



```

1 . cd "C:\Users\eliven\Dropbox\ELLW_2026\code"
   C:\Users\eliven\Dropbox\ELLW_2026\code

2 . doedit "05_main_tables.do"

3 . do "C:\Users\eliven\Dropbox\ELLW_2026\code\05_main_tables.do"

4 . *****
   > *****
5 . *****
   > *****
6 . *****
   > *****
7 . * Table 1 - Table 7 in the main paper
8 .
9 . *****
10 . * Set directory
11 .
12 . global dataset "datasets"

13 . global results "results"

14 .
15 .
16 . *=====
17 . *                Table 1: Descriptive Statistics for Regression Analyses
18 . *=====
19 .
20 . *****Firm-day Level*****
21 . *Import firm-day level dataset
22 . use "$dataset\firm_day_dataset",clear

23 .
24 . * Define events and intermediaries
25 . global events pos_ea neg_ea mef_pos mef_neg event_general

26 . global intermediaries AnaReport_D media_D

27 .
28 . * Identify the regression subsample, Table 1 part 1
29 . qui reghdfe specific_q $events $intermediaries if tradingday == 1 , abs(stkcd date) cl
   > uster(stkcd date)

30 . unique stkcd if e(sample)
    Number of unique values of stkcd is  5122
    Number of records is  596989

31 . winsor2 AbnSpread20to1 illiq AbnVol120to1 vpin AbnRet,replace

32 .
33 . * Export the firmday-level summary statistics, Table 1 part 1
34 . logout, save("$results\table1_firmday_part1") excel dec(3) replace: tabstat specific_q
   > $events $intermediaries logSVI logques_hudong posratio AbnSpread20to1 illiq AbnVol120to1
   > vpin AbnRet dsmvtll average_sentiment_score if e(sample) , stat(N me sd p25 med p75)
   > c(s)

```

Variable	N	Mean	SD	p25	p50	p75
specific_q	596989	1.415941	12.18472	0	0	1
pos_ea	596989	.005372	.0730966	0	0	0
neg_ea	596989	.0045562	.0673457	0	0	0
mef_pos	596989	.001928	.0438668	0	0	0
mef_neg	596989	.0025377	.0503786	0	0	0
event_gene~l	596989	.1370092	.3438574	0	0	0
AnaReport_D	596989	.0200908	.1403111	0	0	0
media_D	596989	.158554	.36526	0	0	0
logSVI_All	596989	5.055211	2.926193	5.036952	6.291569	6.982863
logques_hu~g	596989	.1985969	.4086274	0	0	0
posratio	150333	.0868457	.2243938	0	0	0
AbnSpread2~1	594700	.3165926	3.724795	-1.487382	-.113552	1.515445
illiq	596989	4.411813	6.131595	.7522	2.1957	5.4384

AbnVol20to1	595553	-.065215	2.21585	-.6980284	-.151152	.3083217
vpin	596989	17.80861	5.226129	14.0999	17.3618	21.0293
AbnRet	594844	-.0431306	2.859651	-1.425342	-.0951663	1.260647
dsmvtll	596989	15.59372	1.074949	14.83417	15.37216	16.16926
average_seve	142532	.2467445	.3174501	0	.25	.4517833

results\table1 firmday part1.xml
dir

```

35 .
36 . * Identify the regression subsample, Table 1 part 2
37 . qui reghdfe AbnRet tsc_avg_sent n_tsc_avg_sent dsmvtll $events $intermediaries logSVI_A
> ll logques_hudong , abs(stkcd date) cluster(stkcd date)

38 . unique stkcd if e(sample)
Number of unique values of stkcd is 2463
Number of records is 22714

39 .
40 . * Export the firmday-level summary statistics, Table 1 part 2
41 . logout, save("$results\table1_firmday_part2") excel dec(3) replace: tabstat tsc_avg_sen
> t n_tsc_avg_sent if e(sample) , stat(N me sd p25 med p75) c(s)

```

Variable	N	Mean	SD	p25	p50	p75
tsc_avg_sent	22714	.3088867	.3553345	.0752719	.3333333	.5555556
n_tsc_avg_~t	22714	.2453911	.2709327	.0930048	.2575019	.4041839

results\table1 firmday part2.xml
dir

```

42 .
43 .
44 . *****Query-Answer Level*****
45 . * Import query-answer level dataset
46 . use "$dataset\user_interaction", clear

47 .
48 . * Identify the regression subsample and export summary statistics for Table 1 answer-le
> vel part
49 . global attributes forward_ratio_b subjective_ratio_b sentiment_ratio_b fog topic_divers
> ity_ratio ner_ratio_a logref logwords_a

50 . qui reghdfe pos_fb signal_alignment $attributes if user_id != "", abs(date) cluster(use
> r_id)

51 .
52 . * Export the answer-level summary statistics
53 . logout, save("$results\table1_answer") excel dec(3) replace: tabstat pos_fb thumbsdown
> post_question signal_alignment forward_ratio_b subjective_ratio_b sentiment_ratio_b fog
> topic_diversity_ratio ner_ratio_a total_ref total_word_a if e(sample) , stat(N me sd
> p25 med p75) c(s)

```

Variable	N	Mean	SD	p25	p50	p75
pos_fb	376491	13.51374	34.18707	0	0	0
thumbsdown	376491	.0964166	3.10361	0	0	0
post_quest~n	376491	84.67931	514.9976	2	9	29
signal_ali~t	376491	.3646077	.4774483	0	0	1
forward_r~b	376491	.3760808	.2931148	.125	.3333333	.5849057
subjectiv~b	376491	.4127411	.2523011	.2307692	.4	.5853659
sentiment~b	376491	.3349026	.4280251	.0454545	.375	.6666667
fog	376491	10.69751	2.866906	8.75	10.2549	12.125
topic_dive~o	376491	.1481105	.1131369	.047619	.1363636	.2272727
ner_ratio_a	376491	.1941466	.1316605	.0952839	.1684447	.2665746
total_refe~s	376491	5.482646	8.62087	0	3	6
total_word_a	376491	556.1249	473.2336	288	434	633

results\table1 answer.xml
dir

```

54 .
55 . *****User Level*****
56 . * Import user level dataset
57 . use "$dataset\user_dataset", clear

58 .
59 . preserve

60 .
61 . * Scale specific financial metric variables (0/1 variable) by 100 for clearer descripti
> ve statistics
62 . local vars post_roe_q cum_3_rq cum_3_ra post_solveny_q cum_3_sq cum_3_sa post_ar_q cu
> m_3_aq cum_3_aa

63 . foreach v of local vars {
    2.     replace `v' = `v' * 100
    3. }
(2,537 real changes made)
(564 real changes made)
(2,542 real changes made)
(2,472 real changes made)
(760 real changes made)
(1,281 real changes made)
(2,141 real changes made)
(670 real changes made)
(2,529 real changes made)

64 .
65 . * Define control variables using query characteristics
66 . global query_controls post_sent_qb post_subj_qb post_forward_qb logquery logword

67 .
68 . * Identify the regression subsample and export summary statistics for Table 1 user-leve
> l part
69 . qui reghdfe post_mention_rate_q metric_mention_q_rate metric_mention_a_rate $query_cont
> rols, abs(user_start_month) cluster(user_start_month)

70 . unique user_id if e(sample)
    Number of unique values of user_id is 78887
    Number of records is 78887

71 .
72 . * Export the user-level summary statistics
73 . logout, save("$results\table1_user") excel dec(3) replace: tabstat first3_ner_q_mean f
> irst3_ner_a_mean post_ner_q metric_mention_q_rate metric_mention_a_rate post_mention_ra
> te_q cum_3_rq cum_3_ra post_roe_q cum_3_sq cum_3_sa post_solveny_q cum_3_aq cum_3_aa p
> ost_ar_q if e(sample) , stat(N me sd p25 med p75) c(s) format(%9.3f)

```

Variable	N	Mean	SD	p25	p50	p75
first~q_mean	78453.000	0.320	0.202	0.157	0.287	0.445
first~a_mean	78414.000	0.182	0.109	0.101	0.165	0.244
post_ner_q	78887.000	0.315	0.192	0.171	0.288	0.417
metri~q_rate	78887.000	0.139	0.273	0.000	0.000	0.333
metri~a_rate	78887.000	0.536	0.356	0.333	0.667	1.000
post_men~e_q	78887.000	0.160	0.276	0.000	0.000	0.214
cum_3_rq	78887.000	0.700	8.336	0.000	0.000	0.000
cum_3_ra	78887.000	3.180	17.548	0.000	0.000	0.000
post_roe_q	78887.000	3.201	17.602	0.000	0.000	0.000
cum_3_sq	78887.000	0.956	9.730	0.000	0.000	0.000
cum_3_sa	78887.000	1.605	12.566	0.000	0.000	0.000
post_solve~q	78887.000	3.130	17.412	0.000	0.000	0.000
cum_3_aq	78887.000	0.838	9.115	0.000	0.000	0.000
cum_3_aa	78887.000	3.172	17.524	0.000	0.000	0.000
post_ar_q	78887.000	2.709	16.234	0.000	0.000	0.000

results\table1 user.xml
dir

74 .
 75 . restore

```
76 .
77 . *=====
78 . *           Table 2: User-based Cross-Sections of Query Topics and Tasks
79 . *=====
80 . * Import query-answer level dataset
81 . use "$dataset\user_interaction", clear

82 .
83 . * Table2: Mean of different topics and tasks for all registered users
84 . logout, save("$results\table2_Means") excel replace: ///
    > tabstat financial_perf_q_dummy business_outlook_q_dummy production_ops_q_dummy stoc
    > k_perf_q_dummy product_tech_innov_q_dummy market_comp_q_dummy info_sense_q_dummy info_s
    > earch_q_dummy cause_anal_q_dummy summary_q_dummy compare_q_dummy trend_pred_q_dummy imp
    > act_anal_q_dummy overall_eval_q_dummy fs_anal_q_dummy if user_id != "", statistics(mean
    > ) columns(statistics) format(%6.3f)
```

Variable	Mean
financial_~y	0.279
business_o~y	0.318
production~y	0.307
stock_perf~y	0.237
product_te~y	0.270
market_com~y	0.219
info_sense~y	0.388
info_searc~y	0.327
cause_anal~y	0.168
summary_q_~y	0.033
compare_q_~y	0.136
trend_pred~y	0.136
impact_ana~y	0.187
overall_ev~y	0.158
fs_anal_q_~y	0.146

results\table2 Means.xml
dir

```
85 .
86 . * Table 2: Univariate analysis (T-tests) by AI usage intensity (High vs. Low)
87 . * Grouping variable: month90ge -whether they are top 10% users in their month entry coh
    > orts
88 . logout, save("$results\table2_AIUsage") excel replace: ttable2 financial_perf_q_dummy b
    > usiness_outlook_q_dummy production_ops_q_dummy stock_perf_q_dummy product_tech_innov_q_
    > dummy market_comp_q_dummy info_sense_q_dummy info_search_q_dummy cause_anal_q_dummy sum
    > mary_q_dummy compare_q_dummy trend_pred_q_dummy impact_anal_q_dummy overall_eval_q_dumm
    > y fs_anal_q_dummy, by(month90ge)
```

Variables	G1(0)	Mean1	G2(1)	Mean2	MeanDiff
financia~y	485802	0.171	859372	0.340	-0.168***
business~y	485802	0.311	859372	0.322	-0.012***
producti~y	485802	0.249	859372	0.339	-0.091***
stock_pe~y	485802	0.249	859372	0.230	0.019***
product_~y	485802	0.298	859372	0.255	0.043***
market_c~y	485802	0.228	859372	0.214	0.014***
info_sen~y	644344	0.486	974082	0.323	0.163***
info_sea~y	644344	0.252	974082	0.376	-0.124***
cause_an~y	644344	0.125	974082	0.196	-0.071***
summary_~y	644344	0.023	974082	0.040	-0.017***
compare_~y	644344	0.119	974082	0.148	-0.029***
trend_pr~y	644344	0.168	974082	0.114	0.054***
impact_a~y	644344	0.145	974082	0.215	-0.071***
overall_~y	644344	0.187	974082	0.138	0.048***
fs_anal_~y	644344	0.070	974082	0.196	-0.126***

results\table2 AIUsage.xml
dir

```
89 .
90 . * Table 2: Univariate analysis (T-tests) by financial metrics usage
91 . * Grouping variable: avg_90_qge - whether the user's average metric mention count is in
  > the top 10%
92 . logout, save("$results\table2_FMIUsage") excel replace: ttable2 financial_perf_q_dummy b
  > usiness_outlook_q_dummy production_ops_q_dummy stock_perf_q_dummy product_tech_innov_q_
  > dummy market_comp_q_dummy info_sense_q_dummy info_search_q_dummy cause_anal_q_dummy sum
  > mary_q_dummy compare_q_dummy trend_pred_q_dummy impact_anal_q_dummy overall_eval_q_dumm
  > y fs_anal_q_dummy,by(avg_90_qge)
```

Variables	G1(0)	Mean1	G2(1)	Mean2	MeanDiff
financia~y	1010389	0.156	334785	0.651	-0.495***
business~y	1010389	0.320	334785	0.312	0.008***
producti~y	1010389	0.255	334785	0.462	-0.207***
stock_pe~y	1010389	0.250	334785	0.198	0.052***
product_~y	1010389	0.301	334785	0.179	0.122***
market_c~y	1010389	0.226	334785	0.200	0.026***
info_sen~y	1254375	0.464	364051	0.127	0.337***
info_sea~y	1254375	0.286	364051	0.468	-0.182***
cause_an~y	1254375	0.149	364051	0.233	-0.084***
summary_~y	1254375	0.020	364051	0.081	-0.062***
compare_~y	1254375	0.119	364051	0.195	-0.076***
trend_pr~y	1254375	0.150	364051	0.086	0.064***
impact_a~y	1254375	0.172	364051	0.240	-0.068***
overall_~y	1254375	0.176	364051	0.096	0.080***
fs_anal_~y	1254375	0.068	364051	0.414	-0.346***

results\table2 FMUsage.xml
dir

```
93 .
94 . *=====
95 . * Table 3: Firm-Specific Queries around Corporate Information Events
96 . *=====
97 . * Import firm-day level dataset
98 . use "$dataset\firm_day_dataset",clear

99 .
100 . * Define events and intermediaries
101 . global events pos_ea neg_ea mef_pos mef_neg event_general

102 . global intermediaries AnaReport_D media_D

103 .
104 . * Export the association between firm specific queries and corporate information events
105 . ppmlhdfc specific_q $events if tradingday == 1, abs(stkcd date) cluster(stkcd date)
  (dropped 584 observations that are either singletons or separated by a fixed effect)
  Iteration 1: deviance = 1.4216e+06 eps = . iters = 3 tol = 1.0e-04 min(eta
  > ) = -4.91 P
  Iteration 2: deviance = 1.2019e+06 eps = 1.83e-01 iters = 3 tol = 1.0e-04 min(eta
  > ) = -6.46
  Iteration 3: deviance = 1.1818e+06 eps = 1.71e-02 iters = 2 tol = 1.0e-04 min(eta
  > ) = -8.14
  Iteration 4: deviance = 1.1805e+06 eps = 1.09e-03 iters = 2 tol = 1.0e-04 min(eta
  > ) = -9.57
  Iteration 5: deviance = 1.1804e+06 eps = 6.41e-05 iters = 2 tol = 1.0e-04 min(eta
  > ) = -10.44
  Iteration 6: deviance = 1.1804e+06 eps = 2.20e-06 iters = 2 tol = 1.0e-05 min(eta
  > ) = -10.74
  Iteration 7: deviance = 1.1804e+06 eps = 1.83e-08 iters = 2 tol = 1.0e-06 min(eta
  > ) = -10.78 S
  Iteration 8: deviance = 1.1804e+06 eps = 4.96e-12 iters = 2 tol = 1.0e-07 min(eta
  > ) = -10.78 S 0
```

```
> -----
(legend: p: exact partial-out    s: exact solver    h: step-halving    o: epsilon below tole
> rance)
Converged in 8 iterations and 18 HDfE sub-iterations (tol = 1.0e-08)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.
```

HDfE PPML regression	No. of obs	=	596,405
Absorbing 2 HDfE groups	Residual df	=	116
Statistics robust to heteroskedasticity	Wald chi2(5)	=	115.82
Deviance = 1180384.505	Prob > chi2	=	0.0000
Log pseudolikelihood = -803565.7245	Pseudo R2	=	0.6685

Number of clusters (stkcd) = 5,117
 Number of clusters (date) = 117
 (Std. err. adjusted for 117 clusters in stkcd date)

specific_q	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
pos_ea	.0703447	.0289194	2.43	0.015	.0136637	.1270257
neg_ea	.1308605	.0313819	4.17	0.000	.0693532	.1923679
mef_pos	.1396789	.1178784	1.18	0.236	-.0913586	.3707164
mef_neg	.2851246	.1871981	1.52	0.128	-.0817769	.6520262
event_general	.0682411	.016321	4.18	0.000	.0362526	.1002296
_cons	2.251232	.0016887	1333.09	0.000	2.247922	2.254541

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
stkcd	5117	5117	0	*
date	117	117	0	*

* = FE nested within cluster; treated as redundant for DoF computation

```
106 . outreg2 using "$results\Table3.xls", drop(o* 0*) replace tstat tdec(2) nocons dec(3) ad
> dstat(Pseudo R2, `e(r2_p)') addtext(Firm FE, YES, Date FE, YES)
results\Table3.xls
dir : seeout
```

```
107 . ppmlhdfe specific_q $events $intermediaries if tradingday == 1, abs(stkcd date) cluster
> (stkcd date)
(dropped 584 observations that are either singletons or separated by a fixed effect)
Iteration 1: deviance = 1.4177e+06 eps = . iters = 3 tol = 1.0e-04 min(eta
> ) = -4.93 P
Iteration 2: deviance = 1.1987e+06 eps = 1.83e-01 iters = 3 tol = 1.0e-04 min(eta
> ) = -6.48
Iteration 3: deviance = 1.1787e+06 eps = 1.70e-02 iters = 2 tol = 1.0e-04 min(eta
> ) = -8.16
Iteration 4: deviance = 1.1774e+06 eps = 1.08e-03 iters = 2 tol = 1.0e-04 min(eta
> ) = -9.59
Iteration 5: deviance = 1.1773e+06 eps = 6.44e-05 iters = 2 tol = 1.0e-04 min(eta
> ) = -10.46
Iteration 6: deviance = 1.1773e+06 eps = 2.22e-06 iters = 2 tol = 1.0e-05 min(eta
> ) = -10.76
Iteration 7: deviance = 1.1773e+06 eps = 1.85e-08 iters = 2 tol = 1.0e-06 min(eta
> ) = -10.80 S
Iteration 8: deviance = 1.1773e+06 eps = 5.08e-12 iters = 2 tol = 1.0e-07 min(eta
> ) = -10.80 S 0
```

```
> -----
(legend: p: exact partial-out    s: exact solver    h: step-halving    o: epsilon below tole
> rance)
Converged in 8 iterations and 18 HDfE sub-iterations (tol = 1.0e-08)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.
```

HDFE PPML regression	No. of obs	=	596,405
Absorbing 2 HDFE groups	Residual df	=	116
Statistics robust to heteroskedasticity	Wald chi2(7)	=	170.05
Deviance = 1177308.674	Prob > chi2	=	0.0000
Log pseudolikelihood = -802027.809	Pseudo R2	=	0.6691

Number of clusters (stkcd) = 5,117
 Number of clusters (date) = 117
 (Std. err. adjusted for 117 clusters in stkcd date)

specific_q	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
pos_ea	.0143419	.0270503	0.53	0.596	-.0386757	.0673596
neg_ea	.0602401	.0386761	1.56	0.119	-.0155637	.1360439
mef_pos	.0864365	.1121007	0.77	0.441	-.1332768	.3061499
mef_neg	.2362934	.1858063	1.27	0.203	-.1278804	.6004671
event_general	.0456975	.0137215	3.33	0.001	.0188039	.0725911
AnaReport_D	.0201003	.0271733	0.74	0.459	-.0331584	.073359
media_D	.1490308	.017198	8.67	0.000	.1153234	.1827382
_cons	2.199356	.0059919	367.06	0.000	2.187612	2.2111

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5117	5117	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
108 . outreg2 using "$results\Table3.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> stat(Pseudo R2, `e(r2_p)') addtext(Firm FE, YES, Date FE, YES)
results\Table3.xls
dir : seout
```

```
109 .
110 .
111 . * Export the disclosure informativeness cross-sectional results
112 . ppmlhdfc specific_q pos_ea neg_ea event_general mef_pos c.mef_pos#c.moretopics mef_neg
> c.mef_neg#c.moretopics $intermediaries if tradingday == 1 , abs(stkcd date) cluster(stk
> cd date)
(dropped 584 observations that are either singletons or separated by a fixed effect)
Iteration 1: deviance = 1.4175e+06 eps = . iters = 3 tol = 1.0e-04 min(eta
> ) = -4.93 P
Iteration 2: deviance = 1.1986e+06 eps = 1.83e-01 iters = 3 tol = 1.0e-04 min(eta
> ) = -6.48
Iteration 3: deviance = 1.1786e+06 eps = 1.70e-02 iters = 2 tol = 1.0e-04 min(eta
> ) = -8.16
Iteration 4: deviance = 1.1773e+06 eps = 1.08e-03 iters = 2 tol = 1.0e-04 min(eta
> ) = -9.59
Iteration 5: deviance = 1.1772e+06 eps = 6.44e-05 iters = 2 tol = 1.0e-04 min(eta
> ) = -10.46
Iteration 6: deviance = 1.1772e+06 eps = 2.22e-06 iters = 2 tol = 1.0e-05 min(eta
> ) = -10.76
Iteration 7: deviance = 1.1772e+06 eps = 1.85e-08 iters = 2 tol = 1.0e-06 min(eta
> ) = -10.80 S
Iteration 8: deviance = 1.1772e+06 eps = 5.08e-12 iters = 2 tol = 1.0e-07 min(eta
> ) = -10.80 S 0
```

```
> -----
(legend: p: exact partial-out s: exact solver h: step-halving o: epsilon below tole
> range)
Converged in 8 iterations and 18 HDFE sub-iterations (tol = 1.0e-08)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.
```

HDFE PPML regression	No. of obs	=	596,405
Absorbing 2 HDFE groups	Residual df	=	116
Statistics robust to heteroskedasticity	Wald chi2(9)	=	923.26
Deviance = 1177242.462	Prob > chi2	=	0.0000
Log pseudolikelihood = -801994.703	Pseudo R2	=	0.6691

Number of clusters (stkcd) = 5,117
 Number of clusters (date) = 117
 (Std. err. adjusted for 117 clusters in stkcd date)

specific_q	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
pos_ea	.0136097	.026489	0.51	0.607	-.0383079	.0655272
neg_ea	.0601613	.0387074	1.55	0.120	-.0157038	.1360264
event_general	.0456438	.0137459	3.32	0.001	.0187023	.0725853
mef_pos	.2835492	.0954703	2.97	0.003	.0964309	.4706676
c.mef_pos#c.moretopics	-.4556554	.2011409	-2.27	0.023	-.8498844	-.0614264
mef_neg	.3500381	.2423513	1.44	0.149	-.1249617	.8250379
c.mef_neg#c.moretopics	-.3514876	.2528224	-1.39	0.164	-.8470104	.1440353
AnaReport_D	.0208022	.02739	0.76	0.448	-.0328811	.0744856
media_D	.1488505	.0172464	8.63	0.000	.1150481	.1826528
_cons	2.199393	.0060063	366.18	0.000	2.187621	2.211165

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5117	5117	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```

113 . outreg2 using "$results\Table3.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> stat(Pseudo R2, `e(r2_p)') addtext(Firm FE, YES, Date FE, YES)
results\Table3.xls
dir : seeout

114 . ppmlhdfc specific_q pos_ea neg_ea event_general mef_pos c.mef_pos#c.longmef mef_neg c
> .mef_neg#c.longmef $intermediaries if tradingday == 1 , abs(stkcd date) cluster(stkcd d
> ate)
(dropped 584 observations that are either singletons or separated by a fixed effect)
Iteration 1: deviance = 1.4176e+06 eps = . iters = 3 tol = 1.0e-04 min(eta
> ) = -4.93 P
Iteration 2: deviance = 1.1986e+06 eps = 1.83e-01 iters = 3 tol = 1.0e-04 min(eta
> ) = -6.48
Iteration 3: deviance = 1.1786e+06 eps = 1.70e-02 iters = 2 tol = 1.0e-04 min(eta
> ) = -8.16
Iteration 4: deviance = 1.1774e+06 eps = 1.08e-03 iters = 2 tol = 1.0e-04 min(eta
> ) = -9.59
Iteration 5: deviance = 1.1773e+06 eps = 6.44e-05 iters = 2 tol = 1.0e-04 min(eta
> ) = -10.46
Iteration 6: deviance = 1.1773e+06 eps = 2.22e-06 iters = 2 tol = 1.0e-05 min(eta
> ) = -10.76
Iteration 7: deviance = 1.1773e+06 eps = 1.85e-08 iters = 2 tol = 1.0e-06 min(eta
> ) = -10.80 S
Iteration 8: deviance = 1.1773e+06 eps = 5.08e-12 iters = 2 tol = 1.0e-07 min(eta
> ) = -10.80 S 0

> _____
(legend: p: exact partial-out s: exact solver h: step-halving o: epsilon below tole
> rance)
Converged in 8 iterations and 18 HDFE sub-iterations (tol = 1.0e-08)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.

```


HDFE PPML regression	No. of obs	=	596,405
Absorbing 2 HDFE groups	Residual df	=	116
Statistics robust to heteroskedasticity	Wald chi2(9)	=	223.35
Deviance = 1177271.91	Prob > chi2	=	0.0000
Log pseudolikelihood = -802009.4269	Pseudo R2	=	0.6691

Number of clusters (stkcd) = 5,117
Number of clusters (date) = 117
(Std. err. adjusted for 117 clusters in stkcd date)

specific_q	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
pos_ea	.0137067	.0265796	0.52	0.606	-.0383884	.0658017
neg_ea	.0603658	.0387362	1.56	0.119	-.0155559	.1362874
event_general	.0456577	.013739	3.32	0.001	.0187297	.0725856
mef_pos	.2542618	.1061174	2.40	0.017	.0462755	.4622481
c.mef_pos#c.longmef	-.3956877	.2114938	-1.87	0.061	-.810208	.0188326
mef_neg	.1860216	.137657	1.35	0.177	-.0837813	.4558244
c.mef_neg#c.longmef	.1029525	.2115622	0.49	0.627	-.3117018	.5176069
AnaReport_D	.0204837	.0274367	0.75	0.455	-.0332912	.0742587
media_D	.148913	.0172429	8.64	0.000	.1151176	.1827085
_cons	2.199384	.0060021	366.44	0.000	2.18762	2.211148

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5117	5117	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
115 . outreg2 using "$results\Table3.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> stat(Pseudo R2, `e(r2_p)') addtext(Firm FE, YES, Date FE, YES)
results\Table3.xls
dir : seeout
```

```
116 .
117 . *=====
118 . * Table 4: Answer Attributes, User Feedback, and Continued Engagement
119 . *=====
120 .
121 . * Import query-answer level dataset
122 . use "$dataset\user_interaction", clear

123 .
124 . * Define textual attribute controls
125 . global attributes forward_ratio_b subjective_ratio_b sentiment_ratio_b fog topic_divers
> ity_ratio ner_ratio_a logref logwords_a

126 .
127 . * Export correlations of query attributes with answer attributes
128 . logout, save("$results\Table4_panels.xls") excel replace : ///
> pwcorr_a forward_ratio_qb subjective_ratio_qb sentiment_ratio_qb ner_ratio_q lo
> gwords_q forward_ratio_b subjective_ratio_b sentiment_ratio_b ner_ratio_a logwords_a, f
> ormat(%6.3f) star(0.05)
```

	forwa~qb	subje~qb	senti~qb	ner_ra~q	logwor~q	forwa~_b	subje~_b
forward_r~qb	1.000						
subjectiv~qb	0.307***	1.000					
sentiment~qb	0.159***	0.112***	1.000				
ner_ratio_q	-0.137***	-0.090***	-0.061***	1.000			
logwords_q	0.135***	0.126***	0.103***	-0.220***	1.000		
forward_r~_b	0.383***	0.235***	0.046***	-0.176***	0.126***	1.000	
subjectiv~_b	0.234***	0.420***	0.070***	-0.113***	0.196***	0.471***	1.000
sentiment~_b	0.171***	0.110***	0.343***	-0.038***	-0.010***	0.205***	0.255***
ner_ratio_a	-0.159***	-0.187***	-0.021***	0.379***	-0.155***	-0.379***	-0.395***
logwords_a	0.060***	0.108***	-0.023***	-0.099***	-0.010***	0.285***	0.300***
	senti~_b ner_ra~a logwor~a						
sentiment~_b	1.000						
ner_ratio_a	-0.179***	1.000					
logwords_a	0.234***	-0.417***	1.000				

results\Table4_panela.xml
dir

```

129 .
130 . * Export the association between answer attributes and user feedback
131 . foreach x of varlist pos_fb thumbsdown {
132 .     reghdfe `x' signal_alignment $attributes if user_id != "", abs(date) cluster(u
> ser_id)
133 .     3.      outreg2 using "$results\Table4_panelb.xls", drop(o* 0*) append tstat tdec(
> 2) nocons dec(3) addtext(Date FE, YES)
134 .     4.      }
(MWFE estimator converged in 1 iterations)

```

HDFE Linear regression	Number of obs	=	376,491
Absorbing 1 HDFE group	F(9, 87522)	=	159.53
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.0709
	Adj R-squared	=	0.0705
	Within R-sq.	=	0.0555
Number of clusters (user_id) =	87,523		Root MSE = 32.9604

(Std. err. adjusted for 87,523 clusters in user_id)

pos_fb	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
signal_alignment	.2964251	.1702428	1.74	0.082	-.0372493	.6300995
forward_ratio_b	-1.130327	.7339301	-1.54	0.124	-2.568824	.3081693
subjective_ratio_b	6.878554	.5859274	11.74	0.000	5.730142	8.026967
sentiment_ratio_b	1.950612	.4936253	3.95	0.000	.983111	2.918113
fog	.1109148	.0443719	2.50	0.012	.0239463	.1978834
topic_diversity_ratio	-9.301302	3.518505	-2.64	0.008	-16.19754	-2.405064
ner_ratio_a	-4.034983	1.36024	-2.97	0.003	-6.701041	-1.368925
logref	-7.322884	.4953587	-14.78	0.000	-8.293783	-6.351986
logwords_a	5.268353	.4413244	11.94	0.000	4.403361	6.133345
_cons	-10.56311	2.88227	-3.66	0.000	-16.21234	-4.91389

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
date	174	0	174

results\Table4_panelb.xls

dir : seeout

(MWFE estimator converged in 1 iterations)

HDFE Linear regression
 Absorbing 1 HD FE group
 Statistics robust to heteroskedasticity
 Number of clusters (user_id) = 87,523

Number of obs = 376,491
 F(9, 87522) = 7.02
 Prob > F = 0.0000
 R-squared = 0.0006
 Adj R-squared = 0.0001
 Within R-sq. = 0.0002
 Root MSE = 3.1034

(Std. err. adjusted for 87,523 clusters in user_id)

thumbsdown	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
signal_alignment	-.0012785	.0108184	-0.12	0.906	-.0224824	.0199254
forward_ratio_b	.0018565	.0217978	0.09	0.932	-.040867	.0445799
subjective_ratio_b	.0312631	.0274565	1.14	0.255	-.0225513	.0850775
sentiment_ratio_b	-.0394374	.0136719	-2.88	0.004	-.0662343	-.0126406
fog	.0020502	.001898	1.08	0.280	-.00167	.0057703
topic_diversity_ratio	-.1454802	.0532	-2.73	0.006	-.2497518	-.0412086
ner_ratio_a	.0936772	.0558606	1.68	0.094	-.015809	.2031635
logref	-.0281009	.0063926	-4.40	0.000	-.0406304	-.0155714
logwords_a	.0069141	.009139	0.76	0.449	-.0109982	.0248265
_cons	.0741413	.0601244	1.23	0.218	-.043702	.1919846

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
date	174	0	174

results\Table4 panelb.xls

dir : seeout

```

132 .
133 . * Export the association between answer attributes and continued engagement
134 . ppmlhdfc post_question signal_alignment $attributes if pre_question != 0 & user_id != "
> " , abs(date) cluster(user_id)
Iteration 1: deviance = 1.5864e+08 eps = . iters = 1 tol = 1.0e-04 min(eta
> ) = -4.42 P
Iteration 2: deviance = 1.0692e+08 eps = 4.84e-01 iters = 1 tol = 1.0e-04 min(eta
> ) = -5.53
Iteration 3: deviance = 9.7177e+07 eps = 1.00e-01 iters = 1 tol = 1.0e-04 min(eta
> ) = -6.55
Iteration 4: deviance = 9.6357e+07 eps = 8.51e-03 iters = 1 tol = 1.0e-04 min(eta
> ) = -7.00
Iteration 5: deviance = 9.6345e+07 eps = 1.25e-04 iters = 1 tol = 1.0e-04 min(eta
> ) = -7.05
Iteration 6: deviance = 9.6345e+07 eps = 6.17e-08 iters = 1 tol = 1.0e-04 min(eta
> ) = -7.05
Iteration 7: deviance = 9.6345e+07 eps = 3.47e-14 iters = 1 tol = 1.0e-05 min(eta
> ) = -7.05 S 0

```

> _____
 (legend: p: exact partial-out s: exact solver h: step-halving o: epsilon below tole
 > rance)

Converged in 7 iterations and 7 HD FE sub-iterations (tol = 1.0e-08)

HD FE PPML regression
 Absorbing 1 HD FE group
 Statistics robust to heteroskedasticity
 Deviance = 96344975.57
 Log pseudolikelihood = -48843667.07

No. of obs = 328,676
 Residual df = 62,625
 Wald chi2(9) = 744.58
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.3868

Number of clusters (user_id)= **62,626**
(Std. err. adjusted for **62,626** clusters in user_id)

post_question	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
signal_alignment	.1244353	.0198413	6.27	0.000	.0855471	.1633235
forward_ratio_b	.359531	.1531708	2.35	0.019	.0593218	.6597403
subjective_ratio_b	.1567724	.1659057	0.94	0.345	-.1683968	.4819416
sentiment_ratio_b	.2911895	.0859102	3.39	0.001	.1228087	.4595703
fog	-.0019046	.0103516	-0.18	0.854	-.0221933	.0183842
topic_diversity_ratio	-1.815023	1.17778	-1.54	0.123	-4.123429	.493382
ner_ratio_a	-.3698619	.3318642	-1.11	0.265	-1.020304	.2805799
logref	-.3595856	.1817915	-1.98	0.048	-.7158903	-.0032809
logwords_a	-.137976	.0721679	-1.91	0.056	-.2794224	.0034705
_cons	6.490732	.6213157	10.45	0.000	5.272976	7.708488

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
date	174	0	174

```
135 . outreg2 using "$results\Table4_panelb.xls", drop(o* 0*) append tstat tdec(2) nocons add
> stat(Pseudo R2, `e(r2_p)') dec(3) addtext(Date FE, YES)
results\Table4_panelb.xls
dir : seeout
```

```
136 .
137 . *=====
138 . * Table 5: User Learning from Early Interactions with GenAI
139 . *=====
140 . * Import user level dataset
141 . use "$dataset\user_dataset", clear

142 .
143 . * Define control variables using query characteristics
144 . global query_controls post_sent_qb post_subj_qb post_forward_qb logquery logword

145 .
146 . reghdfe post_ner_q first3_ner_a_mean first3_ner_q_mean $query_controls, abs(user_start_
> month) cluster(user_start_month)
(MWFE_estimator converged in 1 iterations)
warning: missing F statistic; dropped variables due to collinearity or too few clusters
```

HDFE Linear regression	Number of obs	=	78,414
Absorbing 1 HDFE group	F(_____, 5)	=	.
Statistics robust to heteroskedasticity	Prob > F	=	.
	R-squared	=	0.2271
	Adj R-squared	=	0.2270
	Within R-sq.	=	0.2260
Number of clusters (user_start_month) =	6Root MSE	=	0.1688

(Std. err. adjusted for 6 clusters in user_start_month)

post_ner_q	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
first3_ner_a_mean	.0813379	.0109717	7.41	0.001	.0531342	.1095417
first3_ner_q_mean	.3523142	.0250605	14.06	0.000	.2878941	.4167344
post_sent_qb	-.0137343	.0038309	-3.59	0.016	-.0235818	-.0038868
post_subj_qb	-.060299	.002643	-22.81	0.000	-.0670931	-.053505
post_forward_qb	-.0516201	.002537	-20.35	0.000	-.0581417	-.0450985
logquery	.0059382	.002163	2.75	0.041	.000378	.0114984
logword	-.0283243	.0019716	-14.37	0.000	-.0333926	-.023256
_cons	.2821738	.0182084	15.50	0.000	.2353675	.32898

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
user_start_month	6	6	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
147 . outreg2 using "$results\Table5.xls", drop(o* 0*) replace tstat tdec(2) nocons dec(3) ad
> dtext(User-Cohort FE, YES)
results\Table5.xls
dir : seeout
```

```
148 . reghdfe post_mention_rate_q metric_mention_a_rate metric_mention_q_rate $query_controls
> , abs(user_start_month) cluster(user_start_month)
(MWFE estimator converged in 1 iterations)
warning: missing F statistic; dropped variables due to collinearity or too few clusters
```

HDFE Linear regression	Number of obs	=	78,887
Absorbing 1 HDFE group	F(7, 5)	=	.
Statistics robust to heteroskedasticity	Prob > F	=	.
	R-squared	=	0.3705
	Adj R-squared	=	0.3705
	Within R-sq.	=	0.3637
Number of clusters (user_start_month) =	6Root MSE	=	0.2193

(Std. err. adjusted for 6 clusters in user_start_month)

post_mention_rate_q	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
metric_mention_a_rate	.0545418	.0072322	7.54	0.001	.0359508	.0731329
metric_mention_q_rate	.3921102	.0284461	13.78	0.000	.3189872	.4652331
post_sent_qb	-.0282323	.0020804	-13.57	0.000	-.03358	-.0228846
post_subj_qb	-.05531	.011597	-4.77	0.005	-.0851212	-.0254989
post_forward_qb	-.0580609	.0046408	-12.51	0.000	-.0699903	-.0461314
logquery	.0229044	.0038682	5.92	0.002	.0129608	.032848
logword	.093119	.0032538	28.62	0.000	.084755	.1014831
_cons	-.2152874	.0040534	-53.11	0.000	-.2257069	-.2048678

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
user_start_month	6	6	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
149 . outreg2 using "$results\Table5.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> text(User-Cohort FE, YES)
results\Table5.xls
dir : seeout
```

```
150 . reghdfe post_roe_q cum_3_ra cum_3_rq $query_controls, abs(user_start_month) cluster(use
> r_start_month)
(MWFE estimator converged in 1 iterations)
warning: missing F statistic; dropped variables due to collinearity or too few clusters
```

HDFE Linear regression	Number of obs	=	78,887
Absorbing 1 HDFE group	F(7, 5)	=	.
Statistics robust to heteroskedasticity	Prob > F	=	.
	R-squared	=	0.1189
	Adj R-squared	=	0.1188
	Within R-sq.	=	0.1184
Number of clusters (user_start_month) =	6Root MSE	=	0.1652

(Std. err. adjusted for 6 clusters in user_start_month)

post_roe_q	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
cum_3_ra	.050941	.009117	5.59	0.003	.0275049	.074377
cum_3_rq	.3218264	.0144716	22.24	0.000	.2846261	.3590268
post_sent_qb	-.0095874	.0037996	-2.52	0.053	-.0193547	.0001799
post_subj_qb	.0001097	.0037111	0.03	0.978	-.00943	.0096493
post_forward_qb	-.0245581	.0014874	-16.51	0.000	-.0283815	-.0207347
logquery	.0388414	.0032314	12.02	0.000	.0305349	.0471479
logword	.026096	.0027563	9.47	0.000	.0190107	.0331812
_cons	-.1166346	.0062575	-18.64	0.000	-.13272	-.1005493

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
user_start_month	6	6	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
151 . outreg2 using "$results\Table5.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> text(User-Cohort FE, YES)
results\Table5.xls
dir : seeout
```

```
152 . reghdfe post_solveny_q cum_3_sa cum_3_sq $query_controls, abs(user_start_month) cluste
> r(user_start_month)
(MWFE estimator converged in 1 iterations)
warning: missing F statistic; dropped variables due to collinearity or too few clusters
```

HDFE Linear regression	Number of obs	=	78,887
Absorbing 1 HDFE group	F(7, 5)	=	.
Statistics robust to heteroskedasticity	Prob > F	=	.
	R-squared	=	0.1300
	Adj R-squared	=	0.1298
	Within R-sq.	=	0.1289
Number of clusters (user_start_month) =	6Root MSE	=	0.1624

(Std. err. adjusted for 6 clusters in user_start_month)

post_solveny_q	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
cum_3_sa	.0784678	.0107966	7.27	0.001	.0507144	.1062213
cum_3_sq	.4222009	.0265446	15.91	0.000	.3539659	.4904359
post_sent_qb	-.0164649	.0038597	-4.27	0.008	-.0263865	-.0065433
post_subj_qb	.0157216	.0026138	6.01	0.002	.0090026	.0224405
post_forward_qb	-.0196256	.0016649	-11.79	0.000	-.0239054	-.0153457
logquery	.0340736	.0038797	8.78	0.000	.0241006	.0440467
logword	.0181399	.0029554	6.14	0.002	.0105429	.025737
_cons	-.0905864	.0056747	-15.96	0.000	-.1051737	-.0759991

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
user_start_month	6	6	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
153 . outreg2 using "$results\Table5.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> text(User-Cohort FE, YES)
results\Table5.xls
dir : seeout
```

```
154 . reghdfe post_ar_q cum_3_aa cum_3_aq $query_controls, abs(user_start_month) cluster(user
> _start_month)
(MWFE estimator converged in 1 iterations)
warning: missing F statistic; dropped variables due to collinearity or too few clusters
```

HDFE Linear regression	Number of obs	=	78,887
Absorbing 1 HDFE group	F(7, 5)	=	.
Statistics robust to heteroskedasticity	Prob > F	=	.
	R-squared	=	0.1553
	Adj R-squared	=	0.1551
	Within R-sq.	=	0.1548
Number of clusters (user_start_month) =	6Root MSE	=	0.1492

(Std. err. adjusted for 6 clusters in user_start_month)

post_ar_q	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
cum_3_aa	.0730877	.0147147	4.97	0.004	.0352623	.1109131
cum_3_aq	.4135614	.0334149	12.38	0.000	.3276657	.4994571
post_sent_qb	-.0312292	.0060642	-5.15	0.004	-.0468178	-.0156407
post_subj_qb	.0001474	.0027214	0.05	0.959	-.0068482	.0071431
post_forward_qb	-.0136935	.0014325	-9.56	0.000	-.0173758	-.0100111
logquery	.0302017	.0017383	17.37	0.000	.0257333	.0346701
logword	.0257745	.0024384	10.57	0.000	.0195064	.0320427
_cons	-.1080465	.0094266	-11.46	0.000	-.1322782	-.0838148

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
user_start_month	6	6	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
155 . outreg2 using "$results\Table5.xls", drop(o* 0*) append tstat tdec(2) nocons dec(3) add
> text(User-Cohort FE, YES)
results\Table5.xls
dir : seeout
```

```
156 . *=====
157 . * Table 6: User Queries and Market Trading
158 . *=====
159 . *Import firm-day level dataset
160 . use "$dataset\firm_day_dataset",clear

161 .
162 . * Define events and intermediaries
163 . global events pos_ea neg_ea mef_pos mef_neg event_general

164 . global intermediaries AnaReport_D media_D

165 .
166 . * Define base control variables
```

```

167 . local base_vars "logspecific_q dsmvtll $events "
168 .
169 . * List of dependent variables to loop through
170 . local depvars "AbnSpread20to1 illiq AbnVol120to1 vpin"

171 .
172 . preserve

173 . * Loop through each dependent variable
174 . foreach y in `depvars' {
175 .     2. winsor2 `y', replace
176 .     3. reghdfe `y' `base_vars' $intermediaries logSVI_All logques_hudong if tradingday
> == 1, abs(stkcd date) cluster(stkcd date)
175 .     4.
176 .     * Append results to the same Excel file
176 .     outre2 using "$results\Table6.xls", drop(o* 0*) ///
>     append tstat tdec(2) nocons dec(3) ///
>     addtext(Firm FE, YES, Date FE, YES)
175 .     5. }
(dropped 1 singleton observations)
(MWFE estimator converged in 4 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.

```

HDFE Linear regression	Number of obs	=	594,699
Absorbing 2 HDFE groups	F(11, 116)	=	64.32
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.2852
	Adj R-squared	=	0.2789
Number of clusters (stkcd)	Within R-sq.	=	0.0214
Number of clusters (date)	Root MSE	=	3.1630

(Std. err. adjusted for 117 clusters in stkcd date)

AbnSpread20to1	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.2960618	.0285939	10.35	0.000	.2394279	.3526957
dsmvtll	-3.763114	.3874471	-9.71	0.000	-4.530502	-2.995726
pos_ea	.5185135	.0824432	6.29	0.000	.3552244	.6818027
neg_ea	.5425101	.1013894	5.35	0.000	.3416955	.7433247
mef_pos	.2231166	.1036716	2.15	0.033	.0177819	.4284512
mef_neg	.9250836	.1515389	6.10	0.000	.6249417	1.225225
event_general	.0923527	.0203025	4.55	0.000	.0521411	.1325643
AnaReport_D	.0259596	.075988	0.34	0.733	-.1245441	.1764633
media_D	.2797683	.0205071	13.64	0.000	.2391515	.3203852
logSVI_All	.2634291	.0673238	3.91	0.000	.1300857	.3967724
logques_hudong	.1512319	.0217096	6.97	0.000	.1082332	.1942306
_cons	57.48966	5.845618	9.83	0.000	45.91168	69.06764

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5121	5121	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

results\Table6.xlsdir : seout

(MWFE estimator converged in 4 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi

> ller applied.

HDFE Linear regression		Number of obs	=	596,989
Absorbing 2 HDFE groups		F(11, 116)	=	45.69
Statistics robust to heteroskedasticity		Prob > F	=	0.0000
		R-squared	=	0.5653
		Adj R-squared	=	0.5615
Number of clusters (stkcd)	=	5,122		
Number of clusters (date)	=	117		
		Within R-sq.	=	0.0637
		Root MSE	=	4.0605

(Std. err. adjusted for 117 clusters in stkcd date)

illiq	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.2352616	.0531826	4.42	0.000	.1299267	.3405965
dsmvtll	-8.767798	.5993903	-14.63	0.000	-9.954966	-7.58063
pos_ea	.3566966	.1407917	2.53	0.013	.0778409	.6355522
neg_ea	.1428344	.2128053	0.67	0.503	-.2786534	.5643221
mef_pos	-.6579204	.5065412	-1.30	0.197	-1.661189	.3453484
mef_neg	1.020575	.1691547	6.03	0.000	.685543	1.355608
event_general	.0544754	.0304943	1.79	0.077	-.0059224	.1148732
AnaReport_D	-.1419975	.096603	-1.47	0.144	-.3333319	.049337
media_D	.1765014	.0257254	6.86	0.000	.125549	.2274537
logSVI_All	-.1189468	.0883567	-1.35	0.181	-.2939484	.0560548
logques_hudong	-.0031158	.0284978	-0.11	0.913	-.0595592	.0533276
_cons	141.6209	9.217488	15.36	0.000	123.3645	159.8773

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5122	5122	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

results\Table6.xls

dir : seeout

(MWFE estimator converged in 4 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.

HDFE Linear regression		Number of obs	=	595,553
Absorbing 2 HDFE groups		F(11, 116)	=	133.72
Statistics robust to heteroskedasticity		Prob > F	=	0.0000
		R-squared	=	0.1237
		Adj R-squared	=	0.1160
Number of clusters (stkcd)	=	5,110		
Number of clusters (date)	=	117		
		Within R-sq.	=	0.0649
		Root MSE	=	2.0834

(Std. err. adjusted for 117 clusters in stkcd date)

AbnVol20to1	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.3251951	.017178	18.93	0.000	.2911719	.3592182
dsmvtll	.9820481	.1992748	4.93	0.000	.5873592	1.376737
pos_ea	.4941926	.0688759	7.18	0.000	.3577753	.6306099
neg_ea	.2169589	.0447002	4.85	0.000	.1284246	.3054933
mef_pos	.2961795	.1781749	1.66	0.099	-.0567184	.6490774
mef_neg	-.1513681	.0575153	-2.63	0.010	-.2652845	-.0374517
event_general	.110704	.0130685	8.47	0.000	.0848202	.1365878
AnaReport_D	.1241986	.0250257	4.96	0.000	.074632	.1737652
media_D	.3125132	.0188239	16.60	0.000	.2752302	.3497963
logSVI_All	1.177737	.1329927	8.86	0.000	.9143286	1.441146
logques_hudong	.2900238	.0192001	15.11	0.000	.2519956	.3280521
_cons	-21.58911	3.124372	-6.91	0.000	-27.77732	-15.40089

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5110	5110	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

[results\Table6.xls](#)

dir : [seeout](#)

(MWFE estimator converged in 4 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Miller applied.

HDFE Linear regression	Number of obs	=	596,989
Absorbing 2 HDFE groups	F(11, 116)	=	35.60
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.1410
	Adj R-squared	=	0.1334
Number of clusters (stkcd)	=	5,122	Within R-sq. = 0.0025
Number of clusters (date)	=	117	Root MSE = 4.8650

(Std. err. adjusted for 117 clusters in stkcd date)

vpin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.1029943	.0232926	4.42	0.000	.0568604	.1491282
dsmvtll	-1.224985	.3097267	-3.96	0.000	-1.838438	-.6115323
pos_ea	1.100717	.1114469	9.88	0.000	.8799828	1.321452
neg_ea	1.215821	.1350015	9.01	0.000	.9484338	1.483209
mef_pos	1.148083	.1746576	6.57	0.000	.8021517	1.494015
mef_neg	.0694942	.1168213	0.59	0.553	-.1618851	.3008734
event_general	-.0179238	.0250426	-0.72	0.476	-.0675237	.0316762
AnaReport_D	.1371823	.0601099	2.28	0.024	.0181272	.2562375
media_D	.2271514	.0265363	8.56	0.000	.174593	.2797099
logSVI_All	.3403229	.0689899	4.93	0.000	.2036797	.4769661
logques_hudong	.0964406	.0264024	3.65	0.000	.0441473	.1487339
_cons	35.08643	4.695143	7.47	0.000	25.78711	44.38575

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5122	5122	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

[results\Table6.xls](#)

dir : [seeout](#)

177 . restore

178 .

179 . *=====

180 . * Table 7: User Queries and Market Trading Conditional on Positive Answer Feedback

181 . *=====

182 .

183 . * Import firm-day level dataset

```

184 . use "$dataset\firm_day_dataset",clear

185 .
186 . * Define events and intermediaries
187 . global events pos_ea neg_ea mef_pos mef_neg event_general

188 . global intermediaries AnaReport_D media_D

189 .
190 . *****Panel A*****
191 .
192 . * Export the association between user queries and market trading outcome conditional on
    > positive answer feedback
193 . preserve

194 . foreach y in AbnSpread20to1 illiq AbnVol20to1 vpin {
    2.      winsor2 `y', replace
    3.      reghdfe `y' logspecific_q c.logspecific_q#c.posratio dsmvtll $events $intermedi
    > aries logSVI_All logques_hudong if tradingday == 1 , abs(stkcd date) cluster(stkcd date)
    > )
    4.
195 .      outreg2 using "$results\Table7_panels.xls", drop(o* 0*) ///
    >      append tstat tdec(2) nocons dec(3) ///
    >      addtext(Firm FE, YES, Date FE, YES)
    5. }
(dropped 25 singleton observations)
(MWFE estimator converged in 6 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.

```

HDFE Linear regression	Number of obs	=	149,148
Absorbing 2 HDFE groups	F(12, 116)	=	64.77
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.2731
	Adj R-squared	=	0.2468
Number of clusters (stkcd)	Within R-sq.	=	0.0364
Number of clusters (date)	Root MSE	=	2.7793

(Std. err. adjusted for 117 clusters in stkcd date)

AbnSpread20to1	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.1924885	.022938	8.39	0.000	.147057	.2379201
c.logspecific_q# c.posratio	.0718206	.0360554	1.99	0.049	.0004084	.1432329
dsmvtll	-4.382219	.3440629	-12.74	0.000	-5.063679	-3.700759
pos_ea	.3392311	.0805555	4.21	0.000	.1796807	.4987815
neg_ea	.2858311	.1226223	2.33	0.021	.0429621	.5287001
mef_pos	.3247818	.2123659	1.53	0.129	-.0958356	.7453991
mef_neg	1.198689	.2981148	4.02	0.000	.6082355	1.789143
event_general	.195215	.0289973	6.73	0.000	.1377822	.2526479
AnaReport_D	-.0077193	.0545959	-0.14	0.888	-.1158534	.1004149
media_D	.4676678	.0302188	15.48	0.000	.4078157	.5275199
logSVI_All	.3935224	.0974847	4.04	0.000	.2004416	.5866032
logques_hudong	.1532577	.0247987	6.18	0.000	.1041407	.2023747
_cons	68.45261	5.362545	12.76	0.000	57.83142	79.07381

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5090	5090	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

results\Table7 panela.xls

dir : seeout

(dropped 24 singleton observations)

(MWFE estimator converged in 6 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Miller applied.

HDFE Linear regression	Number of obs	=	150,309
Absorbing 2 HDFE groups	F(12, 116)	=	40.41
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.5708
	Adj R-squared	=	0.5554
Number of clusters (stkcd)	=	5,093	Within R-sq. = 0.0609
Number of clusters (date)	=	117	Root MSE = 3.2814

(Std. err. adjusted for 117 clusters in stkcd date)

illiq	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.1505051	.0301441	4.99	0.000	.0908009	.2102092
c.logspecific_q# c.posratio	.069547	.0342919	2.03	0.045	.0016275	.1374665
dsmvtll	-7.051856	.5391287	-13.08	0.000	-8.119669	-5.984044
pos_ea	.3137422	.1351449	2.32	0.022	.0460707	.5814136
neg_ea	.0047479	.116874	0.04	0.968	-.2267358	.2362315
mef_pos	.2416292	.4568743	0.53	0.598	-.663268	1.146526
mef_neg	1.387883	.2736802	5.07	0.000	.845825	1.929941
event_general	.1198135	.0353294	3.39	0.001	.0498392	.1897878
AnaReport_D	-.1490674	.0512318	-2.91	0.004	-.2505385	-.0475964
media_D	.3408615	.030419	11.21	0.000	.2806129	.4011101
logSVI_All	.0567024	.0746386	0.76	0.449	-.0911287	.2045336
logques_hudong	.0596748	.0313941	1.90	0.060	-.0025053	.1218549
_cons	116.3629	8.550825	13.61	0.000	99.42689	133.2989

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5093	5093	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

results\Table7 panela.xls

dir : seeout

(dropped 24 singleton observations)

(MWFE estimator converged in 6 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Miller applied.

HDFE Linear regression	Number of obs	=	150,177
Absorbing 2 HDFE groups	F(12, 116)	=	95.83
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.2007
	Adj R-squared	=	0.1720
Number of clusters (stkcd)	=	5,081	Within R-sq. = 0.1092
Number of clusters (date)	=	117	Root MSE = 2.3534

(Std. err. adjusted for 117 clusters in stkcd date)

AbnVol20to1	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.23255	.022836	10.18	0.000	.1873203	.2777796
c.logspecific_q# c.posratio	.0277719	.0357677	0.78	0.439	-.0430705	.0986143
dsmvtll	2.022262	.2593626	7.80	0.000	1.508561	2.535962
pos_ea	.2907997	.0780841	3.72	0.000	.1361442	.4454551
neg_ea	-.0605863	.1048381	-0.58	0.564	-.2682314	.1470588
mef_pos	.2617864	.2727574	0.96	0.339	-.2784439	.8020168
mef_neg	-.527828	.2254168	-2.34	0.021	-.9742944	-.0813617
event_general	.203217	.0264972	7.67	0.000	.150736	.2556981
AnaReport_D	.051564	.0321119	1.61	0.111	-.0120377	.1151657
media_D	.4394914	.0336689	13.05	0.000	.3728059	.5061768
logSVI_All	1.616778	.1992297	8.12	0.000	1.222179	2.011378
logques_hudong	.3268207	.031452	10.39	0.000	.2645262	.3891153
_cons	-43.29199	3.877923	-11.16	0.000	-50.9727	-35.61127

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5081	5081	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

results\Table7 panela.xls

dir : seeout

(dropped 24 singleton observations)

(MWFE estimator converged in 6 iterations)

Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Miller applied.

HDFE Linear regression	Number of obs	=	150,309
Absorbing 2 HDFE groups	F(12, 116)	=	24.89
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.1506
	Adj R-squared	=	0.1200
Number of clusters (stkcd)	=	5,093	Within R-sq. = 0.0046
Number of clusters (date)	=	117	Root MSE = 4.9254

(Std. err. adjusted for 117 clusters in stkcd date)

vpin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
logspecific_q	.0775897	.0299255	2.59	0.011	.0183185	.1368609
c.logspecific_q# c.posratio	.1909569	.0704734	2.71	0.008	.0513754	.3305383
dsmvtll	-.9858651	.3365113	-2.93	0.004	-1.652368	-.3193621
pos_ea	1.18167	.1752364	6.74	0.000	.8345925	1.528748
neg_ea	1.052607	.2138346	4.92	0.000	.6290804	1.476133
mef_pos	1.091861	.3869405	2.82	0.006	.3254763	1.858245
mef_neg	.6025516	.2936409	2.05	0.042	.0209588	1.184144
event_general	.032883	.0460052	0.71	0.476	-.058236	.1240021
AnaReport_D	.1448581	.0795586	1.82	0.071	-.0127178	.302434
media_D	.4302375	.0434257	9.91	0.000	.3442273	.5162476
logSVI_All	.452043	.0892274	5.07	0.000	.2753169	.6287691
logques_hudong	.0922352	.0388071	2.38	0.019	.0153728	.1690976
_cons	30.24193	5.274103	5.73	0.000	19.79591	40.68796

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5093	5093	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

results\Table7_panela.xls

dir : seeout

196 . restore

197 .

198 .

199 . *****Panel B*****

200 .

201 . * Export the association between overall answer sentiment and abnormal returns and sent
> iment effects conditional on feedback type (Positive vs. Non-Positive FB)

202 . * tsc_avg_sent: Sentiment for positive FB; n_tsc_avg_sent: Sentiment for non-positive F
> B

203 . preserve

204 . winsor2 AbnRet,cut(1 99) replace

205 . reghdfe AbnRet average_sentiment_score dsmvtll \$events \$intermediaries logSVI_All logqu
> es_hudong , abs(stkcd date) cluster(stkcd date)
(dropped 36 singleton observations)
(MWFE estimator converged in 6 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.

HDFE Linear regression	Number of obs	=	142,211
Absorbing 2 HDFE groups	F(11, 116)	=	63.64
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.0923
	Adj R-squared	=	0.0580
Number of clusters (stkcd)	=	5,064	Within R-sq.
Number of clusters (date)	=	117	Root MSE
			= 3.0083

(Std. err. adjusted for 117 clusters in stkcd date)

AbnRet	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
average_sentiment_sc~e	.6012466	.0465784	12.91	0.000	.5089923	.6935009
dsmvtll	1.586652	.3998743	3.97	0.000	.7946505	2.378653
pos_ea	.0780302	.1090304	0.72	0.476	-.1379183	.2939787
neg_ea	-1.151682	.1850655	-6.22	0.000	-1.518228	-.7851365
mef_pos	1.739631	.4901079	3.55	0.001	.7689102	2.710351
mef_neg	-2.096628	.2917108	-7.19	0.000	-2.674398	-1.518858
event_general	.0522808	.0289702	1.80	0.074	-.0050984	.10966
AnaReport_D	.2314879	.0690998	3.35	0.001	.0946271	.3683488
media_D	.5407499	.0356527	15.17	0.000	.4701353	.6113645
logSVI_All	-.1550142	.070035	-2.21	0.029	-.2937274	-.016301
logques_hudong	-.2614112	.0334861	-7.81	0.000	-.3277347	-.1950878
_cons	-24.9508	6.34556	-3.93	0.000	-37.51898	-12.38262

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	5064	5064	0 *
date	117	117	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
206 . outreg2 using "$results\Table7_panelb.xls", drop(o* 0*) append tstat tdec(2) nocons dec
> (3)addtext(Firm FE, YES, Date FE, YES)
results\Table7_panelb.xls
dir : seeout
```

```
207 . reghdfe AbnRet tsc_avg_sent n_tsc_avg_sent dsmvtll $events $intermediaries logSVI_All 1
> ogques_hudong , abs(stkcd date) cluster(stkcd date)
(dropped 1058 singleton observations)
(MWFE estimator converged in 7 iterations)
Warning: VCV matrix was non-positive semi-definite; adjustment from Cameron, Gelbach & Mi
> ller applied.
```

HDFE Linear regression	Number of obs	=	22,714
Absorbing 2 HDFE groups	F(12, 115)	=	20.07
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.2125
	Adj R-squared	=	0.1112
Number of clusters (stkcd)	=	2,463	
Number of clusters (date)	=	116	
	Within R-sq.	=	0.0252
	Root MSE	=	2.8073

(Std. err. adjusted for 116 clusters in stkcd date)

AbnRet	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
tsc_avg_sent	.8086706	.0863738	9.36	0.000	.6375808	.9797604
n_tsc_avg_sent	.4381852	.0977549	4.48	0.000	.2445515	.6318189
dsmvtll	1.653263	.5228767	3.16	0.002	.6175448	2.688981
pos_ea	-.050056	.2632518	-0.19	0.850	-.5715071	.4713951
neg_ea	-.8515995	.3840571	-2.22	0.029	-1.612343	-.0908562
mef_pos	2.231056	.8704774	2.56	0.012	.5068077	3.955304
mef_neg	-2.523725	1.154379	-2.19	0.031	-4.810328	-.2371217
event_general	.1729067	.0723903	2.39	0.019	.0295154	.3162979
AnaReport_D	.0905866	.0783668	1.16	0.250	-.0646429	.2458162
media_D	.481955	.0608691	7.92	0.000	.361385	.6025249
logSVI_All	-.0448673	.1425136	-0.31	0.753	-.3271593	.2374247
logques_hudong	-.2972576	.0593037	-5.01	0.000	-.4147269	-.1797883
_cons	-28.16023	8.644242	-3.26	0.001	-45.28281	-11.03765

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
stkcd	2463	2463	0 *
date	116	116	0 *

* = FE nested within cluster; treated as redundant for DoF computation

```
208 . lincom tsc_avg_sent - n_tsc_avg_sent
```

(1) tsc_avg_sent - n_tsc_avg_sent = 0

AbnRet	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
(1)	.3704854	.1382975	2.68	0.008	.0965448	.6444261

```
209 . outreg2 using "$results\Table7_panelb.xls", drop(o* 0*) append tstat tdec(2) nocons dec
> (3)addtext(Firm FE, YES, Date FE, YES)
results\Table7_panelb.xls
dir : seeout
```

```
210 . restore
```

```
211 .
end of do-file
```

```
212 .
```