more representative case, based on number of articles written, implies magnitudes roughly half to a quarter as large (e.g., O'Brien 20 bp on day $t+1$, Pettit -6 bp on day t).

3.2 Regressions

To more formally characterize the univariate patterns seen in Table 3, we estimate the following linear regression:

$$r_t = c + \sum_{i=1}^{25} (\beta_{i,t-1} \cdot Journalist_{i,t-1} + \beta_{i,t} \cdot Journalist_{i,t}) + \eta \cdot Controls_t + \epsilon_t,$$

where r_t is the excess return of the DJIA index on day t, as defined in Equation (1). All returns are nominal and are reported in basis points.

The variables of interest, the *Journalist* fixed effects, correspond to each of the twenty-five *WSJ* columnists, i, active from 1970-2007. If columnist i is the author of the AOTM for publication the morning of day t, then $Journalist_{i,t}$ takes a value of one, and zero otherwise. Going backward in time, $\beta_{i,t-1}$ captures the effect of columnist i's article published yesterday on today's returns, or alternatively, the effect of journalist i one day after his article is published.

Only a few variables have been shown to be significant predictors of one- or two-day returns: lagged returns, trading volume, lagged volatility, day of the week, and a dummy for the month of January. Of these, perhaps the most important in our context is the day of the week, given that Figure 2 indicates strong intra-week patterns for AOTM columnists. Our specification controls for the effects of these variables on market returns. In particular, the *Controls* vector includes five lags of returns, five lags of detrended daily log NYSE volume, five lags of detrended squared DJIA residuals (i.e. lagged volatility), day-of-the-week dummies, and a dummy variable for the month of January. Newey-West standard errors are calculated using five lags.

We begin with the first column of Table 4, which focuses on the effects of articles published the same day returns are measured (analogous to $\bar{r}_{wrote}$ in Table 3). The introduction of the control variables reduces the statistical significance, but in general, the point estimates are similar to those found in the univariate comparisons. The point estimates have the same sign in 19 of the 25 cases, and for all 6 cases of disagreement, the coefficients are estimated imprecisely. Below the coefficient estimates, we show the p-values for the joint test that $\beta_{i,t} = 0$ for all journalists i. Depending on how standard errors are calculated, the joint significance is between 1% and 4%.

Moving to the right, we see the impact on current returns of articles written for publication yesterday – i.e., we are explaining Wednesday's excess returns as a function of which journalist authored AOTM on Tuesday. This corresponds to the $\bar{r}_{day\ after}$ column in Table 3. As we see, the evidence is even stronger in this column, suggesting that the impact of specific journalists lasts more than one day. We find that seven journalists are now significant at the 10% level, compared to only four in the previous column. O'Brien, Rosenberg, Ip, Pasha, Dorfman, and McGee are all associated with abnormal excess return at better than the 5% level. The p-values for the joint linear restriction test indicate strong statistical significance (p-value < 0.001).

At the bottom of the first column, we present the R^2 of the excess return regression, with and without the full set of journalist fixed effects. The low explanatory power for both the restricted ($\beta_{i,t-1} = \beta_{i,t} = 0$) and unrestricted cases is expected, given that we are examining high frequency returns. Still, percentage wise, the improvement is impressive. Journalist fixed effects explain more than an additional 50% of daily excess returns, relative to that explained by time effects, recent returns, volatility, and trading volume.

Comparing the first and second columns of Table 4, it is interesting that with three exceptions – O'Brien, Pasha, and McGee – the journalists significant on day t are different than those significant on day $t-1$. Although not our main objective, a perhaps interesting observation is that journalists with an immediate impact appear to write simpler or more "digestible" articles. Even including journalists with return effects on both days, we note that day t writers are associated with fewer sentences (89.8 vs. 98.0), fewer words per sentence (11.6 vs. 12.1), a lower % of complex sentences (16.3% vs. 17.4%), and a lower Fog index (11.2 vs. 11.8). The comparison is even more dramatic if we exclude O'Brien, Pasha, and McGee: 85.8 vs. 100.0, 10.4 vs. 11.8, 16.4% vs. 18.0%, and 10.7 vs. 11.9. We stop short of making a causal inference from the differences here, noting only that it is consistent with investors having finite cognitive resources, and therefore needing processing time.[12]

3.3 When Does Financial Journalism Matter?

The tests so far have considered the average marginal impact of each journalist, but have ignored whether their effects are dependent upon market conditions. It is not obvious what we should expect, and largely depends on whether we view financial journalists as mostly providing interpre-

[12]See for example, Plumlee (2003), Petersen (2004), Engelberg (2008), DellaVigna and Pollet (2009) and Hirshleifer, Lim, and Teoh (2009).

tation of underlying events, or as primarily creating de novo content. To draw an analogy with other branches of journalism, should we think of financial journalists as being like sportswriters (who requires a game on which to comment), or to an investigative reporter expected to dig up her own facts and write a groundbreaking story?

Table 5 addresses whether the effects of financial journalism are strongest after days of extreme returns, both positive or negative. We start with the specification in Table 4, but interact each journalist indicator (both on day $t-1$ and day t) with the excess return on day $t-1$. The goal is to gauge the extent to which return conditions on say, Monday, influence how a journalist's whose article is published on Tuesday are perceived by the market, either on Tuesday or Wednesday.

There are two reasons why we might expect Monday's return conditions (in our example) to influence how investors respond to financial journalism on Tuesday or Wednesday. First, returns proxy for how much information is released to the market. To the extent that financial columnists use this information as a basis for their articles, we might expect stronger marginal effects. Second, extreme returns – particularly large negative returns – may proxy for swings in investor sentiment. Borrowing the methodology from Tetlock (2007), García (2010) shows that news-response coefficients in return regressions are stronger in recessions, and interprets this result as investors being more susceptible to slant in reported news during bad times.

Looking first at columns 1 and 2, we see that even in the presence of the interaction terms, the coefficients and statistical significance are similar compared to Table 4. Of more interest are the final two columns, which show how these slopes are affected by the return environment. In the third column, we find that nearly half of the journalists have a significant interaction coefficient, and nearly as many are significant in the final column.

The diagnostic statistics at the bottom of the table formalize the importance of the return interactions in the return regressions. Recall that in Table 4, the inclusion of day t and $t-1$ journalists increased the R^2 from 2.8% to 3.8%. Here, we see a marked improvement with the journalist-return interactions, to over 6%. Note that all specifications (even the baseline without journalists) include lagged returns – instead, it is the interaction with the journalist fixed effects that makes the dramatic difference. At the bottom of the table, we present p-values for the journalist indicators, the return-journalist indicators, and their union. Regardless of how standard errors are calculated, the joint hypothesis that our coefficients are zero is rejected at better than the 0.1%

level.

Given that the journalist-return interactions are important predictors – indeed, even more important than just the journalist indicators alone – it is worth being explicit about their interpretation. While the columnist fixed effects themselves can be interpreted as capturing each columnist's bullishness or bearishness, the interaction terms measure whether a given columnist contributes to, or detracts from, short-term return continuation. In other words, the interactions tell us whether a given journalist amplifies the effects of past returns, or whether he plays an attenuating role.

Take the most frequently credited author, Hillery, as an example. The first two coefficients, β_{t-1} and β_t, indicate that on average, his writings have neither a positive nor negative impact on future returns. However, $\gamma_t = -0.145$ indicates that the days after an excess return of +100 basis points (roughly the standard deviation of the DJIA), the market would be expected to decline by almost 15 basis points on days when Hillery's articles are published. The +0.214 coefficient on the $t-1$ coefficient suggests that this entire effect is (more than reversed) the following day.

Indeed, such reversals are the rule rather the exception. Comparing the interaction coefficients in the last two columns of Table 5, we see that the signs flip in the vast majority of cases. This suggests that although the effects of financial journalism are magnified around extreme returns, the effects are temporary, and unlikely to have permanent impacts on asset prices. This result is not only consistent with the evidence in Tetlock (2007), who finds that the effects of word-based predictability reverse within a week, but also with the intuition that financial journalists are unlikely to possess information advantages about the aggregate stock market.

Taken together, the results in this section paint a clearer picture of the power of individual rhetoric in financial journalism. If we think of journalists as actors with the goal of persuading an audience, the results in Tables 3-4 suggest that an actor's identity *alone* tells us something about the performance he will give. Furthermore, Table 5 indicates that an actor's performance also depends upon the stage he is given. When the stage is set for good news, some journalists make the good news sound even better, while others temper such enthusiasm. By contrast, when the stage is set for bad news, some journalists are somber while others look on the bright side. Interestingly, it is precisely in times with large price movements when the power of rhetoric has the greatest effect on investors.

4 Alternative Specifications and Robustness

4.1 Endogeneity of Journalist Arrival

The evidence in Tables 3-6 indicates persistent correlation between certain journalists writing and subsequent market returns. In this section, we address the possibility that these tests do not identify a causal relation. The specific concern is that the selection of columnists may not be exogenous with respect to future market returns – e.g., an editor choosing a certain journalist around releases of bad news. If not, then the patterns we observe may be spurious, and thus, tell us nothing about the financial media's ability to influence investor behavior. This is clearly not a problem if the process by which journalists are selected is completely exogenous – preset schedules being a special case. However, although there appears to be considerable predictability in how journalists are chosen (see Figure 2), it is equally clear that the selection is not completely deterministic, leaving our specifications open to omitted variable bias.

The traditional solution for this problem is to look for an instrumental variable that is correlated with the potentially endogenous variable (here the vector of journalist dummies), but is otherwise unrelated to the dependent variable (future excess market returns). Intuitively, we want to project each journalist's actual schedule onto explanatory variables that we know cannot be systematically correlated with future returns, and use these projections in place of the potentially endogenous variables. Any observed relation thus results from the part of the endogenous variable that is explained solely by exogenous factors, and thus cannot be susceptible to the endogeneity critique.

We are fortunate to be afforded a nearly perfect instrument: each journalist's recent writing schedule. Because journalists tend to write in relatively short bursts, and often on the same day of the week, we can use past writing activity as an instrument for current writing activity. The key to this being exogenous is the time lag. For a given Tuesday in 1972, for example, we use as instruments which journalist wrote on Monday (yesterday), as well as which journalist wrote the Tuesday one week ago and day-of-the-week dummies for that year. Together, these instruments are powerful predictors of actual writing activity, but have no systematic relation to stock returns.

Table 6 shows the performance of our instrumental variable specification. We run a daily linear probability model for each journalist, over his respective tenure – i.e., for only the years 1970-1984 for Hillery, 1995-2002 for O'Brien, etc. In the second column, we report the R^2 using only year fixed effects; as seen, this generally produces a poor fit. The third column adds market variables which include five lags of recent returns, five lags of de-trended lagged volume, and five lags of squared DJIA residuals. In most cases, these controls lead to a small increase in explanatory power. In the fourth column, in addition to these variables, we include as regressors year fixed effects interacted with day of the week dummies, and two dummies for journalist i – one dummy indicating if journalist i wrote on the previous day and one indicating if journalist i wrote on the same day the previous week. In all cases, these additional controls substantially increase explanatory power, mostly due to the day of the week interactions, which Figure 2 indicates are strong predictors for most journalists.

Column five presents a formal test of whether the market variables help predict journalist arrivals. For six of the twenty-five journalist there is some predictability, which argues for the IV analysis that we shall construct below. The final column shows the additional explanatory power afforded by our instrumental variables, the one-day and one-week lagged $Journalist$ indicators, interacted with the relevant year. In all but one case – Gonzalez – specifying whether a columnist wrote yesterday, or that particular day the previous week dramatically improves fit. Increases in R^2 in the range of 30% are common, and even bigger improvements are seen in some cases. The p-values in the rightmost column show this formally, calculated against the null that the coefficients on the instruments are simultaneously zero. As indicated by their very low values (nearly all are below 0.001), a journalist's past writing schedule is very valuable for predicting his near-term future activity.

It is important to note the difference in terms of fit that the market variables provide, relative to the rotation variables. Even though market variables help predict the arrival of Browning (p-value 6%), the increase in the R^2 of the specification from adding such variables is 5.6%. In contrast, adding the rotation variables increases the R^2 to 45.8% – clearly knowing who wrote last Thursday is more important for determining who writes this Thursday than what happened in the markets over the last week.

In Table 7 we use the fitted values of the LPMs from Table 6, instead of the journalist's actual writing activity, in order to predict DJIA returns. For example, for a given day in 1972, if the journalist rotation model (Table 6) predicts that Hillery will write with 60% probability, we use this fitted value rather than a zero or one, as we did when referring to Hillery's *actual* writing activity in Table 5. During the period when a journalist is not actively writing, these fitted values are automatically set to zero.

Comparing the second column of Table 7 with the second column in Table 5, the similarity is readily apparent. All six columnists with return coefficients significant at the 10% level in Table 5 have point estimates of the same sign in Table 7, and five (except O'Brien) retain similar statistical significance. The evidence is a bit weaker in the first column. Although we observe similar magnitudes in the IV regression for the three significant journalists in the baseline specification (O'Brien -0.109 non-IV vs. -0.194 IV, Pasha 0.475 non-IV vs. 0.291 IV, and McGee 0.370 non-IV vs. 0.358 IV), the noise introduced by the instrumental variable specification reduces the statistical significance below conventional levels for these journalists. However, the joint significance of all columnist coefficients, β_{t-1} and β_t is well below the 1% level, as seen by the linear restriction test directly below the first column. This result holds for standard errors calculated under the normal OLS assumptions, as well as when the White or Newey-West procedure is used.

The third and fourth columns of Table 7 show the coefficients on the return interactions under the IV specification. Like the previous columns, the instrumental variable specification generally produces similar point estimates, but somewhat weaker statistical significance. In the third column, the author interactions associated with the most significant effects – Hillery ($t = 4.2$) Garcia ($t = 10.0$), Rosenberg ($t = 3.9$), and McGee ($t = -3.5$) – all remain significant in the IV regression, with most of the remaining interactions similar across specifications. In the fourth column, 18 of the 25 journalist interactions have the same estimated sign, with nearly unanimous agreement for the most precisely estimated ones. The joint linear restriction test that all interactions are simultaneously zero ($\gamma_{t-1} = \gamma_t = 0$) is soundly rejected ($p < 0.001$ for all standard error assumptions), as is the joint test of all columnist indicators and their interactions. This latter result is unsurprising, given that the improvement in R^2 due to the columnist variables is comparable to that seen in Table 5, increasing from 0.028 to 0.062.[13]

[13]The results are similar if, instead of using the raw probabilities (the outputs of Table 6) as inputs into the IV regression: 1) we take the maximum probability observed across journalists, and assign a value of one to that journalist, or 2) we scale all probabilities so that they sum to one across journalists for every date.

4.2 Different Return Series and Other Robustness

The first five columns of Table 8 presents the results of the Table 4 specification, but vary the return series. Panel A shows the p-values for the linear restriction test that all one-day lagged journalist coefficients are zero, and Panel B shows the corresponding p-values for journalists published on day t. Aggregated linear restriction tests are considered in Panel C.

In the first column, we consider open-to-close DJIA returns. The second and third column present the results when instead we use excess returns on the S&P 500 Index or CRSP value-weighted index as our dependent variable. As seen, the results are similar for each of these alternatives, compared to the DJIA results shown in Table 4. The fourth and fifth columns consider GARCH-adjusted and winsorized returns, respectively, which we conduct to address the concern that a few outliers may be responsible for the return patterns we document. In the regression results reported in the *Winsorized* column returns, volume, and all non-dummy variables are winsorized at the 5 percent level. GARCH-adjusted returns are defined as DJIA close-to-close returns divided by the estimated daily volatility from a GARCH(1,1) model estimated on the same return series. Both procedures strengthen the statistical significance for the day t columnist fixed effects, and lead to similar significance for the day $t-1$ indicators.

The first five columns in Table 9 repeat the interaction regressions of Table 5, but with the alternative return series. As with the previous table, none of our previous conclusions change. In particular, the joint significance of both the fixed effects (β_{t-1} and β_t), and their interactions with returns (γ_{t-1} and γ_t) holds across *all* specifications.

The final column in Tables 8 and 9 repeat the specifications in Tables 4 and 5, expect that we ignore all but the ten most frequently credited authors. As indicated in the summary stats (Table 1), this restriction implies that we are considering only authors with at least 157 written articles. Table 8 indicates that the p-value for joint significance at day $t-1$ is roughly 1%, and about 10% for day t authors. Considered together in Panel C, journalists remain jointly significant at the .4% level. In Table 9, we consider only these authors in the same specification shown in Table 5. Like the previous columns, journalist-return interactions are highly significant predictors of short term market returns.

We end the analysis with a falsification test, and further study whether the effect of journalists' writing can extend over the two-day window that we have focused on this far. In particular, we augment (3) with one more lag and two more leads. The variable $\beta_{i,t-2}$ captures any residual effect from an article published two days prior. Going the other direction, $\beta_{i,t+1}$ and $\beta_{i,t+2}$ are estimated for falsification. They measure the effect of articles written for publication on future dates (e.g., an article for publication on Thursday influencing Wednesday's returns), and consequently, should have no effect.

Table 10 presents the point estimates of such a specification. The rightmost column indicates that generally, any observed journalist-return effects will show up within two business days. Only two journalists – Hillery and Ip – have significant coefficients, and the high p-values at the table's bottom indicate that together, they add little explanatory power to the regression.

The first two columns test for return effects that, under a causal interpretation, should not produce significant results. Each vector β_{t+1} and β_{t+2} measures columnist who write on future days, after controlling for current and past authors. For example, if we are measuring Wednesday's excess returns, columns 5, 4, and 3, respectively, tell us which AOTM author was published on Monday ($t-2$), Tuesday ($t-1$) and Wednesday (t), while columns 1 and 2 pick up who *will* be published on Thursday and Friday respectively. As expected, future authors have no apparent relation to current returns, with Newey-West p-values of 0.52 and $p = 0.66$ for $t+1$ and $t+2$, respectively.

5 Conclusion

There is widespread speculation that the news media has the power to influence financial markets, apart from simply reporting events. Yet, such claims are often based on anecdotal associations that make causal inferences difficult. For example, times of negative financial reporting frequently coincide with bad economic news, and vice versa. Stripping away the effects of only the *reporting* thus requires variation in news content that is unrelated to underlying fundamentals.

The identification strategy of this paper is based on two assumptions. The first is that authors of the *Wall Street Journal's* "Abreast of the Market" column exhibit persistent stylistic differences, such that even for the same set of facts, article content will vary. The second is that the selection of journalists is not systematically related to future returns, an assumption relatively easy to justify given that we are examining returns on a nearly unpredictable market index.

Our results suggest that financial journalists have the potential to influence investor behavior, at least over short time horizons. Adding journalist fixed effects to a daily return regression significantly increases explanatory power, and when these fixed effects are interacted with recent returns, the implied return predictability is even more dramatic. Overall, our results suggest that the interpretation of public news is important, as the effects we uncover are strongest when journalists write about significant market moves.

An important caveat is that although our empirical design permits a causal interpretation, our analysis does not shed light on the specific rhetorical tools that authors use to influence investor behavior. That is, we do not attempt to say whether longer articles, more complex words, or less pessimism leads to a predictable market response. This is not because we cannot quantify a number of content measures, but instead because we think that attempting to gauge a human audience's response may be difficult using computerized algorithms, relative to using statistical stand-ins for human authors. Clearly, we sacrifice the ability to pinpoint specific stylistic techniques, but we hopefully gain by capturing other unobservable elements that vary across journalists.

By documenting causal effects of the media on aggregate market prices, our findings paint a somewhat ominous picture of financial journalism. One recalls Shiller's (2000) less-than-veiled indictment: "The history of speculative bubbles begins roughly with the advent of newspapers" (p. 85). His implication is as clear as it is concerning – if financial journalists can manipulate investor beliefs apart from fundamentals, then their actions and incentives play a direct role in prices and allocations. The evidence in this paper, particularly as it applies to aggregate allocations, calls for a better understanding of these issues.

References

[1] Chevalier, J., Ellison, G., 1999. "Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance," *Journal of Finance* 54 (3), 875-899.

[2] DellaVigna, S., Kaplan, E., 2007. "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics* 122 (3), 1187-1234.

[3] DellaVigna, S., Pollet, J., 2009. "Investor Inattention and Friday Earnings Announcements," *Journal of Finance* 64 (2), 709-49.

[4] Engelberg, J., 2008. "Costly Information Processing: Evidence from Earnings Announcements," *Working Paper*, University of North Carolina.

[5] Engelberg, J., Parsons, C., 2010. "The Causal Impact of Media in Financial Markets," *Forthcoming at the Journal of Finance.*

[6] Fama, E., 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 15 (2), 383-417.

[7] French, K., 1980. "Stock Returns and the Weekend Effect," *Journal of Financial Economics* 8 (1), 55-69.

[8] García, D., 2010. "Sentiment During Recessions," *Working Paper*, University of North Carolina, Kenan-Flagler School of Business.

[9] Gentzkow, M., and Shaprio, J., 2006. "Media Bias and Reputation," *Journal of Political Economy* 114 (2), 280-316.

[10] Gerber, A., Karlan, D., and Bergan, D., 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions," *American Economic Journal: Applied Economics* 1 (2), 35-52.

[11] Hirshleifer, D., Lim, S., Teoh, S., 2009. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News," *Journal of Finance* 64 (5), 2289-2325.

[12] Huberman, G., Regev, T., 2001. "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar," *Journal of Finance* 56 (1), 387-396.

[13] Loughran, T., McDonald, B., 2009. "When is a Liability not a Liability?," *Journal of Finance* , forthcoming.

[14] Milton, J., 1644. "Areopagitica: A Speech of Mr. John Milton for the Liberty of Unlicensed Printing, to the Parliament of England."

[15] Mullainathan, S., Shleifer, A., 2005. "The Market for News," *The American Economic Review* 95 (4), 1031-1053.

[16] Newey, W., West, K. 1987. "A simple, positive definite, heteroscedastic and autocorrelation consistent covariance matrix," *Econometrica* 55, 703708.

[17] Petersen, M., 2004. "Information: Hard and Soft," *Working Paper*, Northwestern University.

[18] Plumlee, M., 2003. "The Effect of Information Complexity on Analysts' Use of that Information," *The Accounting Review* 78 (1), 275-296.

[19] Reuter, J., Zitzewitz, E., 2006. "Do Ads Influence Editors? Advertising and Bias in the Financial Media," *Quarterly Journal of Economics* 121 (1), 197-227.

[20] Shiller, R., 2000. "Irrational Exuberance," *Princeton University Press, Princeton.*

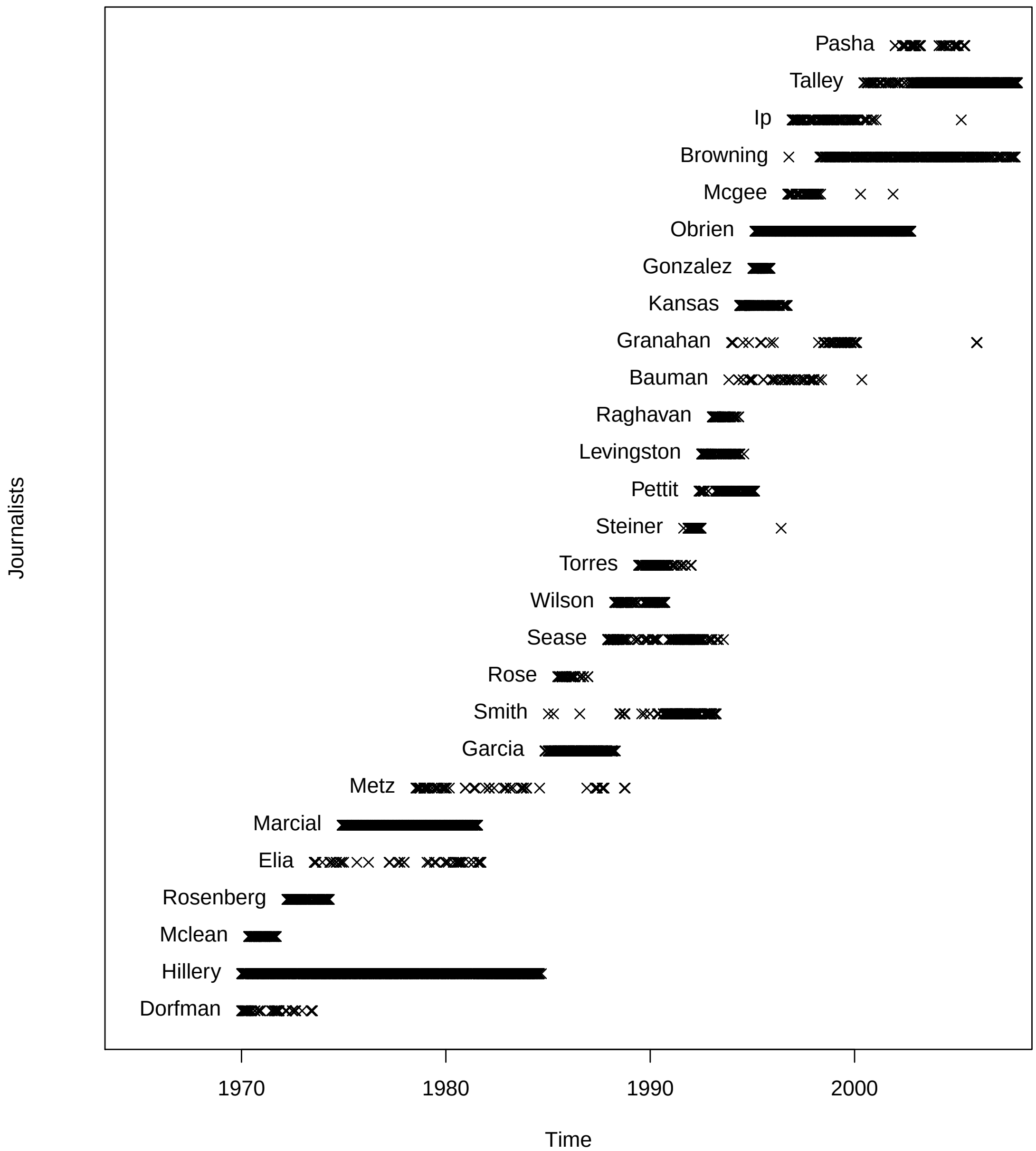
[21] Sialm, C., Shoven, J., 2000. "The Dow Jones Industrial Average: The Impact of Fixing Its Flaws," *The Journal of Wealth Management* 3 (3), 9-18.

[22] Tetlock, P., 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market.," *Journal of Finance* 62 (3), 1139-1168.

[23] Tetlock, P., 2009. "All the News That's Fit to Reprint: Do Investors React to Stale Information?," *Working Paper*, Columbia University.

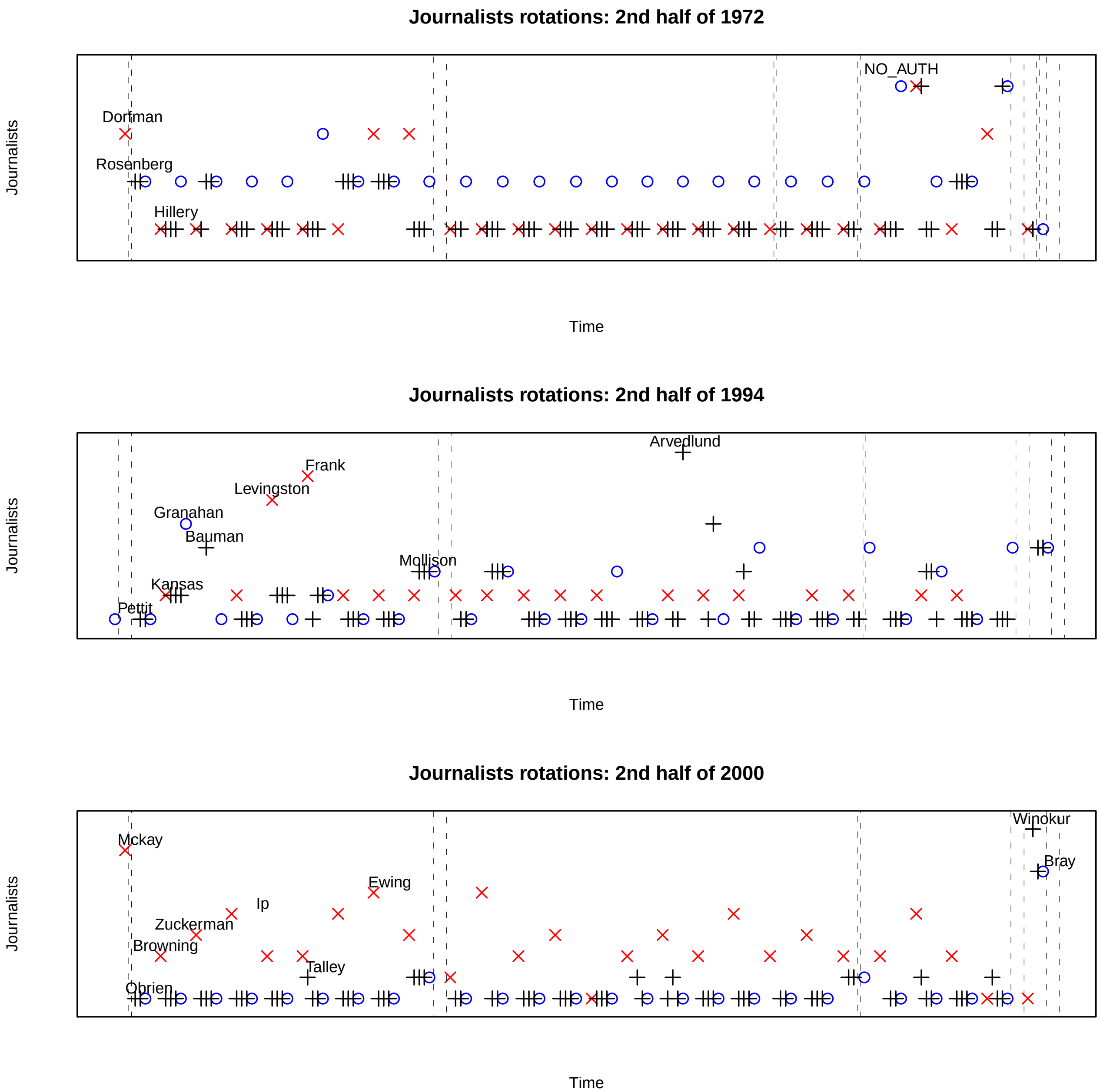
[24] Wooldridge, J., 2009. "Introductory Econometrics," *South-Western College Pub* (4), 269.

Figure 1: Timeline of AOTM journalists



This Figure documents the authorship of the *Wall Street Journal* "Abreast of the Market" column for our full sample time-period. Each point corresponds to an author writing the AOTM column on a given day.

Figure 2: Sample of journalist writing days



This Figure documents the authorship of the *Wall Street Journal* "Abreast of the Market" column for six 6-month subsets of our sample time-periods. For each subsample, we plot as bars of different heights the different authors of the AOTM column.

Table 1: Statistics on journalists' tenure

This table presents statistics for each journalist who wrote more than fifty articles for the AOTM column. In particular, it lists the last names of the journalists, the years they were actively writing for the AOTM column (Years Active), the total number of articles they published (Articles), the total number of consecutive writing days for each journalist (Number of rotations), the average length of these rotations (Average rotations), and the percentage of articles each journalist published on each weekday.

Journalist	Years Active	Articles	Number of rotations	Average length	% Mon.	% Tue.	% Wed.	% Thu.	% Fri.
Hillery	1970 – 1984	2413	708	3.4	0.18	0.23	0.25	0.25	0.09
O'Brien	1995 – 2002	1215	415	2.9	0.01	0.24	0.26	0.25	0.24
Talley	2000 – 2007	915	289	3.2	0.01	0.23	0.26	0.25	0.25
Marcial	1974 – 1981	625	364	1.7	0.23	0.19	0.13	0.12	0.34
Garcia	1984 – 1988	588	70	8.4	0.16	0.21	0.21	0.20	0.21
Smith	1985 – 1993	302	140	2.2	0.02	0.19	0.24	0.25	0.30
Wilson	1988 – 1990	251	97	2.6	0.01	0.23	0.25	0.28	0.23
Browning	1996 – 2007	250	249	1.0	0.90	0.10	0.00	0.00	0.00
Pettit	1992 – 1995	222	109	2.0	0.00	0.21	0.22	0.29	0.28
Sease	1987 – 1993	157	115	1.4	0.69	0.13	0.06	0.06	0.06
Rosenberg	1972 – 1974	125	95	1.3	0.00	0.07	0.13	0.10	0.70
Kansas	1994 – 1996	104	77	1.4	0.61	0.15	0.10	0.09	0.06
McLean	1970 – 1971	103	69	1.5	0.00	0.14	0.12	0.17	0.58
Raghavan	1993 – 1994	93	58	1.6	0.30	0.22	0.22	0.16	0.11
Ip	1996 – 2005	90	80	1.1	0.76	0.12	0.03	0.06	0.03
Gonzalez	1995 – 1995	87	50	1.7	0.18	0.22	0.18	0.23	0.18
Metz	1978 – 1988	80	57	1.4	0.16	0.20	0.15	0.15	0.34
Levingston	1992 – 1994	77	50	1.5	0.51	0.18	0.12	0.10	0.09
Pasha	2001 – 2005	74	36	2.1	0.03	0.23	0.26	0.27	0.22
Rose	1985 – 1986	65	14	4.6	0.14	0.22	0.22	0.23	0.20
Steiner	1991 – 1996	63	33	1.9	0.10	0.37	0.30	0.16	0.08
Dorfman	1970 – 1973	62	56	1.1	0.21	0.06	0.05	0.06	0.61
Bauman	1994 – 2000	52	33	1.6	0.00	0.21	0.27	0.17	0.35
McGee	1996 – 2001	51	51	1.0	0.96	0.04	0.00	0.00	0.00
Elia	1973 – 1981	50	46	1.1	0.68	0.08	0.06	0.02	0.16

Table 2: Article characteristics and journalists

Panel A of this table reports summary statistics for different article characteristics. Pessimism represents the average amount of sentiment in each article as measured by the percentage of negative words minus the percentage of positive words per article. Pessimism is normalized to have a zero mean and unit variance. Syllables is the average number of syllables for all words in an article. WPS is the average number of words per sentence. % Complex is the average percentage of words in each article with three or more syllables. Fog reports the average Fog readability score.

Panel B reports coefficient estimates from the following regression:

$$Characteristic_{j,t} = c + \sum_{i=1}^{25} \beta_i \cdot Journalist_{i,t} + \eta \cdot Controls_t + \epsilon_{j,t}$$

where $Characteristic_j \in \{\text{Syllables, WPS, \% Complex, Fog}\}$ and the $Controls$ vector includes five lags of daily excess DJIA return, five lags of detrended daily log NYSE volume, five lags of detrended squared DIJA residuals, day-of-the-week dummies, year fixed-effects, a dummy variable for the month of January, and year fixed-effects. The table also presents the number of observations, the R-squared for an unreported regression with no journalist fixed effects, R^2_{noJFE}, and the R-squared for the model that includes the journalist fixed effects, R^2_{JFE}. We also present the p-values from an F-test of the following null hypothesis: $\beta_i = 0$, $\forall i$. For each test, p-values are reported for F-tests calculated using the OLS variance/covariance matrix, a heteroscedasticity robust variance/covariance matrix (WHITE), and a Newey-West variance/covariance matrix using five lags (NW5). t-statistics that are statistically significant at the 1% level are indicated by ***, at the 5% level by **, and at the 10% level by *.

	Pessimism		Syllables		WPS		% Complex		Fog	
Panel A. Sample statistics										
Mean	0.00		1.73		10.64		17.12		11.10	
Standard deviation	1.00		0.06		3.41		1.89		1.46	
10% percentile	−1.23		1.70		6.70		14.70		9.50	
90% percentile	1.27		1.80		14.60		19.40		12.80	
Panel B. Multivariate regression coefficients										
Hillery	−0.17***	(−3.1)	0.00	(0.5)	−0.46***	(−2.9)	−0.09	(−0.9)	−0.22***	(−3.0)
O'Brien	0.27***	(5.1)	0.05***	(15.5)	−0.43***	(−2.8)	0.84***	(8.2)	0.16**	(2.2)
Talley	0.08	(1.6)	0.04***	(11.0)	−2.58***	(−15.9)	0.43***	(4.1)	−0.85***	(−11.3)
Marcial	−0.29***	(−4.5)	0.00	(0.9)	1.89***	(9.6)	0.08	(0.6)	0.78***	(8.5)
Garcia	0.10	(1.3)	−0.01	(−1.3)	−0.16	(−0.7)	−0.41***	(−2.8)	−0.23**	(−2.2)
Smith	0.30***	(4.3)	−0.02***	(−3.9)	−0.07	(−0.3)	−0.61***	(−4.5)	−0.28***	(−2.9)
Wilson	−0.01	(−0.1)	0.00	(0.1)	0.98***	(4.8)	−0.31**	(−2.2)	0.27***	(2.8)
Browning	0.05	(0.7)	−0.08***	(−19.3)	2.67***	(13.2)	−2.20***	(−16.5)	0.19**	(2.0)
Pettit	0.48***	(5.9)	0.00	(0.7)	−1.02***	(−4.2)	0.16	(1.0)	−0.35***	(−3.1)
Sease	0.25***	(3.2)	−0.02***	(−5.0)	0.86***	(3.7)	−0.35**	(−2.3)	0.20*	(1.8)
Rosenberg	0.03	(0.3)	0.02***	(2.9)	0.69**	(2.4)	0.42**	(2.2)	0.45***	(3.3)
Kansas	−0.08	(−0.8)	0.03***	(4.7)	0.05	(0.2)	0.60***	(3.1)	0.25*	(1.9)
McLean	−0.19*	(−1.7)	0.01	(1.1)	−0.94***	(−2.9)	0.33	(1.6)	−0.25*	(−1.7)
Raghavan	0.02	(0.2)	−0.01*	(−1.9)	0.23	(0.7)	0.08	(0.4)	0.11	(0.7)
Ip	0.07	(0.7)	0.01	(1.3)	0.99***	(3.2)	0.60***	(3.0)	0.64***	(4.5)
Gonzalez	0.23**	(2.0)	0.00	(−0.1)	0.57	(1.6)	0.62***	(2.6)	0.47***	(2.8)
Metz	−0.07	(−0.6)	−0.01	(−0.9)	−0.09	(−0.3)	−0.44**	(−2.1)	−0.21	(−1.4)
Levingston	0.15	(1.4)	0.01*	(1.8)	−1.17***	(−3.5)	0.47**	(2.1)	−0.29*	(−1.9)
Pasha	0.26**	(2.4)	0.00	(0.4)	−1.08***	(−3.3)	−0.17	(−0.8)	−0.49***	(−3.2)
Rose	−0.05	(−0.4)	−0.01	(−0.8)	−0.60	(−1.6)	−0.40	(−1.6)	−0.40**	(−2.2)
Steiner	−0.08	(−0.6)	0.01*	(1.9)	1.21***	(3.3)	0.88***	(3.6)	0.83***	(4.8)
Dorfman	0.13	(1.0)	0.01	(1.1)	1.12***	(3.0)	0.38	(1.6)	0.60***	(3.5)
Bauman	−0.16	(−1.2)	0.03***	(3.6)	1.28***	(3.4)	0.93***	(3.7)	0.89***	(5.0)
McGee	−0.05	(−0.4)	−0.02*	(−1.8)	1.99***	(5.1)	−0.34	(−1.3)	0.66***	(3.6)
Elia	0.56***	(4.3)	0.00	(0.5)	−1.20***	(−3.1)	−0.16	(−0.6)	−0.55***	(−3.0)
Observations	9552		9552		9552		9552		9552	
R^2_{noJFE}	0.30		0.20		0.39		0.19		0.32	
R^2_{JFE}	0.31		0.30		0.46		0.25		0.37	
p-value OLS	0.00		0.00		0.00		0.00		0.00	
p-value WHITE	0.00		0.00		0.00		0.00		0.00	
p-value NW5	0.00		0.00		0.00		0.00		0.00	

Table 3: Univariate return tests

For each journalist, this table presents the average daily excess return of the DJIA on the days they wrote, $\overline{r}_{wrote}$, the days after they wrote, $\overline{r}_{\text{day after}}$, and for all other days, $\overline{r}_{other}$, during the period they were actively writing for the AOTM column. Column four presents the t-statistic for a test of the difference $\overline{r}_{wrote} - \overline{r}_{other}$, and column six the t-statistic for $\overline{r}_{day\ after} - \overline{r}_{other}$. The t-statistics that are statistically significant at the 1% level are indicated by ***, at the 5% level by **, and at the 10% level by *.

	$\overline{r}_{\text{other}}$	$\overline{r}_{\text{wrote}}$	t-stat	$\overline{r}_{\text{day after}}$	t-stat
Hillery	−0.034	0.007	(1.0)	0.031	(1.2)
O'Brien	−0.036	−0.011	(0.3)	0.202^{***}	(2.6)
Talley	−0.021	0.025	(0.9)	0.066	(1.0)
Marcial	0.046	0.036	(−0.2)	-0.055^{*}	(−1.8)
Garcia	0.005	0.112	(1.1)	−0.249	(−1.0)
Smith	0.035	0.067	(0.5)	0.170	(1.3)
Wilson	0.107	-0.041^{*}	(−1.8)	−0.020	(−1.2)
Browning	0.029	0.066	(0.5)	0.002	(−0.4)
Pettit	0.050	-0.057^{**}	(−2.0)	0.084	(0.5)
Sease	0.026	0.053	(0.3)	0.133	(1.2)
Rosenberg	0.099	-0.095^{*}	(−1.9)	-0.304^{***}	(−3.6)
McLean	0.076	0.119	(0.6)	−0.005	(−1.0)
Kansas	0.103	0.069	(−0.3)	−0.035	(−1.1)
Raghavan	−0.024	0.200^{***}	(2.9)	−0.021	(0.0)
Ip	0.013	0.195	(1.4)	0.118	(0.8)
Gonzalez	0.010	0.107	(1.2)	0.225^{**}	(2.3)
Metz	0.029	0.031	(0.0)	0.041	(0.1)
Levingston	0.029	0.042	(0.2)	−0.031	(−0.7)
Pasha	0.011	0.275^{*}	(1.9)	-0.431^{**}	(−2.3)
Rose	0.123	0.034	(−0.8)	−0.084	(−0.9)
Dorfman	0.057	−0.076	(−1.6)	0.142	(0.7)
Steiner	0.027	0.028	(0.0)	-0.258^{***}	(−2.6)
Bauman	0.073	0.001	(−0.5)	−0.064	(−0.7)
McGee	0.003	0.411^{**}	(2.4)	0.406^{**}	(2.4)
Elia	−0.005	−0.191	(−1.3)	0.027	(0.2)

Table 4: Multivariate return regressions

This table presents OLS coefficient estimates for

$$r_t = c + \sum_{i=1}^{25} \{\beta_{i,t-1} \cdot Journalist_{i,t-1} + \beta_{i,t} \cdot Journalist_{i,t}\} + \eta \cdot Controls_t + \epsilon_t$$

The *Controls* vector is as defined in Table 2. Coefficient estimates are recorded on the left-hand side of each column and their corresponding OLS t-statistics are presented in parentheses on the right-hand side. This table also presents the number of observations in each regression, the R-squared for an unreported regression with no journalist fixed effects, R^2_{noJFE}, and the R-squared for the reported regression that includes the journalist fixed effects, R^2_{JFE}. Also recorded are p-values from F-tests testing the following null hypotheses: $\beta_{i,t-1} = 0$, $\forall i$; $\beta_{i,t} = 0$, $\forall i$; and $\beta_{i,t-1} = \beta_{i,t} = 0$, $\forall i$. For each test, p-values are reported for F-tests calculated using the OLS variance/covariance matrix, a heteroscedasticity robust variance/covariance matrix (WHITE), and a Newey-West variance/covariance matrix using five lags (NW5). t-statistics that are statistically significant at the 1% level are indicated by ***, at the 5% level by **, and at the 10% level by *.

	publication date		day after	
	β_t	t-stat	β_{t-1}	t-stat
Hillery	0.038	(0.6)	−0.007	(−0.1)
O'Brien	−0.121*	(−1.9)	0.202***	(3.2)
Talley	0.010	(0.2)	0.039	(0.6)
Marcial	0.078	(1.0)	−0.105	(−1.4)
Garcia	0.204*	(2.0)	−0.118	(−1.1)
Smith	−0.026	(−0.3)	0.144*	(1.7)
Wilson	−0.032	(−0.4)	−0.133	(−1.5)
Browning	0.064	(0.8)	0.024	(0.3)
Pettit	−0.098	(−1.0)	0.111	(1.1)
Sease	0.010	(0.1)	0.072	(0.8)
Rosenberg	−0.047	(−0.4)	−0.304***	(−2.6)
Kansas	0.030	(0.3)	0.050	(0.4)
McLean	0.091	(0.7)	−0.026	(−0.2)
Raghavan	0.175	(1.3)	0.037	(0.3)
Ip	0.110	(0.9)	0.255**	(2.1)
Gonzalez	−0.076	(−0.5)	0.206	(1.4)
Metz	0.073	(0.6)	−0.081	(−0.6)
Levingston	0.038	(0.3)	−0.068	(−0.5)
Pasha	0.487***	(3.4)	−0.296**	(−2.1)
Rose	0.285	(1.3)	−0.323	(−1.5)
Steiner	−0.134	(−0.9)	0.027	(0.2)
Dorfman	0.050	(0.3)	−0.321**	(−2.2)
Bauman	−0.016	(−0.1)	−0.134	(−0.8)
McGee	0.386**	(2.5)	0.536***	(3.4)
Elia	−0.118	(−0.8)	0.060	(0.4)
Observations	9592			
R^2_{noJFE}	0.028			
R^2_{JFE}	0.038			
H_0 :	$\beta_{t-1} = 0$	$\beta_t = 0$		$\beta_{t-1} = \beta_t = 0$
p-value OLS	0.000	0.011		0.000
p-value WHITE	0.000	0.042		0.000
p-value NW5	0.000	0.025		0.000

Table 5: Multivariate return regressions with interactions

This table presents OLS coefficient estimates for

$$
\begin{aligned}
r_t \quad = \quad & c + \sum_{i=1}^{25} \{\ \beta_{i,t-1} \cdot \mathit{Journalist}_{i,t-1} + \beta_{i,t} \cdot \mathit{Journalist}_{i,t} + \gamma_{i,t-1} \cdot r_{t-1} \mathit{Journalist}_{i,t-1} \\
& + \gamma_{i,t} \cdot r_{t-1} \mathit{Journalist}_{i,t}\ \} + \eta \cdot \mathit{Controls}_t + \epsilon_t .
\end{aligned}
$$

The *Controls* vector is as in Table 2. Coefficient estimates are recorded on the left-hand side of each column and their corresponding OLS t-statistics are presented in parentheses on the right-hand side. The table also presents the number of observations in each regression, the R-squared for an unreported regression with no journalist fixed effects, R^2 no JD, and the R-squared for the reported regression that includes the journalist fixed effects, $R^2 JD$. Also recorded are p-values from F-tests testing the following null hypotheses: $\beta_{i,t-1} = \beta_{i,t} = 0$, $\forall i$; $\gamma_{i,t-1} = \gamma_{i,t} = 0$, $\forall i$; and $\beta_{i,t-1} = \beta_{i,t} = \gamma_{i,t-1} = \gamma_{i,t} = 0$, $\forall i$. For each test, p-values are reported for F-tests calculated using the OLS variance/covariance matrix, a heteroscedasticity robust variance/covariance matrix (WHITE), and a Newey-West variance/covariance matrix using five lags (NW5). The t-statistics that are statistically significant at the 1% level are indicated by ***, at the 5% level by **, and at the 10% level by *.

	publication date		day after		publication date		day after	
	β_t	t-stat	β_{t-1}	t-stat	γ_t	t-stat	γ_{t-1}	t-stat
Hillery	0.024	(0.4)	0.002	(0.0)	-0.145^{***}	(-2.8)	0.214^{***}	(4.2)
O'Brien	-0.109^{*}	(-1.7)	0.196^{***}	(3.2)	-0.112^{**}	(-2.5)	0.054	(1.2)
Talley	0.014	(0.2)	0.040	(0.6)	-0.162^{***}	(-2.8)	0.015	(0.3)
Marcial	0.062	(0.8)	-0.098	(-1.3)	-0.026	(-0.4)	0.138^{**}	(2.1)
Garcia	0.132	(1.3)	-0.085	(-0.8)	-0.543^{***}	(-8.1)	0.679^{***}	(10.0)
Smith	-0.029	(-0.4)	0.142^{*}	(1.7)	-0.069	(-0.8)	0.127	(1.5)
Wilson	-0.007	(-0.1)	-0.136	(-1.5)	0.092	(1.1)	-0.166^{**}	(-2.2)
Browning	0.065	(0.8)	0.015	(0.2)	-0.088	(-1.2)	0.028	(0.4)
Pettit	-0.100	(-1.0)	0.101	(1.0)	0.145	(1.0)	-0.123	(-0.9)
Sease	0.005	(0.1)	0.048	(0.5)	0.182^{**}	(2.0)	-0.145	(-1.6)
Rosenberg	-0.042	(-0.4)	-0.270^{**}	(-2.3)	-0.231^{**}	(-2.0)	0.430^{***}	(3.9)
Kansas	0.032	(0.3)	0.027	(0.2)	-0.006	(0.0)	0.078	(0.4)
McLean	0.069	(0.5)	-0.012	(-0.1)	0.172	(1.3)	0.291^{**}	(2.1)
Raghavan	0.186	(1.4)	0.024	(0.2)	-0.051	(-0.2)	0.022	(0.1)
Ip	0.117	(1.0)	0.236^{*}	(1.9)	0.041	(0.4)	0.090	(0.8)
Gonzalez	-0.070	(-0.5)	0.191	(1.3)	-0.025	(-0.1)	0.115	(0.5)
Metz	0.093	(0.7)	-0.083	(-0.6)	-0.163	(-1.3)	0.288^{*}	(2.0)
Levingston	0.039	(0.3)	-0.063	(-0.5)	0.109	(0.6)	-0.155	(-0.9)
Pasha	0.475^{***}	(3.3)	-0.231	(-1.6)	-0.115	(-1.0)	-0.082	(-0.8)
Rose	0.287	(1.3)	-0.344	(-1.6)	-0.709^{***}	(-3.1)	0.493^{**}	(2.3)
Steiner	-0.142	(-0.9)	0.018	(0.1)	0.133	(0.6)	-0.136	(-0.9)
Dorfman	0.045	(0.3)	-0.321^{**}	(-2.2)	-0.055	(-0.4)	0.459^{***}	(3.0)
Bauman	-0.047	(-0.3)	-0.128	(-0.8)	0.203	(1.0)	0.227	(1.2)
McGee	0.370^{**}	(2.4)	0.732^{***}	(4.6)	-0.273^{**}	(-2.4)	-0.350^{***}	(-3.5)
Elia	-0.163	(-1.1)	0.056	(0.4)	0.040	(0.3)	0.144	(1.0)
Observations	9592							
R^2_{noJFE}	0.028							
R^2_{JFE}	0.061							
H_0 :	$\beta_{t-1} = \beta_t = 0$		$\gamma_{t-1} = \gamma_t = 0$		$\beta_{t-1} = \beta_t = \gamma_{t-1} = \gamma_t = 0$			
p-value OLS	0.000		0.000		0.000			
p-value WHITE	0.000		0.000		0.000			
p-value NW5	0.000		0.000		0.000			

Table 6: Linear probability models – forecasting journalists arrivals

This table presents R^2 for the following linear probability models:

$$\begin{aligned}
\text{Model 1} : \mathit{Journalist}_{j,t} &= c_j + \sum_{i=1}^{38} \lambda_i \cdot Y_{i,t} + \epsilon_{j,t} \\
\text{Model 2} : \mathit{Journalist}_{j,t} &= c_j + \sum_{i=1}^{38} \lambda_i \cdot Y_{i,t} + \eta \cdot \mathit{Market\ Variables}_t + \epsilon_{j,t} \\
\text{Model 3} : \mathit{Journalist}_{j,t} &= c_j + \sum_{i=1}^{38} \lambda_i \cdot Y_{i,t} + \eta \cdot \mathit{Market\ Variables}_t + \\
& \sum_{i=1}^{38} (\psi_i \cdot Y_{i,t} \times \mathit{Journalist}_{j,t-1} + \rho_i \cdot Y_{i,t} \times \mathit{Journalist}_{j,t-7} + \zeta_i \cdot D_t \times Y_{i,t}) + \epsilon_{j,t}
\end{aligned}$$

for all $j = 1 \ldots 25$. Y_i is a vector of year fixed effects for year i, while D_t is a matrix of day of the week dummies. *Market Variables* represents five lags of DJIA returns, five lags of volume, and five lags of DJIA squared residuals as previously defined. The variable $\mathit{Journalist}_{j,t-7}$ represents an indicator variable that equals 1 if journalist j wrote on the same day the previous week, and 0 otherwise. Each regression is run using only data for the period during which the corresponding journalist was actively writing. Column 5 presents the p-value for an F-test of $\mathrm{H}_0 : \eta = 0$, and Column 6 reports the OLS p-value for the F-test of $\mathrm{H}_0 : \psi_i = \rho_i = \zeta_i = 0$, $\forall i$.

	Model 1 R^2	Model 2 R^2	Model 3 R^2	$\mathrm{H}_0 : \eta = 0$ p-value	$\mathrm{H}_0 : \psi_i = \rho_i = \zeta_i = 0$ p-value
Hillery	0.038	0.062	0.389	0.592	0.000
O'Brien	0.028	0.088	0.429	0.434	0.000
Talley	0.287	0.318	0.618	0.847	0.000
Marcial	0.004	0.029	0.192	0.985	0.000
Garcia	0.103	0.118	0.476	0.735	0.000
Smith	0.273	0.281	0.508	0.965	0.000
Wilson	0.086	0.137	0.411	0.272	0.000
Browning	0.020	0.076	0.458	0.060	0.000
Pettit	0.072	0.117	0.320	0.675	0.000
Sease	0.081	0.125	0.446	0.002	0.000
Rosenberg	0.011	0.100	0.613	0.437	0.000
McLean	0.073	0.174	0.462	0.028	0.000
Kansas	0.023	0.111	0.586	0.355	0.000
Raghavan	0.050	0.081	0.174	0.855	0.001
Ip	0.085	0.102	0.350	0.990	0.000
Gonzalez	0.000	0.043	0.100	0.907	0.779
Metz	0.040	0.044	0.192	0.759	0.000
Levingston	0.122	0.176	0.349	0.864	0.000
Pasha	0.044	0.088	0.353	0.027	0.000
Rose	0.021	0.033	0.580	0.801	0.000
Dorfman	0.224	0.241	0.406	0.010	0.000
Steiner	0.008	0.029	0.355	0.465	0.000
Bauman	0.082	0.091	0.232	0.986	0.000
McGee	0.067	0.115	0.526	0.048	0.000
Elia	0.018	0.031	0.169	0.741	0.000

Table 7: Multivariate return regressions with IV

This table presents coefficient estimates for

$$r_t = c + \sum_{i=1}^{25} \{ \beta_{i,t-1} \cdot IVJournalist_{i,t-1} + \beta_{i,t} \cdot IVJournalist_{i,t} + \gamma_{i,t-1} \cdot r_{t-1} IVJournalist_{i,t-1} + \gamma_{i,t} \cdot r_{t-1} IVJournalist_{i,t} \} + \eta \cdot Controls_t + \epsilon_t.$$

where *IVJournalist* is defined as the fitted fitted values from the LPM model of Table 6, column 4. The *Controls* vector is as in Table 2. Coefficient estimates are recorded on the left-hand side of each column and their corresponding OLS t-statistics are presented in parentheses on the right-hand side. The table also presents the number of observations in each regression, the R-squared for an unreported regression with no journalist fixed effects, R^2 no JD, and the R-squared for the reported regression that includes the journalist fixed effects, $R^2 JD$. Also recorded are p-values from F-tests testing the following null hypotheses: $\beta_{i,t-1} = \beta_{i,t} = 0$, $\forall i$; $\gamma_{i,t-1} = \gamma_{i,t} = 0$, $\forall i$; and $\beta_{i,t-1} = \beta_{i,t} = \gamma_{i,t-1} = \gamma_{i,t} = 0$, $\forall i$. For each test, p-values are reported for F-tests calculated using the OLS variance/covariance matrix, a heteroscedasticity robust variance/covariance matrix (WHITE), and a Newey-West variance/covariance matrix using five lags (NW5). The t-statistics that are statistically significant at the 1% level are indicated by ***, at the 5% level by **, and at the 10% level by *.

	publication date		day after		publication date		day after	
	β_t	t-stat	β_{t-1}	t-stat	γ_t	t-stat	γ_{t-1}	t-stat
Hillery	−0.033	(−0.4)	0.058	(0.7)	−0.047	(−0.7)	0.170**	(2.4)
O'Brien	−0.194	(−1.6)	0.118	(1.0)	−0.067	(−0.8)	0.066	(0.8)
Talley	0.061	(0.5)	0.007	(0.1)	−0.140	(−1.5)	0.028	(0.3)
Marcial	0.092	(0.7)	−0.037	(−0.3)	0.091	(0.8)	0.156	(1.3)
Garcia	0.251	(1.6)	−0.155	(−1.0)	−0.411***	(−3.2)	0.648***	(5.1)
Smith	−0.222	(−1.4)	0.385**	(2.4)	0.067	(0.4)	0.154	(1.0)
Wilson	−0.156	(−1.1)	−0.066	(−0.4)	0.034	(0.3)	−0.026	(−0.2)
Browning	0.201	(1.1)	−0.079	(−0.5)	−0.193	(−1.4)	0.130	(1.0)
Pettit	−0.159	(−0.7)	0.164	(0.8)	0.057	(0.2)	−0.042	(−0.1)
Sease	−0.184	(−1.0)	0.294*	(1.7)	0.314**	(2.0)	−0.297**	(−2.0)
Rosenberg	−0.089	(−0.6)	−0.287*	(−1.9)	−0.109	(−0.8)	0.457***	(3.3)
McLean	−0.122	(−0.6)	0.069	(0.4)	0.029	(0.1)	0.193	(0.7)
Kansas	0.007	(0.0)	0.195	(1.2)	0.258*	(1.7)	0.213	(1.3)
Raghavan	0.288	(0.8)	0.052	(0.1)	0.063	(0.1)	0.035	(0.1)
Ip	−0.391	(−1.5)	0.196	(0.8)	0.537**	(2.4)	−0.019	(−0.1)
Gonzalez	−0.259	(−0.7)	0.175	(0.4)	0.754	(1.0)	−0.551	(−0.8)
Metz	−0.243	(−0.8)	0.062	(0.2)	0.021	(0.1)	−0.572**	(−2.1)
Levingston	0.030	(0.1)	−0.113	(−0.4)	−0.186	(−0.5)	−0.043	(−0.1)
Pasha	0.291	(1.1)	−0.484*	(−1.9)	−0.556***	(−2.9)	0.229	(1.2)
Rose	−0.183	(−0.7)	0.113	(0.4)	−0.736**	(−2.5)	0.569*	(1.9)
Dorfman	−0.077	(−0.3)	0.129	(0.4)	0.372	(1.1)	−0.303	(−0.9)
Steiner	0.029	(0.1)	−0.562**	(−2.3)	0.335	(1.3)	0.347	(1.5)
Bauman	−0.547	(−1.5)	0.156	(0.4)	0.733*	(1.8)	0.412	(1.0)
McGee	0.358	(1.4)	0.712***	(2.8)	−0.509*	(−1.8)	−0.601***	(−2.6)
Elia	−0.549	(−1.4)	−0.434	(−1.1)	0.588	(1.6)	−0.579*	(−1.9)
Observations	9592							
R^2_{noJFE}	0.028							
R^2_{JFE}	0.062							
H_0 :	$\beta_{t-1} = \beta_t = 0$		$\gamma_{t-1} = \gamma_t = 0$		$\beta_{t-1} = \beta_t = \gamma_{t-1} = \gamma_t = 0$			
p-value OLS	0.020		0.000		0.000			
p-value WHITE	0.001		0.000		0.000			
p-value NW5	0.000		0.002		0.000			

Table 8: Multivariate return regressions

This table presents F-test p-values for the same regression reported in Table 4 only in this instance using different return series. In particular, the *Open-Close* column uses DJIA open-to-close daily excess returns, the *S&P 500* column uses daily S&P 500 excess returns, the *CRSP VWTD* column uses daily CRSP value-weighted excess returns, and the *GARCH-Adj.* column uses GARCH-adjusted returns which are defined as DJIA close-to-close returns divided by the estimated daily volatility from a GARCH(1,1) model estimated on the same return series. In addition, to different return series the last two columns also use slightly different regressors. The results in the *Winsorized* column use returns, volume, and all non-dummy variables that are winsorized at the 5 percent level, and the results in the Ten Authors column use only journalist indicators for the ten most prolific writers in our sample.

	Open-Close	S&P 500	CRSP VWTD	GARCH-Adj.	Winsorized	Ten Authors
Panel A: $\beta_{t-1} = 0$						
p-value OLS	0.000	0.000	0.000	0.000	0.000	0.010
p-value WHITE	0.000	0.000	0.000	0.000	0.000	0.005
p-value NW5	0.000	0.000	0.000	0.000	0.000	0.005
Panel B: $\beta_t = 0$						
p-value OLS	0.015	0.046	0.097	0.033	0.027	0.089
p-value WHITE	0.052	0.214	0.264	0.046	0.012	0.114
p-value NW5	0.034	0.172	0.183	0.035	0.008	0.101
Panel C: $\beta_{t-1} = \beta_t = 0$						
p-value OLS	0.000	0.000	0.001	0.000	0.000	0.004
p-value WHITE	0.000	0.000	0.000	0.000	0.000	0.004
p-value NW5	0.000	0.000	0.000	0.000	0.000	0.003

Table 9: Multivariate return regressions with interactions

This table presents F-test p-values for the same regression reported in Table 5 only in this instance using different return series. In particular, the *Open-Close* column uses DJIA open-to-close daily excess returns, the *S&P 500* column uses daily S&P 500 excess returns, the *CRSP VWTD* column uses daily CRSP value-weighted excess returns, and the *GARCH-Adj.* column uses GARCH-adjusted returns which are defined as DJIA close-to-close returns divided by the estimated daily volatility from a GARCH(1,1) model estimated on the same return series. In addition, to different return series the last two columns also use slightly different regressors. The results in the *Winsorized* column use returns, volume, and all non-dummy variables that are winsorized at the 5 percent level, and the results in the Ten Authors column use only journalist indicators for the ten most prolific writers in our sample.

	Open-Close	S&P 500	CRSP VWTD	GARCH-Adj.	Winsorized	Ten Authors
Panel A: $\beta_{t-1} = \beta_t = 0$						
p-value OLS	0.000	0.000	0.001	0.000	0.000	0.013
p-value WHITE	0.000	0.000	0.000	0.000	0.000	0.008
p-value NW5	0.000	0.007	0.000	0.000	0.000	0.007
Panel B: $\gamma_{t-1} = \gamma_t = 0$						
p-value OLS	0.000	0.000	0.000	0.000	0.000	0.000
p-value WHITE	0.000	0.000	0.000	0.000	0.000	0.000
p-value NW5	0.000	0.000	0.000	0.000	0.000	0.000
Panel C: $\beta_{t-1} = \beta_t = \gamma_{t-1} = \gamma_t = 0$						
p-value OLS	0.000	0.000	0.000	0.000	0.000	0.000
p-value WHITE	0.000	0.000	0.000	0.000	0.000	0.000
p-value NW5	0.000	0.000	0.000	0.000	0.000	0.000

Table 10: Multivariate return regression and falsification

This table presents coefficient estimates for

$$r_t = c + \sum_{i=0}^{25} \{ \beta_{i,t-2} \cdot Journalist_{i,t-2} + \beta_{i,t-1} \cdot Journalist_{i,t-1} + \beta_{i,t} \cdot Journalist_{i,t} + \beta_{i,t+1} \cdot Journalist_{i,t+1} + \beta_{i,t+2} \cdot Journalist_{i,t+2} \} + \eta \cdot Controls_t + \epsilon_t$$

The *Controls* vector is as in Table 2. Coefficient estimates are recorded on the left-hand side of each column and their corresponding t-statistics are presented in parentheses on the right-hand side. Also recorded are p-values from F-tests of the following null hypotheses: $\beta_{i,t+k} = 0$, $\forall i$ and $k \in \{-2, -1, 0, 1, 2\}$. For each test, p-values are reported for F-tests calculated using the OLS variance/covariance matrix, a heteroscedasticity robust variance/covariance matrix (WHITE), and a Newey-West variance/covariance matrix using five lags (NW5). t-statistics that are statistically significant at the 1% level are indicated by ***, at the 5% level by **, and at the 10% level by *.

	2 days before		1 day before		publication date		1 day after		2 days after	
	$\beta_{i,t+2}$		$\beta_{i,t+1}$		$\beta_{i,t}$		$\beta_{i,t-1}$		$\beta_{i,t-2}$	
Hillery	0.113*	(1.7)	−0.028	(−0.4)	−0.005	(−0.1)	−0.057	(−0.9)	0.132**	(2.0)
O'Brien	0.005	(0.1)	0.122*	(1.9)	−0.130**	(−2.0)	0.229***	(3.6)	−0.026	(−0.4)
Talley	0.009	(0.1)	0.103	(1.5)	−0.013	(−0.2)	0.033	(0.5)	0.069	(1.0)
Marcial	0.101	(1.3)	0.019	(0.2)	0.033	(0.4)	−0.136*	(−1.7)	0.074	(0.9)
Garcia	0.039	(0.4)	0.020	(0.2)	0.226*	(1.9)	−0.058	(−0.5)	−0.142	(−1.3)
Smith	−0.171**	(−2.0)	0.114	(1.3)	−0.013	(−0.1)	0.129	(1.5)	−0.008	(−0.1)
Wilson	0.079	(0.9)	0.052	(0.5)	−0.064	(−0.7)	−0.118	(−1.2)	−0.070	(−0.8)
Browning	−0.043	(−0.5)	0.041	(0.5)	0.024	(0.3)	−0.012	(−0.1)	0.002	(0.0)
Pettit	−0.031	(−0.3)	0.015	(0.1)	−0.094	(−0.9)	0.105	(1.0)	0.031	(0.3)
Sease	−0.112	(−1.2)	0.096	(1.0)	0.027	(0.3)	0.068	(0.7)	0.044	(0.5)
Rosenberg	0.218*	(1.8)	0.035	(0.3)	−0.113	(−0.9)	−0.355***	(−3.0)	0.168	(1.4)
Kansas	−0.026	(−0.2)	0.180	(1.5)	0.008	(0.1)	0.045	(0.4)	−0.073	(−0.6)
McLean	0.209	(1.6)	−0.039	(−0.3)	0.032	(0.2)	−0.068	(−0.5)	0.058	(0.4)
Raghavan	−0.162	(−1.2)	0.146	(1.1)	0.177	(1.3)	0.046	(0.3)	0.023	(0.2)
Ip	−0.172	(−1.4)	−0.018	(−0.1)	0.088	(0.7)	0.248**	(2.0)	−0.323***	(−2.6)
Gonzalez	−0.006	(0.0)	0.069	(0.5)	−0.083	(−0.6)	0.219	(1.5)	−0.026	(−0.2)
Metz	0.189	(1.4)	0.220*	(1.7)	−0.003	(0.0)	−0.113	(−0.9)	−0.064	(−0.5)
Levingston	−0.110	(−0.8)	0.042	(0.3)	0.059	(0.4)	−0.081	(−0.6)	0.064	(0.5)
Pasha	0.088	(0.6)	0.097	(0.6)	0.452***	(2.9)	−0.257	(−1.6)	−0.008	(−0.1)
Rose	0.094	(0.4)	0.014	(0.1)	0.270	(1.0)	−0.354	(−1.3)	−0.001	(0.0)
Steiner	−0.005	(0.0)	0.181	(1.1)	−0.164	(−1.0)	−0.008	(0.0)	−0.029	(−0.2)
Dorfman	0.098	(0.7)	−0.064	(−0.4)	0.022	(0.1)	−0.352**	(−2.4)	0.041	(0.3)
Bauman	−0.035	(−0.2)	0.131	(0.8)	−0.017	(−0.1)	−0.145	(−0.9)	0.057	(0.4)
McGee	−0.264*	(−1.7)	−0.057	(−0.4)	0.303*	(1.9)	0.461***	(2.9)	−0.188	(−1.2)
Elia	−0.008	(0.0)	0.173	(1.1)	−0.142	(−0.9)	0.033	(0.2)	0.011	(0.1)
Observations	9592									
R^2_{noJFE}	0.028									
R^2_{JFE}	0.044									
p-value OLS	0.814		0.842		0.077		0.001		0.672	
p-value WHITE	0.671		0.583		0.153		0.000		0.541	
p-value NW5	0.659		0.519		0.115		0.000		0.580	