Internet Appendix for

Investor Attention and Stock Returns

Jian Chen

School of Economics,
Xiamen University

Guohao Tang

College of Finance and Statistics,
Hunan University

Jiaquan Yao[*]

School of Management,
Jinan University

Guofu Zhou

Olin School of Business,
Washington University in St. Louis

December 2020

This Internet Appendix repots the results for supplementary and robustness tests as described below:

Table IA 1. In-sample Forecasting Results for Individual Attention Proxies

This table reports results from following predictive regression,

$$R_{t+h} = \alpha + \beta A_t + \varepsilon_{t+h},$$

where R_{t+h} is the average stock market excess return over the prediction horizon h, $h = 1$, 6, and 12 months, and A_t denotes one of the 12 attention proxies: abnormal trading volume (A^{AVol}), extreme returns (A^{ERet}), past returns (A^{PRet}), nearness to the Dow 52-week high (A^{52wH}), nearness to the Dow historical high (A^{HisH}), analyst coverage ($A^{\#AC}$), change in advertising expenses (A^{CAD}), mutual fund inflow (A^{Inflow}), mutual fund outflow ($A^{Outflow}$), media coverage (A^{Media}), Google search volume (A^{Google}), and search-traffic on Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system (A^{EDGAR}). Panels A, B, and C show the results for A^{AVol}, A^{ERet}, A^{PRet}, A^{52wH}, A^{HisH}, $A^{\#AC}$, and A^{CAD} from January 1980 to December 2017; for A^{Inflow}, $A^{Outflow}$, A^{Media}, and A^{Google} from January 2004 to December 2017; and for A^{EDGAR} from January 2004 to June 2017, respectively. In each panel, the estimates of regression slopes (β), t-statistics (t-stat.) based on the Hodrick (1992) standard errors, and R^2s are reported.

	$h = 1$			$h = 6$			$h = 12$		
	β (%)	t-stat.	R^2 (%)	β (%)	t-stat.	R^2 (%)	β (%)	t-stat.	R^2 (%)
Panel A: 1980:01-2017:12									
A^{AVol}	0.15	0.67	0.12	0.12	1.02	0.42	0.04	-0.60	0.08
A^{ERet}	-0.28	-0.91	0.41	-0.14	-0.85	0.56	-0.09	-1.21	0.49
A^{PRet}	-0.22	-1.51	0.26	-0.27	-2.17	2.06	-0.29	-2.00	4.84
A^{52wH}	-0.15	-0.89	0.12	-0.18	-2.02	0.95	-0.22	-2.46	2.82
A^{HisH}	-0.32	-1.95	0.54	-0.36	-4.42	3.61	-0.32	-2.84	5.63
$A^{\#AC}$	0.29	1.48	0.45	0.18	0.91	0.87	0.18	1.04	1.80
A^{CAD}	-0.23	-1.29	0.28	-0.22	-1.38	1.32	-0.26	-2.06	3.81
Panel B: 2004:01-2017:12									
A^{Inflow}	-0.54	-2.61	1.79	-0.58	-1.71	8.49	-0.18	-1.20	1.70
$A^{Outflow}$	-0.76	-1.59	3.45	-0.44	-1.05	4.76	-0.20	-1.03	2.12
A^{Media}	0.06	0.20	0.02	-0.09	-0.70	0.21	-0.06	-0.47	0.19
A^{Google}	-0.09	-0.40	0.05	-0.27	-0.97	1.84	-0.28	-1.15	3.20
Panel C: 2004:01-2017:06									
A^{EDGAR}	0.40	1.31	0.92	0.34	1.08	2.75	0.32	1.00	4.64

Table IA 2. Additional Results of Comparison with Economic Variables

This table reports results from following predictive regression,

$$R_{t+h} = \alpha + \beta A_t + \phi X_t + \varepsilon_{t+h},$$

where R_{t+h} is the average stock market excess return over the prediction horizon h, $h = 1, 3, 6,$ and 12 months, A_t is one of the attention measures A^{PLS}, A^{SPCA}, and A^{PCA} at time t, and X_t represents a vector of economic variables from Goyal and Welch (2008), including dividend–payout ratio (DE), net equity expansion (NTIS), treasury bill rate (TBL), long-term yield (LTY), default yield spread (DFY), and inflation (INFL). Panels A, B, and C show results for A^{PLS}, A^{SPCA}, and A^{PCA}, respectively. In each panel, reported are estimates of regression slopes and adjusted R^2s in percentage form. Brackets below the slope estimates report the t-statistics based on the Hodrick (1992) standard errors. The sample period is from January 1980 to December 2017.

	Panel A: Results for A^{PLS}			Panel B: Results for A^{SPCA}			Panel C: Results for A^{PCA}		
	$h=1$	$h=6$	$h=12$	$h=1$	$h=6$	$h=12$	$h=1$	$h=6$	$h=12$
A	-0.63	-0.55	-0.34	-0.47	-0.39	-0.24	-0.18	-0.23	-0.24
	[-2.35]	[-2.28]	[-2.13]	[-3.06]	[-2.59]	[-2.08]	[-0.98]	[-1.64]	[-1.93]
DE	0.03	0.17	0.09	0.03	0.16	0.08	0.00	0.12	0.03
	[0.14]	[0.75]	[0.56]	[0.16]	[0.95]	[0.64]	[-0.02]	[0.63]	[0.23]
NTIS	0.13	0.24	0.13	-0.11	0.03	0.00	-0.02	0.11	0.05
	[0.46]	[0.74]	[0.52]	[-0.41]	[0.10]	[0.00]	[-0.05]	[0.31]	[0.20]
TBL	0.04	0.31	-0.40	-0.52	-0.19	-0.72	-0.48	-0.15	-0.68
	[0.07]	[0.44]	[-0.87]	[-0.87]	[-0.32]	[-1.63]	[-0.77]	[-0.23]	[-1.47]
LTY	-0.19	-0.26	0.42	0.23	0.11	0.66	0.18	0.05	0.61
	[-0.26]	[-0.40]	[0.93]	[0.35]	[0.19]	[1.46]	[0.25]	[0.08]	[1.30]
DFR	0.42	0.09	0.02	0.39	0.07	0.00	0.44	0.10	0.03
	[2.70]	[1.70]	[0.42]	[1.99]	[0.82]	[0.07]	[2.50]	[1.22]	[0.50]
INFL	0.14	-0.21	-0.16	0.16	-0.20	-0.15	0.20	-0.17	-0.14
	[0.39]	[-1.02]	[-1.38]	[0.42]	[-1.01]	[-1.36]	[0.52]	[-0.85]	[-1.22]
Adj. R^2	1.74	8.65	11.60	1.12	5.53	9.28	0.19	3.01	9.35

Table IA 3. Additional Out-of-sample Forecasting Results for Individual Attention Proxies

This table reports the out-of-sample R_{OS}^2's and *MSFE-adjusted* statistics for predicting the average stock market returns over the prediction horizon h based on one of the 7 attention proxies: abnormal trading volume (A^{AVol}), extreme returns (A^{ERet}), past returns (A^{PRet}), nearness to the Dow 52-week high (A^{52wH}), nearness to the Dow historical high (A^{HisH}), analyst coverage ($A^{\#AC}$), change in advertising expenses (A^{CAD}). Panels A, B, and C show the results for $h = 1$, 6, and 12 months, respectively. All of the predictors and regression slopes are estimated recursively using the data available at the forecast formation time t. The out-of-sample period is from January 1995 to December 2017. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: $h = 1$		Panel B: $h = 6$		Panel C: $h = 12$	
	R_{OS}^2 (%)	*MSFE-adjusted*	R_{OS}^2 (%)	*MSFE-adjusted*	R_{OS}^2 (%)	*MSFE-adjusted*
A^{AVol}	-0.77	-0.73	-0.45	-0.02	-0.76	-0.99
A^{ERet}	-0.13	0.26	-0.44	0.19	-1.47	-2.03
A^{PRet}	-0.39	0.18	-3.11	0.04	-4.32	0.85
A^{52wH}	-0.63	-0.65	-4.41	-3.80	1.12**	2.15
A^{HisH}	-0.66	-0.12	-3.23	0.04	2.49***	3.57
$A^{\#AC}$	-0.20	1.04	-0.93	-0.08	-3.55	-1.94
A^{CAD}	-0.51	-0.25	-3.07	-1.82	2.40***	3.85