Internet Appendix for "Sentiment During Recessions"

Diego García*

In this appendix, we provide additional results to supplement the evidence included in the published version of the paper. In Section 1, we give a description of the data that the paper uses, which is available from the *Journal of Finance* webpage. In Section 2, we include the full output to the main table of the paper (Table 4 in the paper), as well as three other specifications (one setting the lag-operator to $s = 1$, one changing the recession indicator variable for a variable based on volatility, and one giving the output used in Figure 4 of the paper). Section 3 considers two further tests: one looks at "hard news," proxied by the amount of numbers (figures) included in the news article, while the other interacts media content on a given day with lagged media content.

1 The Data Set

The data set used in the paper is constructed using the New York Times Article Archive. In essense, it consists of financial columns from the New York Times (NYT) for the 1905 to 2005 period. The bulk of the sample is constructed using two columns, labelled for most of the 20th century by "Financial Markets" and "Topics in Wall Street." By manually choosing articles, using their titles as coded in the electronic version of the NYT Article Archive, we collect a database with a total of 55,307 articles. In addition to the two mentioned columns, it includes their predecessors/successors, labeled "Along the Highways of Finance," "The Financial Week," "The Financial Situation," "News, Comment and Incident on the Stock Exchange," "Sidelights on the Financial and Business Developments of the Day," and "Market Place." For the period August 30, 1978 through December 1, 1978, 128 articles from the Washington Post are used (the New York Times was not printed due to a union strike).

The file `DataFinNewsNYT.csv` has a total of five columns: trading date, total number of words, number of negative words, number of positive words, and number of figures (strings

*Citation format: García, Diego, Internet Appendix for "Sentiment During Recessions," *Journal of Finance* [DOI string]. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the author of the article.

containing a numeric character) in the article. For each trading date, the file gives the aggregate of number of words, positive words, etc. for all articles published between the date following the previous trading day to (and including) the given trading date (see the published paper for details). The words and numbers (strings with a number character) are computed using the output from ABBYY (software capable of large-scale OCR tasks) of originally scanned copies of the articles (from the NYT Article Archive). Positive and negative words are as defined in Loughran and McDonald (2011)

2 Other Specifications

Table IA.1 considers the main model in the paper:

$$\begin{aligned} R_t = (1 - D_t) \left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2) \right) \\ + D_t \left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2) \right) + \boldsymbol{\eta} X_t + \epsilon_t, \end{aligned} \tag{IA.1}$$

where D_t is a dummy that takes the value one if and only if date t is during a recession (as defined by the NBER). Table IA.1 gives the full estimates of the model — the rows related to the media variables, the parameters β, are those reported in Table 4 of the paper.

Table IA.2 considers the model in (IA.1), using NBER recessions to define D_t, but letting the lag operator on the media variables have $s = 1$. It mimics Table 4 in the paper, setting $s = 1$ instead of $s = 5$. The econometric specification in the main body of the paper is robust to different types of right-hand-side perturbations.

Table IA.3 gives the point estimates reported in Figure 4 of the paper. In particular, it studies the specification

$$R_t = \boldsymbol{\beta} \mathcal{L}_s(M_t) + \boldsymbol{\gamma} \mathcal{L}_s(R_t) + \boldsymbol{\psi} \mathcal{L}_s(R_t^2) + \boldsymbol{\eta} X_t + \epsilon_t \tag{IA.2}$$

within each business cycle. The pessimism factor loads with a negative sign on 18 of the 20 recessions in our sample, whereas it has negative sign on 17 out of the 21 expansionary periods. The effect of positive words is particularly strong during the Great Depression (1929:08 to 1933:02 and 1937:05 to 1938:05), but point estimates are also large in the second half of the sample (i.e., 1957:08 to 1958:03, 1969:12 to 1970:10). Overall, the evidence in Table IA.3 suggests that the predictability of stock returns using media content is strong throughout both the first and the second halves of our sample.

Table IA.4 considers the model in (IA.1), where instead of using NBER recessions for D_t, we define D_t as a dummy that takes the value one if and only if the estimated volatility from a GARCH (1,1) model is in its upper quartile. While correlated with the business cycle, the volatility estimates allow us to see if it is times of uncertainty per se, rather than economic conditions, that drive our results. The results reported in Table IA.4 suggest that NBER recessions are the periods in which the media has more bite. In particular, we see that the point estimates on the leading terms during periods of low volatility (Panel A) are lower than those during high volatility (Panel B). But the differences are not large enough in economic or statistical terms.

3 Further Tests

Our next test focuses on the nature of the information in the columns. Clearly, many investors learned about firms and their decisions via the press during our time period, since there were no other important media outlets. Financial data itself seem to be the most relevant for investors — in the columns under study the NYT regularly published tables with dividends, stock prices, earnings, even discussions on single figures. We divide the NYT articles based on the number of figures, that is, the fraction of words that contain numbers.[1] The underlying idea is that articles with more "hard data" are more likely to contain information that is relevant for investors.

The NYT columns under study underwent significant changes over the years: some years they would include tables with stock prices, whereas other years they would not include the tables as part of the column itself.[2] To classify our columns in terms of the amount of hard figures they provided, we first estimate the following model:

$$N_t = \boldsymbol{\beta} Y_t + \boldsymbol{\eta} X_t + \epsilon_t, \tag{IA.3}$$

where N_t denotes the fraction of words that contain numbers, Y_t is a matrix of year-month indicator variables, and the exogenous variables X_t include day-of-the-week dummies. We use the residuals from the above model as a proxy for the amount of hard information provided by the columns. This allows us to control for any patterns in the format of the columns through time and/or weekly conventions.

More specifically, we define an indicator variable I_t that takes the value one if and only if

[1] We should note that during this time period stock prices, interest rates, dividends, etc. were given as fractions. The OCR software struggles with the way the NYT printed such fractions, so the text of the pdf images will typically have numbers and letters together. Clearly, in such instances the original text meant to refer to a number. We define a "number" as any string that contains any of the ASCII characters 0 to 9.

[2] It should be noted that this heterogeneity is part of the classification of news that the NYT undertook when digitizing the newspaper. "Cutting up" news stories when there are hanging tables in the newspaper (i.e., above or below the actual column) clearly calls for some human judgement.

the estimated residuals from (IA.3) are positive. We then estimate the model

$$\begin{aligned}R_t = (1-I_t)\left[(1-D_t)\left(\boldsymbol{\beta}_1\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_1\mathcal{L}_s(R_t)+\boldsymbol{\psi}_1\mathcal{L}_s(R_t^2)\right)+D_t\left(\boldsymbol{\beta}_2\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_2\mathcal{L}_s(R_t)+\boldsymbol{\psi}_2\mathcal{L}_s(R_t^2)\right)\right]\\ +I_t\left[(1-D_t)\left(\boldsymbol{\beta}_3\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_3\mathcal{L}_s(R_t)+\boldsymbol{\psi}_3\mathcal{L}_s(R_t^2)\right)+D_t\left(\boldsymbol{\beta}_4\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_4\mathcal{L}_s(R_t)+\boldsymbol{\psi}_4\mathcal{L}_s(R_t^2)\right)\right]\\ +\boldsymbol{\eta}X_t+\epsilon_t, \qquad \text{(IA.4)}\end{aligned}$$

which boils down to (IA.1) with all the coefficients interacted with the figures indicator I_t. Table IA.5 gives the estimates of the leading coefficients on the media variables. In expansions we find that there is an effect irrespective of the amount of hard figures. The pessimism factor, for example, loads significantly with a coefficient of -3.1 basis points, whereas the coefficient on days low on figures is only -3.8 basis points. During recessions we also find no significant differences among articles with high or low figure counts. On days that the financial columns lack hard data, the pessimism factor moves the DJIA by -10.9 basis points versus -12.7 basis points on days with more figures. We conclude that hard data do not drive the relationship between our media content measures and stock returns. While this rules out some theories based on information, it could very well be the case that hard information and soft information are actually substitutes, or independent from each other, and that our word counts contain facts that investors could not infer via the information contained in earnings or dividend announcements.

The interpretation of media content as sentiment changes, implicit in Tetlock (2007) and in our paper, could depend on the past few sets of information signals. Informational models could certainly generate such implications as well, in which the value of a signal depends on some underlying variable correlated with past signal realizations.

Table IA.6 further interacts the media content measure written prior to trading with different metrics of past market performance. It estimates a statistical model as in (IA.4), where the indicator variable I_t now takes the value one if and only if prior lagged media content (over one, four, or 19 days) was particularly large. Our model captures in a parsimonious fashion the interactions of lagged news on their future impact on stock returns.

In particular, we estimate the model in (IA.4), where $I_t = 1$ if and only if $\bar{M}_{t-1} \geq ks(\bar{M}_{t-1})$, with $s(Y_t)$ denoting the standard deviation of the process Y_t. In Panel A of Table IA.6 we use $\bar{M}_{t-1} = M_{t-2}$, whereas in Panels B and C we use, respectively, $\bar{M}_{t-1} = \sum_{s=t-5}^{t-2} M_s/4$ (Panel B), and $\bar{M}_{t-1} = \sum_{s=t-20}^{t-2} M_s/19$ (Panel C). The test discriminates days that were preceded by particularly good (bad) news, that is, news of the flavor "the economy is going to hell in a hand basket again," versus days in which the tone was not particularly positive (negative) the previous trading day. In the table, we set $k = 1/2$, which given the normalization of the media metrics boils down to requiring that $\bar{M}_{t-1}$ is one-half of a standard deviation above its unconditional mean. Other normalizations yield similar results.

Table IA.6 shows that there is no difference between days with previously negative or positive news, at least in terms of the impact of news on same-day stock prices. The point estimates are virtually the same between "extreme days" (for which news were particularly negative/positive on previous days) and "regular days."

References

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Table IA.1
Feedback from News Content to the DJIA along the Business Cycle

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = (1 - D_t)\left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t.$$

All independent variables are as defined in Table III of the paper. The dependent variable R_t is the log-return on the DJIA from 1905 to 2005. The sample period comprises 27,449 trading days, of which 6,467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Panel A. Media variables ($\boldsymbol{\beta}$)						
$(1 - D_t) \times M_{t-1}$	0.024	3.3	−0.028	−3.5	−0.035	−4.2
$(1 - D_t) \times M_{t-2}$	0.004	0.6	0.004	0.5	0.001	0.1
$(1 - D_t) \times M_{t-3}$	−0.004	−0.6	0.005	0.7	0.006	0.8
$(1 - D_t) \times M_{t-4}$	−0.012	−1.7	0.006	0.8	0.011	1.5
$(1 - D_t) \times M_{t-5}$	−0.004	−0.6	0.006	0.8	0.007	0.9
$D_t \times M_{t-1}$	0.085	3.9	−0.087	−3.4	−0.117	−4.4
$D_t \times M_{t-2}$	0.004	0.2	−0.005	−0.2	−0.004	−0.2
$D_t \times M_{t-3}$	−0.021	−1.0	0.010	0.4	0.020	0.8
$D_t \times M_{t-4}$	−0.009	−0.4	0.016	0.7	0.019	0.8
$D_t \times M_{t-5}$	−0.005	−0.2	0.028	1.2	0.026	1.1
Panel B. Lagged returns ($\boldsymbol{\gamma}$)						
$(1 - D_t) \times R_{t-1}$	0.053	3.7	0.052	3.6	0.046	3.1
$(1 - D_t) \times R_{t-2}$	−0.035	−2.6	−0.034	−2.4	−0.036	−2.5
$(1 - D_t) \times R_{t-3}$	0.011	0.8	0.010	0.8	0.011	0.8
$(1 - D_t) \times R_{t-4}$	0.020	1.7	0.018	1.6	0.020	1.7
$(1 - D_t) \times R_{t-5}$	0.004	0.4	0.005	0.4	0.006	0.5
$D_t \times R_{t-1}$	−0.003	−0.1	−0.005	−0.2	−0.017	−0.5
$D_t \times R_{t-2}$	−0.016	−0.6	−0.017	−0.7	−0.021	−0.8
$D_t \times R_{t-3}$	0.026	1.1	0.023	0.9	0.024	1.0
$D_t \times R_{t-4}$	0.076	3.0	0.075	3.0	0.076	2.9
$D_t \times R_{t-5}$	0.025	1.0	0.031	1.2	0.031	1.2
Panel C. Lagged squared-returns ($\boldsymbol{\psi}$)						
$(1 - D_t) \times R^2_{t-1}$	0.010	5.2	0.010	5.2	0.010	5.1
$(1 - D_t) \times R^2_{t-2}$	0.010	1.3	0.010	1.3	0.010	1.3
$(1 - D_t) \times R^2_{t-3}$	−0.005	−0.8	−0.006	−0.8	−0.005	−0.8
$(1 - D_t) \times R^2_{t-4}$	0.002	0.4	0.002	0.4	0.002	0.4
$(1 - D_t) \times R^2_{t-5}$	−0.010	−1.5	−0.010	−1.5	−0.010	−1.5
$D_t \times R^2_{t-1}$	0.004	0.3	0.005	0.4	0.005	0.4
$D_t \times R^2_{t-2}$	0.024	2.7	0.024	2.7	0.024	2.7
$D_t \times R^2_{t-3}$	0.004	0.6	0.004	0.6	0.004	0.6
$D_t \times R^2_{t-4}$	−0.016	−2.1	−0.016	−2.1	−0.016	−2.1
$D_t \times R^2_{t-5}$	−0.017	−2.5	−0.018	−2.5	−0.018	−2.5
Panel D. Day-of-the-week and recession dummies ($\boldsymbol{\eta}$)						
I_{Tue}	0.145	6.5	0.137	6.1	0.140	6.3
I_{Wed}	0.151	6.7	0.143	6.3	0.144	6.3
I_{Thu}	0.125	5.5	0.116	5.1	0.116	5.2
I_{Fri}	0.160	7.1	0.152	6.7	0.152	6.7
I_{Sat}	0.182	7.0	0.176	6.7	0.174	6.6
D_t	−0.075	−2.9	−0.081	−3.0	−0.078	−3.1

Table IA.2
Feedback from News Content to the DJIA along the Business Cycle

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = (1 - D_t)\left(\beta_1 M_{t-1} + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\beta_2 M_{t-1} + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t,$$

where the lag-operator $\mathcal{L}_s$ for the media variable is set at $s = 1$. It differs with respect to Table IV in the paper only along this respect (in the paper $s = 5$).

Panel A. Expansions ($\boldsymbol{\beta}_1$)

	Positive		Negative		Pessimism	
$(1 - D_t) \times M_{t-1}$	0.022	3.1	−0.025	−3.2	−0.032	−3.8

Panel B. Recessions ($\boldsymbol{\beta}_2$)

	Positive		Negative		Pessimism	
$D_t \times M_{t-1}$	0.081	3.7	−0.079	−3.2	−0.107	−4.1

Panel C. Tests

	Positive		Negative		Pessimism	
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	7.2	0.007	5.0	0.025	8.6	0.003

Table IA.3
Media Sentiment by Business Cycle

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = \boldsymbol{\beta}\mathcal{L}_s(M_t) + \boldsymbol{\gamma}\mathcal{L}_s(R_t) + \boldsymbol{\psi}\mathcal{L}_s(R_t^2) + \boldsymbol{\eta}X_t + \epsilon_t.$$

It corresponds to an estimation of Table 3 in the paper for each business-cycle.

			Positive		Negative		Pessimism	
Start date	End date	N	β	t-stat	β	t-stat	β	t-stat
Panel A. Expansions								
190408	190704	703	0.030	0.6	−0.052	−1.0	−0.068	−1.1
190806	190912	473	0.016	0.3	0.051	0.9	0.036	0.6
191201	191212	301	−0.017	−0.5	−0.013	−0.2	−0.014	−0.3
191412	191807	1090	0.000	0.0	−0.032	−0.7	−0.025	−0.5
191903	191912	245	−0.015	−0.1	−0.209	−1.3	−0.136	−0.9
192107	192304	548	0.029	0.6	0.015	0.3	−0.009	−0.2
192407	192609	677	−0.017	−0.4	−0.099	−1.8	−0.055	−1.1
192711	192907	518	0.125	2.0	0.025	0.3	−0.068	−0.9
193303	193704	1241	−0.165	−2.5	0.042	0.6	0.132	1.6
193806	194501	2006	−0.006	−0.3	−0.046	−1.9	−0.031	−1.2
194510	194810	869	0.057	2.2	−0.073	−2.0	−0.088	−2.5
194910	195306	1033	0.027	1.2	0.017	0.7	−0.000	−0.0
195405	195707	819	0.026	0.9	−0.061	−1.6	−0.065	−1.8
195804	196003	505	0.053	1.6	0.028	0.6	−0.025	−0.6
196102	196911	2194	0.035	1.9	−0.039	−1.6	−0.054	−2.0
197011	197310	757	0.034	1.1	0.054	1.5	0.020	0.6
197503	197912	1222	0.040	1.8	−0.003	−0.1	−0.022	−0.9
198007	198106	252	0.052	0.9	−0.021	−0.3	−0.056	−0.8
198211	199006	1938	0.015	0.6	−0.032	−1.2	−0.032	−1.2
199103	200102	2526	0.028	1.4	−0.045	−2.6	−0.051	−2.7
200111	200512	1050	−0.030	−1.0	0.002	0.1	0.015	0.5
Panel B. Recessions								
190705	190805	327	0.029	0.4	−0.073	−0.7	−0.054	−0.5
191001	191112	595	0.000	0.0	−0.058	−1.0	−0.060	−1.0
191301	191411	473	0.010	0.2	−0.080	−1.4	−0.070	−1.2
191808	191902	172	−0.123	−1.2	−0.048	−0.5	0.029	0.2
192001	192106	448	0.049	0.6	−0.158	−1.9	−0.155	−1.7
192305	192406	351	0.112	1.4	−0.076	−0.9	−0.110	−1.3
192610	192710	326	−0.018	−0.3	−0.103	−1.2	−0.060	−0.7
192908	193302	1064	0.138	1.4	−0.005	−0.0	−0.091	−0.8
193705	193805	324	0.257	1.5	0.066	0.4	−0.089	−0.6
194502	194509	189	0.069	1.4	−0.055	−1.0	−0.076	−1.4
194811	194909	258	0.008	0.2	−0.047	−1.0	−0.047	−0.8
195307	195404	209	0.012	0.3	−0.030	−0.5	−0.018	−0.3
195708	195803	168	0.228	2.0	−0.116	−1.1	−0.245	−1.9
196004	196101	210	0.040	0.7	−0.065	−0.9	−0.083	−1.2
196912	197010	234	0.209	2.9	−0.065	−1.0	−0.180	−2.3
197311	197502	335	0.083	0.8	0.114	1.2	0.047	0.5
198001	198006	126	0.053	0.5	−0.227	−2.0	−0.190	−1.5
198107	198210	338	0.008	0.2	−0.030	−0.5	−0.031	−0.5
199007	199102	168	0.048	0.4	−0.145	−1.4	−0.156	−1.4
200103	200110	167	0.011	0.1	−0.029	−0.3	−0.027	−0.2

Table IA.4
Volatility and News

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = (1 - D_t)\left(\boldsymbol{\beta}_1 M_{t-1} + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_2 M_{t-1} + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t;$$

where D_t is a dummy variable taking on the value 1 if and only if the estimated volatility from a GARCH(1,1) process on the log-returns of the DJIA is in its upper quartile. It essentially reproduces Table 4 of the paper where the interaction dummy is high-volatility periods instead of NBER recessions.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Panel A. Low volatility ($\boldsymbol{\beta}_1$)						
$(1 - D_t) \times M_{t-1}$	0.026	4.6	−0.031	−5.0	−0.040	−6.3
$(1 - D_t) \times M_{t-2}$	−0.007	−1.2	0.005	0.8	0.008	1.2
$(1 - D_t) \times M_{t-3}$	−0.000	−0.0	0.001	0.2	0.001	0.1
$(1 - D_t) \times M_{t-4}$	−0.003	−0.5	0.002	0.3	0.004	0.6
$(1 - D_t) \times M_{t-5}$	0.001	0.2	−0.001	−0.1	−0.001	−0.1
Panel B. High volatility ($\boldsymbol{\beta}_2$)						
$D_t \times M_{t-1}$	0.051	2.2	−0.045	−1.9	−0.064	−2.6
$D_t \times M_{t-2}$	0.026	1.2	0.008	0.3	−0.006	−0.3
$D_t \times M_{t-3}$	−0.028	−1.3	0.014	0.6	0.025	1.1
$D_t \times M_{t-4}$	−0.039	−1.7	0.017	0.7	0.032	1.4
$D_t \times M_{t-5}$	−0.023	−1.0	0.039	1.9	0.042	1.9
Panel C. Tests						
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\beta_{11} = \beta_{21}$	1.2	0.278	0.4	0.549	0.9	0.342
Test $\sum_{j=2}^{5} \beta_{1j} = 0$	0.8	0.382	0.6	0.452	1.2	0.274
Test $\sum_{j=2}^{5} \beta_{2j} = 0$	2.5	0.114	4.5	0.034	5.8	0.016

Table IA.5
Hard News and the Effect of Media Content on Stock Returns

The table reports the estimated coefficients β from the model

$$\begin{aligned} R_t \;=\;& I_t \left[(1-D_t)\left(\boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\beta}_1 \mathcal{L}_s(M_t)\right) + D_t \left(\boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\beta}_2 \mathcal{L}_s(M_t)\right)\right] \\ & + (1-I_t)\left[(1-D_t)\left(\boldsymbol{\gamma}_3 \mathcal{L}_s(R_t) + \boldsymbol{\beta}_3 \mathcal{L}_s(M_t)\right) + D_t \left(\boldsymbol{\gamma}_4 \mathcal{L}_s(R_t) + \boldsymbol{\beta}_4 \mathcal{L}_s(M_t)\right)\right] + \boldsymbol{\eta} X_t + \epsilon_t, \end{aligned}$$

where $\mathcal{L}_s$ denotes an s-lag operator, namely $\mathcal{L}_s(R_t) = \{R_{t-1}, \ldots, R_{t-s}\}$, and D_t is a dummy variable taking on the value 1 if and only if date t is during a recession. The dependent variable R_t is the log-return on the DJIA index from 1905–2005. The variable M_t denotes one of our media measures described in Table 4 in the paper. In Panel A the variable I_t is an indicator variable that takes on the value one if and only if the news on date t contain more figures (numbers) than the average date in the month, controlling for possible day of the week patterns. In Panel B, the dummy variable I_t takes on the value one if and only if date $t-1$ was not a trading date. As the set of exogeneous variables X_t we include a constant term, day-of-the-week dummies, and a dummy for whether date t belongs to a recession or an expansion. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Expansion, high-information, β_{11}	0.020	1.9	−0.026	−2.1	−0.031	−2.6
Expansion, low-information, β_{21}	0.026	3.0	−0.029	−3.0	−0.038	−3.8
Recession, high-information, β_{31}	0.098	3.1	−0.094	−2.6	−0.127	−3.4
Recession, low-information, β_{41}	0.075	2.9	−0.082	−2.8	−0.109	−3.7
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	0.3	0.611	0.1	0.819	0.2	0.644
$\beta_{31} = \beta_{41}$	0.4	0.532	0.1	0.779	0.2	0.662
$\beta_{11} = \beta_{31}$	5.7	0.017	3.1	0.076	6.1	0.014
$\beta_{21} = \beta_{41}$	3.2	0.075	2.9	0.086	5.2	0.022

Table IA.6
Repeated Media Content Days and Stock Returns

The table reports the estimated coefficients β from the model

$$\begin{aligned} R_t = (1 - I_t) \left[(1 - D_t) \left(\boldsymbol{\beta}_1 M_{t-1} + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2) \right) + D_t \left(\boldsymbol{\beta}_3 M_{t-1} + \boldsymbol{\gamma}_3 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_3 \mathcal{L}_s(R_t^2) \right) \right] \\ + I_t \left[(1 - D_t) \left(\boldsymbol{\beta}_2 M_{t-1} + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2) \right) + D_t \left(\boldsymbol{\beta}_4 M_{t-1} + \boldsymbol{\gamma}_4 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_4 \mathcal{L}_s(R_t^2) \right) \right] \\ + \boldsymbol{\eta} X_t + \epsilon_t, \end{aligned}$$

where $I_t = 1$ if and only if $\bar{M}_{t-1} \geq ks(\bar{M}_{t-1})$, with $s(Y_t)$ denoting the standard deviation of the process Y_t. In Panel A of Table IA.6 we use $\bar{M}_{t-1} = M_{t-2}$, whereas in Panels B and C we use, respectively, $\bar{M}_{t-1} = \sum_{s=t-5}^{t-2} M_s/4$ (Panel B), and $\bar{M}_{t-1} = \sum_{s=t-20}^{t-2} M_s/19$ (Panel C). We set $k = 1/2$ throughout. All other independent variables are defined as in Table IV in the paper. The dependent variable R_t is the log-return on the DJIA index from 1905 to 2005. The sample period comprises 27,449 trading days, of which 6,467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

Panel A. Conditioning variable $\bar{M}_{t-1} = M_{t-2}$

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Expansion, extreme day, β_1	0.024	2.3	−0.028	−1.9	−0.024	−1.6
Expansion, regular day, β_2	0.021	2.5	−0.021	−2.4	−0.034	−3.6
Recession, extreme day, β_3	0.076	2.2	−0.046	−1.1	−0.101	−2.4
Recession, regular day, β_4	0.083	3.3	−0.094	−3.4	−0.108	−3.9

Panel B. Conditioning variable $\bar{M}_{t-1} = \sum_{s=t-5}^{t-2} M_s/4$

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Expansion, extreme day, β_1	0.020	2.0	−0.017	−1.2	−0.034	−2.2
Expansion, regular day, β_2	0.023	2.7	−0.027	−3.2	−0.029	−3.2
Recession, extreme day, β_3	0.118	3.7	−0.078	−1.8	−0.072	−1.6
Recession, regular day, β_4	0.067	2.6	−0.078	−3.1	−0.127	−5.1

Panel C. Conditioning variable $\bar{M}_{t-1} = \sum_{s=t-20}^{t-2} M_s/19$

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Expansion, extreme day, β_1	0.032	3.2	−0.025	−1.8	−0.017	−1.1
Expansion, regular day, β_2	0.016	1.8	−0.023	−2.6	−0.038	−4.4
Recession, extreme day, β_3	0.107	3.1	−0.066	−1.5	−0.103	−2.4
Recession, regular day, β_4	0.074	2.9	−0.086	−3.6	−0.108	−4.3