Media and the stock market

Diego García
UNC at Chapel Hill

Serbian Central Bank
Belgrade, June 31st, 2011

Media as a financial intermediary

- Shiller (2000)'s account of the media: hyping, tagging-along, entertainment at best.
  - "The history of speculative bubbles begins roughly with the advent of newspapers" (Shiller, 2000).
- Tetlock(2007)'s study of "Abreast of the market" column, 1984–1999: negative word counts move stock returns by 5-8bps, most of the effect reverses over 5 days.
  - Barber and Odean (2008) show individual investors trade following news coverage, attention-driven buying.

Today's talk

- Summary of two papers:
  - "Sentiment during recessions."
  - "Journalists and the stock market," joint with Casey Dougal, Joey Engelberg, and Chris Parsons.
- Two new datasets:
  - NYT financial news 1905–2005.
  - WSJ "Abreast of the market" *authorship* 1970–2007.
- Two research questions:
  - Can the media really influence market prices? When?
  - Can we claim a *causal* relationship?

Motivation I

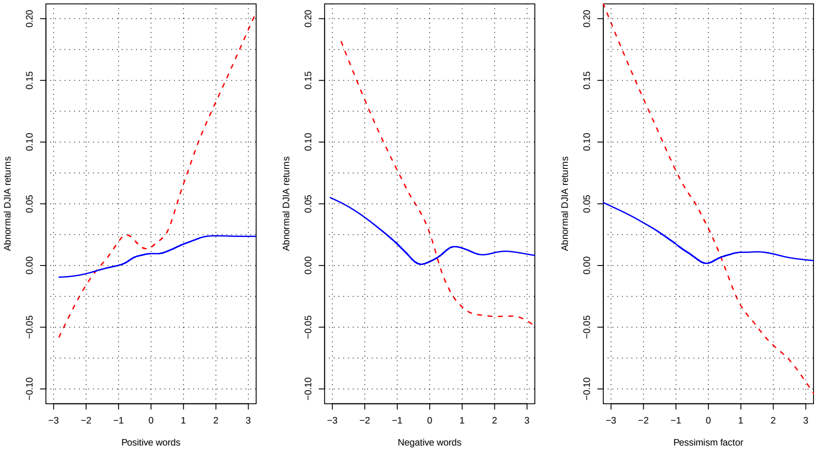
First project essentially reproduces Tetlock (2007) by constructing measure of financial news content by counting positive and negative words from two columns in the NYT (1905–2005).

- Evidence from psychology showing people react more to information when primed into negative mood states.
- Akerlof and Shiller (2009): "we conceive of the link between **changes in confidence** [...] as being especially **large and critical when economies are going into a downturn**, but not so important at other times."
- Pre-1984 the media was much more concentrated, so we have an outlet that virtually every investor read.
- Independent time-series, significantly higher statistical power.

Main results I

- Media predicts stock returns at the daily frequency, particularly so **during recessions**.
  - Over a day, a one-SD increase in pessimism makes DJIA drop by 12 basis points in recessions.
  - Effect is 3.5 basis points in expansions.
- The predictability lasts into the afternoon, and the initial effect is gone in five days.
  - Rules out hypothesis based on information quickly getting impounded into prices.
  - Reversal points into sentiment interpretation.
- Ancillary results:
  - Both positive and negative words bite.
  - The effect is particularly pronounced Mondays.

Non-parametric estimates by business cycle



Motivation II

- Growing literature on the effects of media on asset prices.
  - Tetlock (2007) shows that the content of Abreast-of-the-market (AOTM) column helps predict DJIA returns.
- But it is difficult to say much about causation.
  - Underlying events could cause both news stories and returns, or, alternatively, journalists' spin could move asset prices.
  - "It is unclear whether the financial news media induces, amplifies, or simply reflects' investors interpretations of stock market performance." (Tetlock, 2007)
- Goal of this project: show that there is a *causal* link between the journalists' writing and aggregate market returns.

Research design

- Study **authorship** of the AOTM column from the WSJ (1970–2007).
  - Same column as Tetlock (2007).
- **What** journalist write may be endogenous, but **when** they write is not.
  - Who writes on a given date is not related to market conditions, but rather scheduling (rotations).
  - This exogenous variation gives us the instrument we need to argue causality.
- The identity of the journalist may be as informative as the content of the article itself.
  - Jon Stewart vs. Diego García telling the same joke.
  - Remark: we get causality at the cost of having a “black-box” (no content analysis).

Main results II

- Methodology: F-test of joint significance of 25 journalist dummies.
  - Joint test over day-of and day-after publication has p-values on the 0.01% ballpark (rare, very rare).
- Stronger when interacted with lagged returns (the "material" used by the journalists when writing).
  - Journalists attenuate or amplify existing market conditions.
- Results go through with simple IV that takes advantage of predictability of rotations.

Economic variables

Study covers 1905–2005 time-period.

- Focus on the total return index for the Dow Jones Industrial average.
    - CRSP data not available for 20+ years in my sample.
- NBER recession data.
    - Twenty different recessions during these 101 years, including the Great Depression.
- Intraday DJIA data from Global Financial Data (1933–2005).

New York Times columns

- Focus on two columns from the New York Times: "Financial Markets" and "Topics in Wall Street" ("Sidelights of the day," eventually "Market Place").
  - Both ran virtually uninterrupted from 1905–2005.
  - The later is slightly longer, around 1000 words versus 700.
  - Discuss anything from the stock market to industry conditions to details on a specific company.
  - A total of over 55,000+ columns.
- Obtained copies of their pdf images via `ProQuest`, from the NYT Historical Archive.
  - Also available to any subscriber of the NYT.
- Convert the pdf images to text via "optical character recognition" (OCR).
  - Use a version of ABBYY available at the Carolina Digital Library and Archives (CDLA).

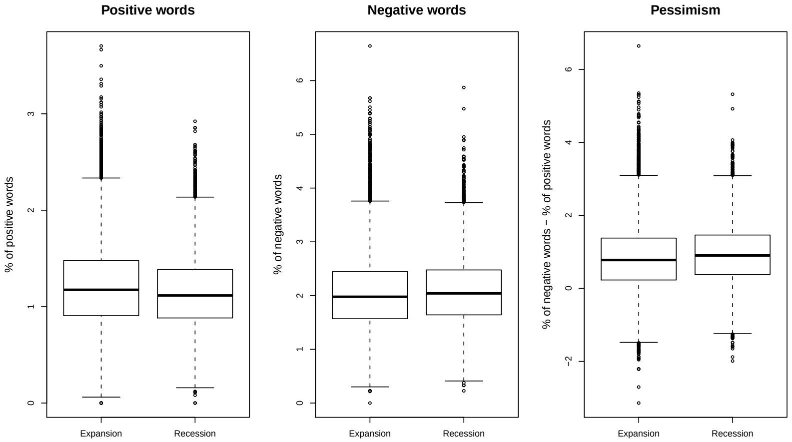
Media measures

- For each article I count the total number of positive words g_{it} and negative words b_{it}.
    - Using McDonald's dictionaries for signing words.
- Normalize by the total number of words w_{it} to create a measure of positive/negative news.
- Aggregate all the news by taking the average of all articles written from market close to market open to create two time-series, G_t and B_t, that run on the same time domain as the DJIA.
    - Essentially grab the news written on the afternoon and published the next day.
    - Need the close-to-open to get some news published on Mondays that dealt with the stock market on Saturdays.

Sample statistics, media content

News	Mean	Median	25%-qu.	75%-qu.	SD
A. All dates					
Positive	1.20	1.16	0.90	1.46	0.42
Negative	2.06	1.99	1.59	2.45	0.67
Pessimism	0.86	0.81	0.26	1.40	0.88
B. Recessions					
Positive	1.15	1.12	0.88	1.38	0.39
Negative	2.09	2.04	1.64	2.48	0.64
Pessimism	0.94	0.90	0.38	1.46	0.84
C. Expansions					
Positive	1.21	1.17	0.91	1.48	0.43
Negative	2.05	1.98	1.57	2.45	0.68
Pessimism	0.84	0.78	0.23	1.38	0.89

Boxplot of media variables



DJIA returns, sample stats

A. Sample stats

	Mean	Median	25%-qu.	75%-qu.	SD
All dates	0.020	0.044	-0.450	0.526	1.071
Expansions	0.042	0.056	-0.410	0.517	0.943
Recessions	-0.053	-0.011	-0.637	0.565	1.408

Simple model of asset returns

$$R_t = (1 - D_t)\gamma_1 \mathcal{L}_s(R_t) + D_t \gamma_2 \mathcal{L}_s(R_t) + \boldsymbol{\eta} X_t + \epsilon_t;$$

where $\mathcal{L}_s(R_t) = \{R_{t-1}, \ldots, R_{t-s}\}$.

DJIA autocorrelations

B. Time-series regression

Expansions	γ_1	t-stat	Recessions	γ_2	t-stat
$(1-D_t)\times R_{t-1}$	0.052	3.1	$D_t \times R_{t-1}$	0.024	0.8
$(1-D_t)\times R_{t-2}$	-0.045	-2.7	$D_t \times R_{t-2}$	-0.019	-0.7
$(1-D_t)\times R_{t-3}$	0.004	0.3	$D_t \times R_{t-3}$	0.004	0.2
$(1-D_t)\times R_{t-4}$	0.005	0.5	$D_t \times R_{t-4}$	0.062	2.6
$(1-D_t)\times R_{t-5}$	0.011	0.7	$D_t \times R_{t-5}$	0.022	0.9
	η	t-stat		η	t-stat
I_{Tue}	0.140	6.3	I_{Fri}	0.167	7.4
I_{Wed}	0.153	6.7	I_{Sat}	0.189	7.5
I_{Thu}	0.125	5.6	D_t	-0.091	-5.0

Econometric approach - news to returns

Estimate the following model of stock returns

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

where

- $\mathcal{L}_s$ denotes an s-lag operator.
- M_t denotes one of our media measures i.e. $M_t = G_t$ in the case of positive news, $M_t = B_t$ in the case of negative news, and $M_t = B_t - G_t$ in the case of our pessimism factor.
- X_t includes a constant term, day-of-the-week dummies, and a dummy for recessions or an expansion.
- Standard errors as in White (1980).

Feedback news to stock returns

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
A. All dates						
M_{t-1}	0.039	5.2	-0.043	-5.2	-0.055	-6.3
M_{t-2}	0.003	0.4	0.003	0.3	0.001	0.2
M_{t-3}	-0.008	-1.1	0.005	0.7	0.008	1.0
M_{t-4}	-0.013	-1.8	0.008	1.0	0.013	1.6
M_{t-5}	-0.005	-0.6	0.009	1.2	0.010	1.3
B. Tests	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_1 = 0$	26.9	0.000	26.8	0.000	40.0	0.000
$\sum_{j=2}^{5} \beta_j = 0$	3.1	0.077	3.6	0.059	5.6	0.018

Feedback news to stock returns along the business cycle

Same as before, but interacting variables with business cycle indicators:

$$\begin{aligned} R_t &= (1 - D_t)\left(\beta_1 \mathcal{L}_s(M_t) + \gamma_1 \mathcal{L}_s(R_t) + \psi_1 \mathcal{L}_s(R_t^2)\right) \\ &\quad + D_t\left(\beta_2 \mathcal{L}_s(M_t) + \gamma_2 \mathcal{L}_s(R_t) + \psi_2 \mathcal{L}_s(R_t^2)\right) \\ &\quad + \boldsymbol{\eta} X_t + \epsilon_t. \end{aligned}$$

Basic test: differences in influence of media variables, $\boldsymbol{\beta}_1$ versus $\boldsymbol{\beta}_2$.

Feedback news to stock returns

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
A. Expansions (β_1)						
$(1 - D_t) \times M_{t-1}$	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
B. Recessions (β_2)						
$D_t \times M_{t-1}$	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

Feedback news to stock returns

	Positive		Negative		Pessimism	
C. Tests	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
$\sum_{j=2}^{5} \beta_{1j} = 0$	1.6	0.205	2.6	0.109	3.4	0.066
$\sum_{j=2}^{5} \beta_{2j} = 0$	0.7	0.403	1.6	0.212	2.3	0.132

- Leading coefficients much larger in recessions.
- Some evidence of reversals.

GARCH(1,1) adjusted returns

	Positive		Negative		Pessimism	
A. GARCH-adjusted returns						
	β	t-stat	β	t-stat	β	t-stat
$(1-D_t)\times M_{t-1}$	0.022	3.1	-0.025	-3.3	-0.033	-4.1
$D_t \times M_{t-1}$	0.051	3.5	-0.070	-4.3	-0.087	-5.1
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11}=\beta_{21}$	3.1	0.079	6.1	0.014	8.2	0.004
$\sum_{j=2}^{5}\beta_{1j}=0$	1.3	0.254	1.9	0.170	3.0	0.085
$\sum_{j=2}^{5}\beta_{2j}=0$	0.0	0.868	2.7	0.103	2.1	0.148

Using residuals of media regressions

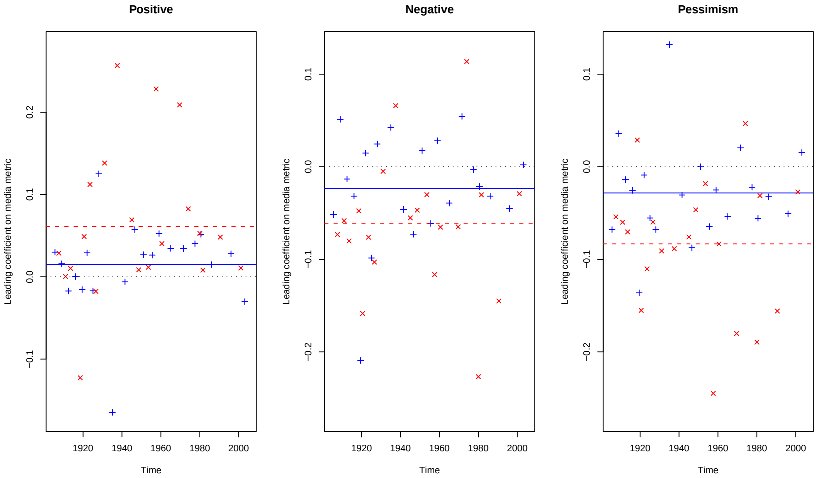
	Positive		Negative		Pessimism	
B. Orthogonal media measures						
	β	t-stat	β	t-stat	β	t-stat
$(1-D_t)\times M_{t-1}$	0.022	3.3	-0.027	-3.8	-0.032	-4.4
$D_t \times M_{t-1}$	0.078	3.8	-0.068	-3.1	-0.094	-4.1
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11}=\beta_{21}$	6.9	0.008	3.1	0.080	6.6	0.010
$\sum_{j=2}^{5}\beta_{1j}=0$	0.1	0.815	0.2	0.667	0.2	0.688
$\sum_{j=2}^{5}\beta_{2j}=0$	0.0	0.898	0.0	0.867	0.0	0.917

Robust regressions

C. Robust regression

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
$(1-D_t)\times M_{t-1}$	0.024	4.0	-0.025	-4.0	-0.034	-5.3
$D_t \times M_{t-1}$	0.055	4.6	-0.086	-7.0	-0.101	-7.9
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11}=\beta_{21}$	5.4	0.020	19.7	0.000	22.2	0.000
$\sum_{j=2}^{5}\beta_{1j}=0$	3.2	0.075	1.3	0.258	3.4	0.065
$\sum_{j=2}^{5}\beta_{2j}=0$	1.2	0.282	7.7	0.005	3.5	0.063

Estimates by business cycle



Econometric approach - stock to news

Estimate the following model

$$\begin{aligned} M_t &= (1 - D_t)\left(\beta_1 \mathcal{L}_s(M_t) + \lambda_1 R_t + \gamma_1 \mathcal{L}_s(R_t) + \psi_1 \mathcal{L}_s(R_t^2)\right) \\ &\quad + D_t\left(\beta_2 \mathcal{L}_s(M_t) + \lambda_2 R_t + \gamma_2 \mathcal{L}_s(R_t) + \psi_2 \mathcal{L}_s(R_t^2)\right) \\ &\quad + \boldsymbol{\eta} X_t + \epsilon_t. \end{aligned}$$

- System of equations not strictly a VAR, since one should include contemporaneous returns.
    - Columns are finished after the market closed.
- If we pick up differences in reporting, perhaps that explains the differential effect of media on stock returns.

Feedback stock returns to news

	Positive		Negative		Pessimism	
	λ, β	t-stat	λ, β	t-stat	λ, β	t-stat
A. Using raw returns $(\lambda_1, \beta_1, \lambda_2, \beta_2)$						
$(1 - D_t) \times R_t$	0.335	32.3	-0.332	-30.7	-0.414	-34.2
$(1 - D_t) \times R_{t-1}$	0.046	6.1	-0.056	-7.6	-0.059	-7.8
$D_t \times R_t$	0.197	16.8	-0.221	-19.5	-0.263	-20.3
$D_t \times R_{t-1}$	0.048	5.4	-0.045	-5.4	-0.052	-5.7
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\lambda_1 = \lambda_2$	28.5	0.000	20.2	0.000	28.6	0.000

Feedback stock returns to news

	Positive		Negative		Pessimism	
	λ, β	t-stat	λ, β	t-stat	λ, β	t-stat
B. Returns normalized by GARCH(1,1) $(\lambda_1, \beta_1, \lambda_2, \beta_2)$						
$(1 - D_t) \times R_t$	0.345	48.7	-0.342	-50.2	-0.427	-61.7
$(1 - D_t) \times R_{t-1}$	0.040	5.6	-0.044	-6.4	-0.047	-6.8
$D_t \times R_t$	0.324	26.3	-0.338	-31.4	-0.413	-37.0
$D_t \times R_{t-1}$	0.065	5.5	-0.050	-4.7	-0.063	-5.6
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\lambda_1 = \lambda_2$	1.9	0.086	0.9	0.474	0.8	0.564

Hard news versus word counts

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Exp, high-inf, β_{11}	0.020	1.9	-0.026	-2.1	-0.031	-2.6
Exp, low-inf, β_{21}	0.026	3.0	-0.029	-3.0	-0.038	-3.8
Rec, high-inf, β_{31}	0.098	3.1	-0.094	-2.6	-0.127	-3.4
Rec, low-inf, 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

Mondays and holidays

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Exp, Mon/holidays, β_{11}	0.056	3.3	-0.062	-3.5	-0.079	-4.1
Exp, weekday, β_{21}	0.015	1.9	-0.019	-2.1	-0.024	-2.6
Rec, Mon/holidays, β_{31}	0.188	4.0	-0.208	-3.8	-0.267	-4.9
Rec, weekdays, β_{41}	0.062	2.8	-0.061	-2.3	-0.085	-3.1
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	4.7	0.031	4.8	0.029	6.6	0.010
$\beta_{31} = \beta_{41}$	6.2	0.012	6.0	0.014	10.0	0.002
$\beta_{11} = \beta_{31}$	6.9	0.008	6.2	0.013	10.5	0.001
$\beta_{21} = \beta_{41}$	4.0	0.046	2.3	0.128	4.5	0.034

Returns from 11am to close

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
$(1-D_t)\times M_{t-1}$	-0.001	-0.1	-0.015	-2.3	-0.012	-1.8
$D_t \times M_{t-1}$	0.057	3.1	-0.047	-2.6	-0.065	-3.5
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\beta_{11}=\beta_{21}$	8.9	0.003	2.9	0.088	7.3	0.007

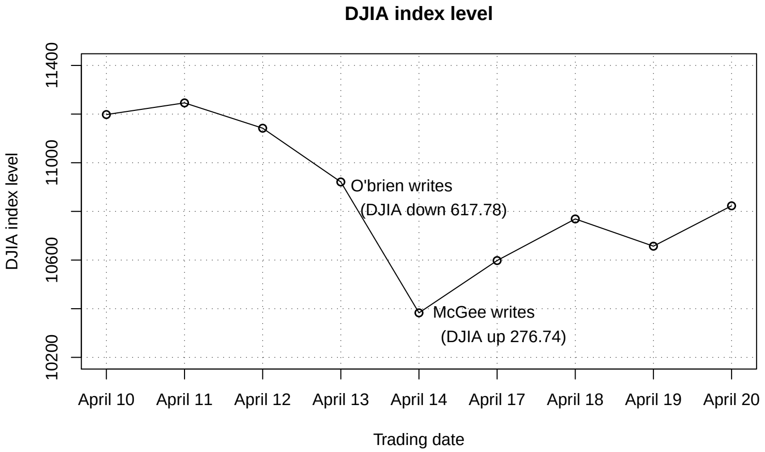
Conclusion

- Paper presents evidence of strong asymmetry in the reaction of DJIA returns to news across the business cycle.
  - Effect particularly strong in **recessions**.
- Evidence is consistent with **sentiment** playing a more important role during economic downturns.
  - Concentration on Monday, afternoon predictability.

Conclusion

- Paper presents evidence of strong asymmetry in the reaction of DJIA returns to news across the business cycle.
  - Effect particularly strong in **recessions**.
- Evidence is consistent with **sentiment** playing a more important role during economic downturns.
  - Concentration on Monday, afternoon predictability.
- As a research agenda:
  - Study low-frequency components of news measures, and their relationship to the business cycle.
  - Create a time-series of sentiment dating back to the XIX century (WSJ and FT archives now open).

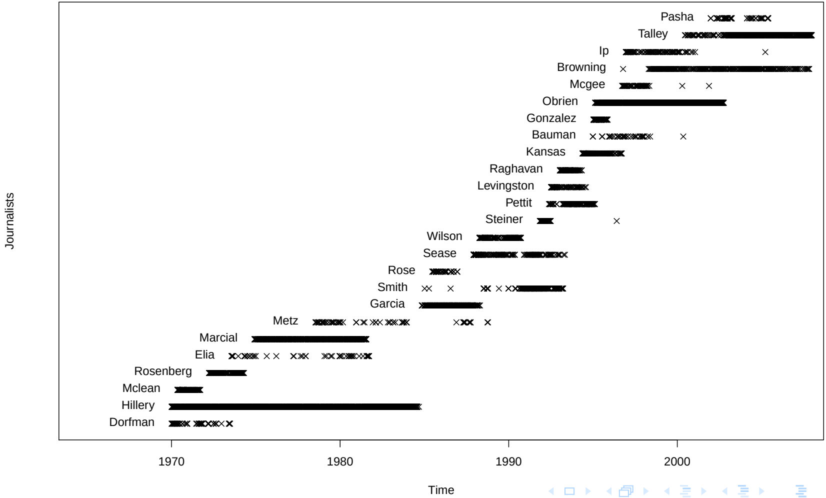
Journalists' spin and the DJIA



Data

- AOTM column, from ProQuest.
  - Widely read market summary column in the US.
  - Data pre-1984 in image format – use OCR to convert pdf files into readable text.
- Time period 1970–2007.
  - Prior to 1970 the WSJ did not consistently print the author of the AOTM column.
- 9592 AOTM articles in total, with 25 journalists that wrote 50 or more times (5% of articles without an author).
- Focus on excess DJIA returns (also CRSP and S&P500).

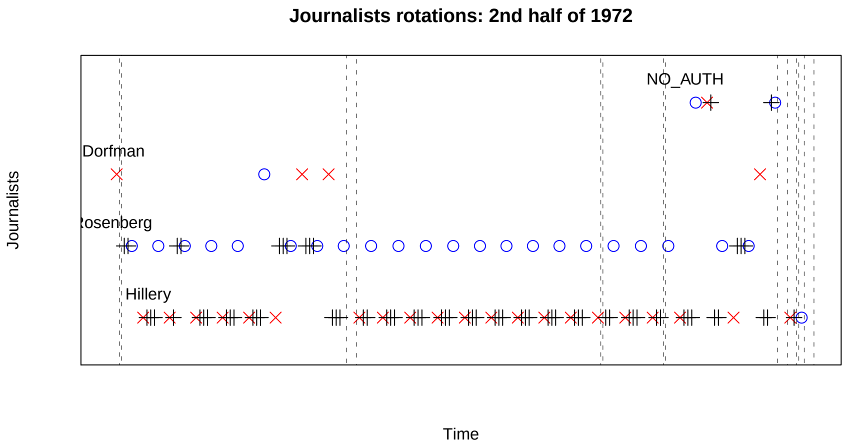
Pretty timeline



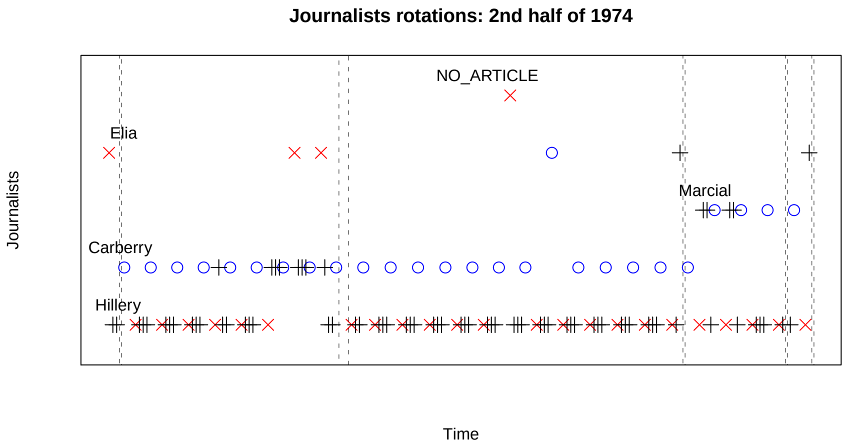
Journalists

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

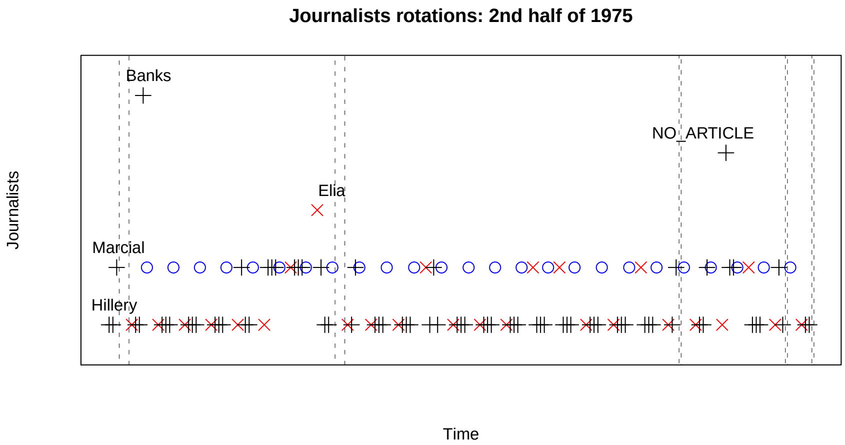
Journalists' rotations



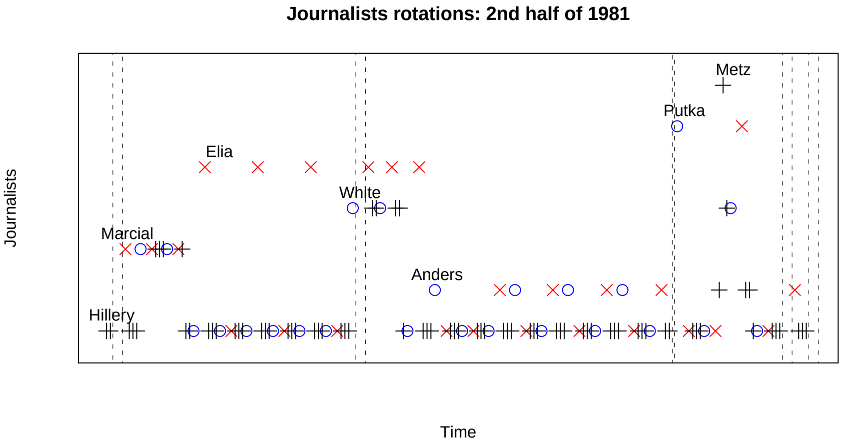
Journalists' rotations



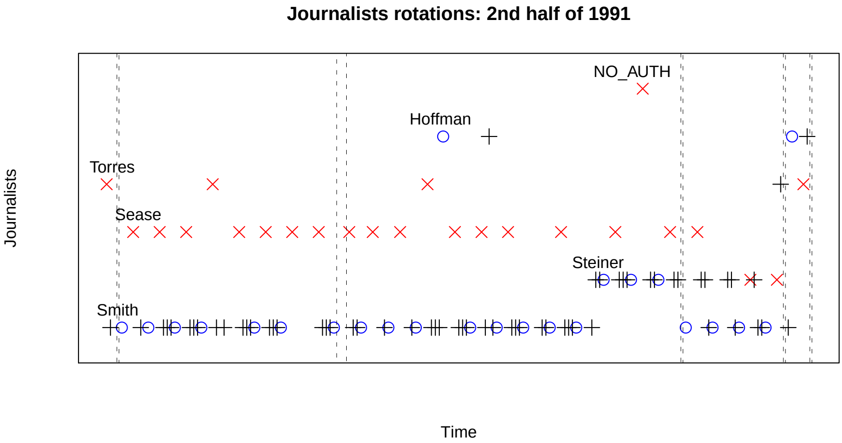
Journalists' rotations



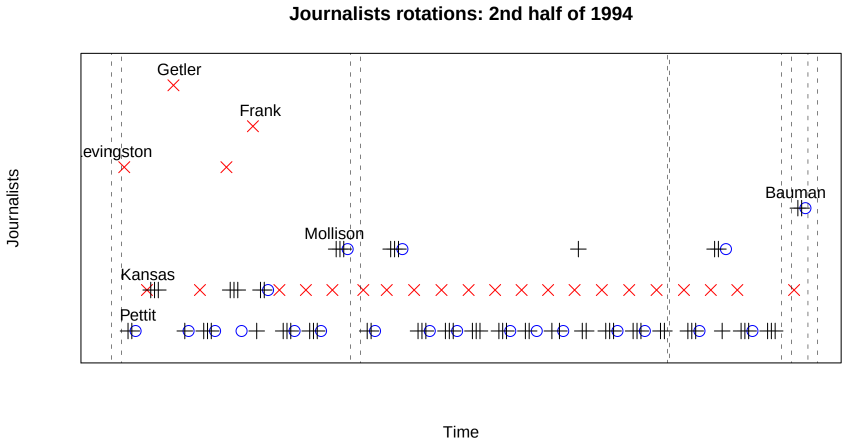
Journalists' rotations



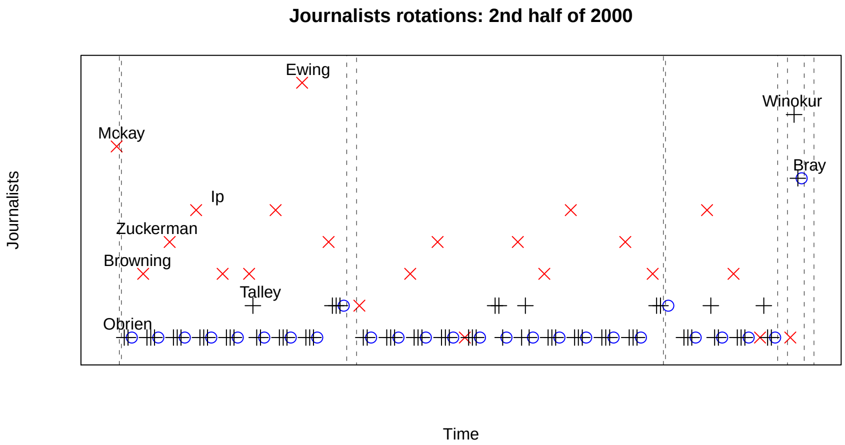
Journalists' rotations



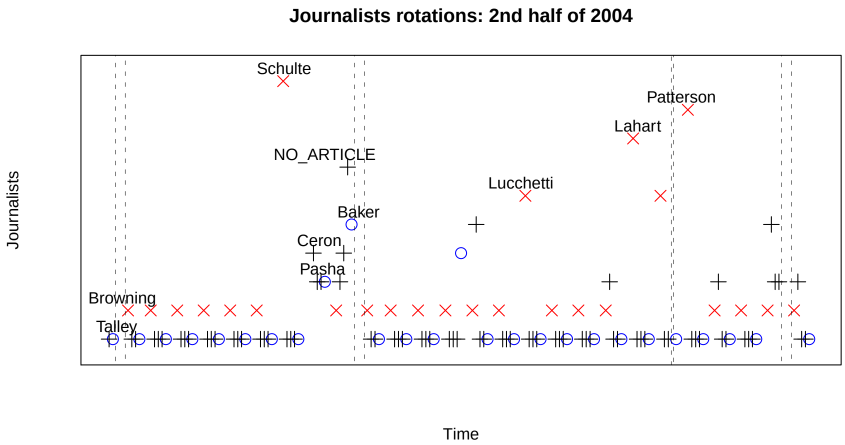
Journalists' rotations



Journalists' rotations



Journalists' rotations



Journalists' writing styles

We examine stylistic differences among journalists estimating:

$$Characteristic_{i,t} = c + \sum_{i=1}^{26} \beta_i \cdot Journalist_{i,t} + \sum_{j=1}^{5} \alpha_j \cdot r_{t-j} + \gamma \cdot Weekday_t + \epsilon_t$$

where $Journalist_{i,t} = 1$ if and only if journalist i has a column published in date t.

	Pessimism	Syllables	WPS	Complex	Fog
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

Magnitudes are statistically and economically (linguistically) large.

Univariate evidence

	$\bar{r}_{\text{wrote}}$	t-stat	$\bar{r}_{\text{day after}}$	t-stat	$\bar{r}_{\text{no journ}}$
Hillery	0.007	(1.0)	0.031	(1.2)	−0.034
O'Brien	−0.011	(0.3)	0.202***	(2.6)	−0.036
Talley	0.025	(0.8)	0.066	(1.0)	−0.020
Marcial	0.036	(−0.2)	−0.055*	(−1.8)	0.046
Garcia	0.112	(1.1)	−0.249	(−1.0)	0.005
Smith	0.067	(0.5)	0.170	(1.3)	0.035
Browning	−0.041*	(−1.8)	−0.020	(−1.2)	0.108
Pettit	0.066	(0.5)	0.002	(−0.4)	0.029
Wilson	−0.057**	(−2.0)	0.084	(0.5)	0.050
Sease	0.053	(0.3)	0.134	(1.2)	0.026
Rosenberg	−0.095*	(−1.9)	−0.304***	(−3.6)	0.099
McLean	0.119	(0.6)	−0.005	(−0.9)	0.075
Kansas	0.069	(−0.3)	−0.032	(−1.1)	0.103
Ip	0.200***	(2.9)	−0.021	(0.0)	−0.025
Gonzalez	0.195	(1.4)	0.118	(0.8)	0.013
Torres	0.107	(1.2)	0.225**	(2.3)	0.010
Raghavan	0.031	(0.0)	0.041	(0.1)	0.029
Levingston	0.042	(0.2)	−0.031	(−0.7)	0.029
Metz	0.275*	(1.9)	−0.431**	(−2.3)	0.011
Pasha	0.034	(−0.7)	−0.142	(−1.1)	0.123
Granahan	−0.076	(−1.6)	0.142	(0.7)	0.057
Rose	0.028	(0.0)	−0.258***	(−2.6)	0.027
Dorfman	0.001	(−0.5)	−0.064	(−0.7)	0.073
Bauman	0.411**	(2.4)	0.428**	(2.5)	0.003
Steiner	−0.191	(−1.3)	0.027	(0.2)	−0.005

Basic specification

Simple model of stock returns:

$$
\begin{aligned}
r_t \quad = \quad & c + \sum_{i=1}^{25} \left\{ \beta_{i,t-1} \cdot Journalist_{i,t-1} + \beta_{i,t} \cdot Journalist_{i,t} \right\} \\
& + \eta \cdot Controls_t + \epsilon_t .
\end{aligned}
$$

where $Journalist_{i,t} = 1$ if and only if journalist i has a column published in date t.

$Controls_t$: lagged returns; lagged squared returns; lagged volume; day-of-the-week dummies; January dummy; year-fixed effects.

Simple specification

	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	

Formal tests

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

- Journalists are jointly significantly different from zero.
- Four are individually significant the publication day, seven the day after.

Lagged returns and journalists' spin

- Journalists need “ingredients” to write about.
- It is plausible that the “spin” given by authors changes with market conditions.
- Next set of tests captures this idea.
- Simple interaction of journalists dummies with lagged returns:

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

Conditional on $r_{t-1} = \pm 1$, a journalists marginal impact on date t's stock returns is $\beta_{i,t} \pm \gamma_{i,t}$, and $\beta_{i,t-1} \pm \gamma_{i,t-1}$,

Interaction with lagged returns

	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					

Formal tests

H_0 :	$\beta_{t-1} = \beta_t = 0$	$\gamma_{t-1} = \gamma_t = 0$	Both
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

- Journalists are jointly significantly different from zero. Very, very rare to find our results under the null (1 in 10,000).
- Eight of the interactions are individually significant the publication day, ten the day after (reversals?).

IV approach

- Skeptical reader may wonder if editoral decisions may not be correlated with stock returns.
  - Put the more senior guy after a market drop (or more bullish. . .).
  - If journalists identity is correlated with market conditions, then causality is out of the window.
- IV approach:
  - Estimate journalist arrival as a function of historical patterns: (a) day-of-the-week for each year ("schedule"), (b) who wrote yesterday, (c) who wrote that day last week.
  - Use fitted values ($\in [0, 1]$) in the predictability regressions.

IV fit – rotations versus market outcomes

	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
Kansas	0.073	0.174	0.462	0.028	0.000
McLean	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
Steiner	0.224	0.241	0.406	0.010	0.000
Dorfman	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

DJIA predictability with IVs

	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	0.3	0.058	0.6	0.047	0.5	0.170^{*}	1.8
O'Brien	0.194	1.3	0.118	0.9	0.067	0.6	0.066	0.6
Talley	0.061	0.5	0.007	0.1	0.140	1.2	0.028	0.2
Marcial	0.092	0.7	0.037	0.3	0.091	0.7	0.156	1.2
Garcia	0.251	1.5	0.155	0.9	0.411	1.5	0.648^{*}	1.8
Smith	0.222	1.4	0.385^{**}	2.6	0.067	0.4	0.154	0.9
Wilson	0.156	0.9	0.066	0.4	0.034	0.2	0.026	0.2
Browning	0.201	1.0	0.079	0.4	0.193	1.0	0.130	0.8
Pettit	0.159	1.2	0.164	1.1	0.057	0.3	0.042	0.2
Sease	0.184	0.7	0.294	1.3	0.314	1.5	0.297	1.4
Rosenberg	0.089	0.6	0.287^{**}	2.1	0.109	0.6	0.457^{***}	3.0
Kansas	0.122	0.8	0.069	0.5	0.029	0.2	0.193	0.9
McLean	0.007	0.0	0.195	1.1	0.258	1.6	0.213	1.1
Raghavan	0.288	1.0	0.052	0.2	0.063	0.1	0.035	0.1
Ip	0.391	1.2	0.196	0.6	0.537	1.6	0.019	0.1
Gonzalez	0.259	1.0	0.175	0.8	0.754^{*}	1.8	0.551	1.5
Metz	0.243	0.8	0.062	0.2	0.021	0.0	0.572	1.3
Levingston	0.030	0.2	0.113	0.6	0.186	0.6	0.043	0.2
Pasha	0.291	0.9	0.484	1.3	0.556^{**}	2.1	0.229	0.8
Rose	0.183	0.8	0.113	0.5	0.736^{**}	2.5	0.569^{**}	2.0
Steiner	0.077	0.3	0.129	0.5	0.372	1.1	0.303	0.8
Dorfman	0.029	0.1	0.562^{***}	3.2	0.335	1.5	0.347^{**}	2.4
Bauman	0.547	1.5	0.156	0.5	0.733^{*}	1.9	0.412	1.0
McGee	0.358	1.1	0.712^{***}	2.9	0.509	1.1	0.601^{**}	2.0
Elia	0.549	1.2	0.434	1.1	0.588	1.1	0.579^{*}	1.9
Observations	9592							
R^2_{noJFE}	0.028	R^2_{JFE}	0.054					

Formal tests

H_0 :	$\beta_{t-1} = \beta_t = 0$	$\gamma_{t-1} = \gamma_t = 0$	Both
p-value BOOT	0.000	0.000	0.000

- The IVs are sufficient to predict the DJIA.
- Journalists writing (authorship) has a direct causal effect on aggregate asset prices.

Robustness

Estimate the baseline model and p-values of different hypothesis using:

- Bootstrap F-stats to address small-sample biases.
- Only ten authors.
- Open-close returns.
- CRSP VW returns.
- S&P500 returns.
- GARCH-adjustments.
- Winsorized returns.

Robustness

	Open-Close	S&P 500	CRSP VW	GARCH	Winsorized	10 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

- Lots of zeros.

Robustness

	Open-Close	S&P 500	CRSP VW	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

- They are all zero.

Conclusions

- “The history of speculative bubbles begins roughly with the advent of newspapers,” (Shiller, 2000).
- Wide-spread belief that media can actually drive aggregate asset prices.
    - Causality part is hard to argue, since coverage is endogenous to market events.
- Our contribution: show there is a causal link from what financial reporters write and market stock returns.
    - Use exogenous rotation of journalists to have an instrument that allows us to claim causality.

Conclusions

- "The history of speculative bubbles begins roughly with the advent of newspapers," (Shiller, 2000).
- Wide-spread belief that media can actually drive aggregate asset prices.
  - Causality part is hard to argue, since coverage is endogenous to market events.
- Our contribution: show there is a causal link from what financial reporters write and market stock returns.
  - Use exogenous rotation of journalists to have an instrument that allows us to claim causality.
- Lots left to do for future research:
  - What is the mechanism?
  - When does the press really matter?
  - Are they imposing negative externalities?

Conclusions

- "The history of speculative bubbles begins roughly with the advent of newspapers," (Shiller, 2000).
- Wide-spread belief that media can actually drive aggregate asset prices.
  - Causality part is hard to argue, since coverage is endogenous to market events.
- Our contribution: show there is a causal link from what financial reporters write and market stock returns.
  - Use exogenous rotation of journalists to have an instrument that allows us to claim causality.

Conclusions

- "The history of speculative bubbles begins roughly with the advent of newspapers," (Shiller, 2000).
- Wide-spread belief that media can actually drive aggregate asset prices.
  - Causality part is hard to argue, since coverage is endogenous to market events.
- Our contribution: show there is a causal link from what financial reporters write and market stock returns.
  - Use exogenous rotation of journalists to have an instrument that allows us to claim causality.
- Lots left to do for future research:
  - What is the mechanism?
  - When does the press really matter?
  - Are they imposing negative externalities?