



```

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

2 . doedit "01_process_csmar_daily.do"

3 . do "C:\Users\eliven\Dropbox\ELLW_2026\code\01_process_csmar_daily.do"

4 . *****
5 . ***# 0.Defining Default Paths
6 .
7 . global path "datasets\daily"

8 . global rawdata "$path\rawdata"

9 . global data "$path\data"

10 .
11 . *****
12 . *****
13 . ***# 1. importing rawdata
14 . // 1.1 Daily Return
15 . import excel "$rawdata\日个股回报率文件2224\TRD_Dalyr.xlsx", sheet("sheet1") firstrow c
   > ase(lower) clear
   (24 vars, 1,000,002 obs)

16 . labone, nrow(1) // ssc install labone

17 . drop in 1/2
   (2 observations deleted)

18 . destring _all,replace
   stkcd: all characters numeric; replaced as long
   trddt: contains nonnumeric characters; no replace
   opnprc: all characters numeric; replaced as double
   hiprc: all characters numeric; replaced as double
   loprc: all characters numeric; replaced as double
   clsprc: all characters numeric; replaced as double
   dnshrtrd: all characters numeric; replaced as long
   dnvaltrd: all characters numeric; replaced as double
   dsmvosd: all characters numeric; replaced as double
   dsmvtll: all characters numeric; replaced as double
   dretwd: all characters numeric; replaced as double
   dretnd: all characters numeric; replaced as double
   adjprc wd: all characters numeric; replaced as double
   adjprcnd: all characters numeric; replaced as double
   markettype: all characters numeric; replaced as byte
   capchgdt: contains nonnumeric characters; no replace
   trdsta: all characters numeric; replaced as byte
   ahshrtrd_d: all characters numeric; replaced as long
   ahvaltrd_d: all characters numeric; replaced as double
   precloseprice: all characters numeric; replaced as double
   changeratio: all characters numeric; replaced as double
   limitdown: all characters numeric; replaced as double
   (127 missing values generated)
   limitup: all characters numeric; replaced as double
   (127 missing values generated)
   limitstatus: all characters numeric; replaced as byte
   (127 missing values generated)

```

```

19 .
20 . compress
    variable trddt was str12 now str10
    variable capchgdt was str24 now str10
    (16,000,000 bytes saved)

21 . save "$data\2224dret_1.dta",replace
    (file datasets\daily\data\2224dret_1.dta not found)
    file datasets\daily\data\2224dret_1.dta saved

22 . *****
23 . import excel "$rawdata\日个股回报率文件2224\TRD_Dalyr1.xlsx", sheet("sheet1") firstrow
    > case(lower) clear
    (24 vars, 1,000,002 obs)

24 . labone, nrow(1) // ssc install labone

25 . drop in 1/2
    (2 observations deleted)

26 . destring _all,replace
    stkcd: all characters numeric; replaced as long
    trddt: contains nonnumeric characters; no replace
    opnprc: all characters numeric; replaced as double
    hiprc: all characters numeric; replaced as double
    loprc: all characters numeric; replaced as double
    clsprc: all characters numeric; replaced as double
    dnshrtrd: all characters numeric; replaced as double
    dnvaltrd: all characters numeric; replaced as double
    dsmvosd: all characters numeric; replaced as double
    dsmvtll: all characters numeric; replaced as double
    dretwd: all characters numeric; replaced as double
    dretnd: all characters numeric; replaced as double
    adjprc wd: all characters numeric; replaced as double
    adjprcnd: all characters numeric; replaced as double
    markettype: all characters numeric; replaced as byte
    capchgdt: contains nonnumeric characters; no replace
    trdsta: all characters numeric; replaced as byte
    ahshrtrd_d: all characters numeric; replaced as long
    ahvaltrd_d: all characters numeric; replaced as double
    precloseprice: all characters numeric; replaced as double
    changeratio: all characters numeric; replaced as double
    limitdown: all characters numeric; replaced as double
    (1507 missing values generated)
    limitup: all characters numeric; replaced as double
    (1497 missing values generated)
    limitstatus: all characters numeric; replaced as byte
    (1508 missing values generated)

27 .
28 . compress
    variable trddt was str12 now str10
    variable capchgdt was str24 now str10
    (16,000,000 bytes saved)

29 . save "$data\2224dret_2.dta",replace
    (file datasets\daily\data\2224dret_2.dta not found)
    file datasets\daily\data\2224dret_2.dta saved

30 . *****

```

```

31 . import excel "$rawdata\日个股回报率文件2224\TRD_Dalyr2.xlsx", sheet("sheet1") firstrow
    > case(lower) clear
    (24 vars, 1,000,002 obs)

32 . labone, nrow(1) // ssc install labone

33 . drop in 1/2
    (2 observations deleted)

34 . destring _all,replace
    stkcd: all characters numeric; replaced as long
    trddt: contains nonnumeric characters; no replace
    opnprc: all characters numeric; replaced as double
    hiprc: all characters numeric; replaced as double
    loprc: all characters numeric; replaced as double
    clsprc: all characters numeric; replaced as double
    dnshrtrd: all characters numeric; replaced as double
    dnvaltrd: all characters numeric; replaced as double
    dsmvosd: all characters numeric; replaced as double
    dsmvtll: all characters numeric; replaced as double
    dretwd: all characters numeric; replaced as double
    dretnd: all characters numeric; replaced as double
    adjprc wd: all characters numeric; replaced as double
    adjprcnd: all characters numeric; replaced as double
    markettype: all characters numeric; replaced as byte
    capchgdt: contains nonnumeric characters; no replace
    trdsta: all characters numeric; replaced as byte
    ahshrtrd_d: all characters numeric; replaced as long
    ahvaltrd_d: all characters numeric; replaced as double
    precloseprice: all characters numeric; replaced as double
    changeratio: all characters numeric; replaced as double
    limitdown: all characters numeric; replaced as double
    (621 missing values generated)
    limitup: all characters numeric; replaced as double
    (621 missing values generated)
    limitstatus: all characters numeric; replaced as byte
    (694 missing values generated)

35 .
36 . compress
    variable trddt was str12 now str10
    variable capchgdt was str24 now str10
    (16,000,000 bytes saved)

37 . save "$data\2224dret_3.dta",replace
    (file datasets\daily\data\2224dret_3.dta not found)
    file datasets\daily\data\2224dret_3.dta saved

38 .
39 . *****
40 . import excel "$rawdata\日个股回报率文件2224\TRD_Dalyr3.xlsx", sheet("sheet1") firstrow
    > case(lower) clear
    (24 vars, 209,543 obs)

41 . labone, nrow(1) // ssc install labone

42 . drop in 1/2
    (2 observations deleted)

43 . destring _all,replace
    stkcd: all characters numeric; replaced as long
    trddt: contains nonnumeric characters; no replace
    opnprc: all characters numeric; replaced as double
    hiprc: all characters numeric; replaced as double
    loprc: all characters numeric; replaced as double
    clsprc: all characters numeric; replaced as double
    dnshrtrd: all characters numeric; replaced as long
    dnvaltrd: all characters numeric; replaced as double
    dsmvosd: all characters numeric; replaced as double
    dsmvtll: all characters numeric; replaced as double
    dretwd: all characters numeric; replaced as double
    dretnd: all characters numeric; replaced as double

```

```

adjprcnd: all characters numeric; replaced as double
adjprcnd: all characters numeric; replaced as double
markettype: all characters numeric; replaced as byte
capchgd: contains nonnumeric characters; no replace
trdsta: all characters numeric; replaced as byte
ahshrtrd_d: all characters numeric; replaced as long
ahvaltrd_d: all characters numeric; replaced as double
preclosetprice: all characters numeric; replaced as double
changeratio: all characters numeric; replaced as double
limitdown: all characters numeric; replaced as double
(1420 missing values generated)
limitup: all characters numeric; replaced as double
(1265 missing values generated)
limitstatus: all characters numeric; replaced as byte
(1420 missing values generated)

```

```

44 .
45 . compress
    variable trddt was str12 now str10
    variable capchgd was str24 now str10
    (3,352,656 bytes saved)

46 . save "$data\2224dret_4.dta",replace
    (file datasets\daily\data\2224dret_4.dta not found)
    file datasets\daily\data\2224dret_4.dta saved

47 .
48 . // appending daily return
49 . use "$data\2224dret_1.dta",clear

50 . append using "$data\2224dret_2.dta"
    (variable dnshrtrd was long, now double to accommodate using data's values)

51 . append using "$data\2224dret_3.dta"

52 . append using "$data\2224dret_4.dta"

53 . save "$data\2224dret.dta",replace
    file datasets\daily\data\2224dret.dta saved

54 .
55 . erase "$data\2224dret_1.dta"

56 . erase "$data\2224dret_2.dta"

57 . erase "$data\2224dret_3.dta"

58 . erase "$data\2224dret_4.dta"

59 .
60 . *****
61 . // 1.2 Earnings Announcement
62 . import excel "$rawdata\年中季报事件表000538951\ER_RelForcDate.xlsx", sheet("sheet1") fi
    > rstrow case(lower) clear
    (12 vars, 21,380 obs)

63 . labone, nrow(1) // ssc install labone

64 . drop in 1/2
    (2 observations deleted)

```

```

65 . destring _all,replace
    institutionid: all characters numeric; replaced as long
    symbol: all characters numeric; replaced as long
    securityid: all characters numeric; replaced as double
    shortname: contains nonnumeric characters; no replace
    reptyp: all characters numeric; replaced as byte
    accouperi: contains nonnumeric characters; no replace
    fullname: contains nonnumeric characters; no replace
    firrelforcdte: contains nonnumeric characters; no replace
    firchgrelforcdte: contains nonnumeric characters; no replace
    secchgrelforcdte: contains nonnumeric characters; no replace
    thrchgrelforcdte: contains nonnumeric characters; no replace
    actreldate: contains nonnumeric characters; no replace

66 .
67 . rename symbol stkcd

68 . rename accouperi qtr_end

69 . rename actreldate date

70 . gen EA = 1

71 .
72 . keep stkcd reptyp qtr_end date EA

73 . compress
    variable EA was float now byte
    variable qtr_end was str18 now str10
    variable date was str15 now str10
    (342,048 bytes saved)

74 .
75 . save "$data\EA.dta",replace
    file datasets\daily\data\EA.dta saved

76 .
77 . *****
78 . // 1.3 Events (excluding quarterly report)
79 . import excel "$rawdata\公司公告事件表150217655\ER_Announcement.xlsx", sheet("sheet1") f
    > irstrow case(lower) clear
    (19 vars, 633,978 obs)

80 . labone, nrow(1) // ssc install labone

81 . drop in 1/2
    (2 observations deleted)

82 . destring _all,replace
    announcementid: all characters numeric; replaced as long
    institutionid: all characters numeric; replaced as long
    symbol: all characters numeric; replaced as long
    securityid: all characters numeric; replaced as double
    shortname: contains nonnumeric characters; no replace
    declaredate: contains nonnumeric characters; no replace
    fullname: contains nonnumeric characters; no replace
    title: contains nonnumeric characters; no replace
    classid: contains nonnumeric characters; no replace
    classname: contains nonnumeric characters; no replace
    classifyid: contains nonnumeric characters; no replace
    classifyname: contains nonnumeric characters; no replace
    filesize: all characters numeric; replaced as long
    industrycode: all characters numeric; replaced as byte
    (633976 missing values generated)
    industryname: all characters numeric; replaced as byte
    (633976 missing values generated)
    industrycode1: contains nonnumeric characters; no replace
    industryname1: contains nonnumeric characters; no replace
    areacode: all characters numeric; replaced as long
    (20681 missing values generated)
    area: contains nonnumeric characters; no replace

```

```

83 .
84 . rename symbol stkcd

85 . split classid, parse(,) generate(eventid)
variables created as string:
eventid1 eventid2 eventid3 eventid4

86 . destring eventid*,replace
eventid1: all characters numeric; replaced as byte
(245970 missing values generated)
eventid2: all characters numeric; replaced as byte
(616045 missing values generated)
eventid3: all characters numeric; replaced as byte
(633470 missing values generated)
eventid4: all characters numeric; replaced as byte
(633970 missing values generated)

87 .
88 . gen event_ea = (eventid1 == 2 | eventid2 == 2 | eventid3 == 2 | eventid4 == 2)

89 . gen event_general = (eventid1 != 2 & eventid2 != 2 & eventid3 != 2 & eventid4 != 2)

90 .
91 . gen event_financing = (eventid1 == 3 | eventid2 == 3 | eventid3 == 3 | eventid4 == 3 |
> eventid1 == 4 | eventid2 == 4 | eventid3 == 4 | eventid4 == 4 | eventid1 == 5 | eventid
> 2 == 5 | eventid3 == 5 | eventid4 == 5 |eventid1 == 6 | eventid2 == 6 | eventid3 == 6 |
> eventid4 == 6)

92 . gen event_merge = (eventid1 == 7 | eventid2 == 7 | eventid3 == 7 | eventid4 == 7)

93 . gen event_transaction = (eventid1 == 8 | eventid2 == 8 | eventid3 == 8 | eventid4 == 8)

94 . gen event_meeting = (eventid1 == 9 | eventid2 == 9 | eventid3 == 9 | eventid4 == 9)

95 . gen event_pledge = (eventid1 == 10 | eventid2 == 10 | eventid3 == 10 | eventid4 == 10)

96 . gen event_litigation = (eventid1 == 11 | eventid2 == 11 | eventid3 == 11 | eventid4 ==
> 11)

97 . gen event_director = (eventid1 == 12 | eventid2 == 12 | eventid3 == 12 | eventid4 == 12
> )

98 . gen event_emergency = (eventid1 == 13 | eventid2 == 13 | eventid3 == 13 | eventid4 == 1
> 3)

99 . gen event_misconduct = (eventid1 == 14 | eventid2 == 14 | eventid3 == 14 | eventid4 ==
> 14)

100 . gen event_transition = (eventid1 == 15 | eventid2 == 15 | eventid3 == 15 | eventid4 ==
> 15)

101 . gen event_donate = (eventid1 == 16 | eventid2 == 16 | eventid3 == 16 | eventid4 == 16)

102 . gen event_other = (eventid1 == 99 | eventid2 == 99 | eventid3 == 99 | eventid4 == 99 |
> eventid1 == . | eventid2 == . | eventid3 == . | eventid4 == .)

103 .
104 . ***
105 . local events financing merge transaction meeting pledge litigation director emergency m
> isconduct transition donate other general

```

```

106 .
107 . foreach ev of local events {
    2.      gen double event_`ev'_file = cond(event_`ev'==1, filesize, .)
    3. }
(601,715 missing values generated)
(600,033 missing values generated)
(592,163 missing values generated)
(572,791 missing values generated)
(633,975 missing values generated)
(630,996 missing values generated)
(633,972 missing values generated)
(633,906 missing values generated)
(633,974 missing values generated)
(632,352 missing values generated)
(633,782 missing values generated)
(5 missing values generated)
(23,352 missing values generated)

108 .
109 . foreach ev in financing merge transaction meeting pledge litigation director emergency
> misconduct transition donate other general {
    2.      replace event_`ev'_file = 0 if missing(event_`ev'_file)
    3. }
(601,715 real changes made)
(600,033 real changes made)
(592,163 real changes made)
(572,791 real changes made)
(633,975 real changes made)
(630,996 real changes made)
(633,972 real changes made)
(633,906 real changes made)
(633,974 real changes made)
(632,352 real changes made)
(633,782 real changes made)
(5 real changes made)
(23,352 real changes made)

110 . ***
111 .
112 . collapse ///
>      (sum)   event_ea event_general event_financing event_merge event_transaction ///
>             event_meeting event_pledge event_litigation event_director event_emergency
> ///
>             event_misconduct event_transition event_donate event_other ///
>      (mean)  event_financing_file event_merge_file event_transaction_file ///
>             event_meeting_file event_pledge_file event_litigation_file event_director_f
> ile ///
>             event_emergency_file event_misconduct_file event_transition_file ///
>             event_donate_file event_other_file event_general_file ///
>      (count) n_ann = filesize, ///
>      by(stkcd declaredate)

113 .
114 .
115 . global events "event_ea event_general event_financing event_merge event_transaction eve
> nt_meeting event_pledge event_litigation event_director event_emergency event_misconduct
> t event_transition event_donate event_other"

116 . foreach eve in $events{
    2.      replace `eve' = 1 if `eve' > 1
    3. }
(5,912 real changes made)
(79,405 real changes made)
(4,222 real changes made)
(3,684 real changes made)
(6,629 real changes made)
(19,118 real changes made)
(0 real changes made)
(68 real changes made)
(0 real changes made)
(1 real change made)
(0 real changes made)

```

(15 real changes made)  
 (12 real changes made)  
 (80,677 real changes made)

```

117 .
118 . label variable event_ea "Earnings Annoucement (periodic report)"

119 . label variable event_general "Events excluding EA"

120 . label variable event_financing "Events about Financing"

121 . label variable event_merge "Events about M&A"

122 . label variable event_transaction "Events about Transaction"

123 . label variable event_meeting "Events about Shareholder Meeting"

124 . label variable event_pledge "Events about Pledging"

125 . label variable event_litigation "Events about Lawsuits"

126 . label variable event_director "Events about Diretor Board"

127 . label variable event_emergency "Emergency"

128 . label variable event_misconduct "Misconduct"

129 . label variable event_transition "Director/Manager Transition"

130 . label variable event_donate "Events about Donation"

131 . label variable event_other "Other or Unclassified Events"

132 .
133 . rename declaredate date

134 .
135 . compress
    variable n_ann was long now byte
    variable event_ea was double now byte
    variable event_general was double now byte
    variable event_financing was double now byte
    variable event_merge was double now byte
    variable event_transaction was double now byte
    variable event_meeting was double now byte
    variable event_pledge was double now byte
    variable event_litigation was double now byte
    variable event_director was double now byte
    variable event_emergency was double now byte
    variable event_misconduct was double now byte
    variable event_transition was double now byte
    variable event_donate was double now byte
    variable event_other was double now byte
    variable event_pledge_file was double now int
    variable date was str12 now str10
    (21,320,291 bytes saved)

136 . save "$data\Events_new.dta",replace
    file datasets\daily\data\Events_new.dta saved
  
```



```

137 .
138 . *****
139 . // Revision: use the earnings briefings as EAs
140 . import excel "$rawdata\业绩快报表(季)122242701\FIN_F_QuiTraFin.xlsx", sheet("sheet1") f
    > instrow case(lower) clear
    (42 vars, 25,295 obs)

141 . labone, nrow(1) // ssc install labone

142 . drop in 1/2
    (2 observations deleted)

143 . qui destring _all,replace

144 . rename publirate date

145 . rename stockcode stkcd

146 . sort stkcd date numquitrafinreport

147 . bys stkcd date: keep if _n == _N
    (48 observations deleted)

148 . gen neg_bf = eps < epslast & epslast!= .

149 . gen pos_bf = eps >= epslast

150 . replace pos_bf = 0 if eps == .
    (41 real changes made)

151 . replace neg_bf = 0 if eps == .
    (0 real changes made)

152 . label variable neg_bf "Negative earnings briefings"

153 . label variable pos_bf "Postive earnings briefings"

154 . keep date stkcd pos_bf neg_bf

155 . compress
    variable neg_bf was float now byte
    variable pos_bf was float now byte
    variable date was str21 now str10
    (429,165 bytes saved)

156 . save "$data\Briefings.dta",replace
    file datasets\daily\data\Briefings.dta saved

157 .
158 . *****
159 . // 1.4 Managerial earnings forecast
160 . import excel "$rawdata\业绩预告表(季)235813487\FIN_F_ForecFin.xlsx", sheet("sheet1") fi
    > rstrow case(lower) clear
    (7 vars, 5,257 obs)

161 . labone, nrow(1) // ssc install labone

162 . drop in 1/2
    (2 observations deleted)

```

```

163 . destring _all,replace
    stockcode: all characters numeric; replaced as long
    source: all characters numeric; replaced as byte
    stockname: contains nonnumeric characters; no replace
    numforecfinreport: all characters numeric; replaced as byte
    accperiod: contains nonnumeric characters; no replace
    publidate: contains nonnumeric characters; no replace
    forecfinreporttype: contains nonnumeric characters; no replace

164 .
165 . rename stockcode stkcd

166 .
167 . gen mef = 1

168 . gen mef_pos = (forecfinreporttype == "大增" | forecfinreporttype == "扭亏" | forecfinre
    > porttype == "略增" | forecfinreporttype == "续盈")

169 . gen mef_neg = (forecfinreporttype == "大降" | forecfinreporttype == "略降" | forecfinre
    > porttype == "续亏" | forecfinreporttype == "转亏")

170 . gen mef_neu = (forecfinreporttype == "不确定")

171 . gen meg_extpos = forecfinreporttype == "大增"

172 . gen meg_extneg = forecfinreporttype == "大降"

173 .
174 . collapse (sum) mef mef_pos mef_neg mef_neu meg_extpos meg_extneg, by(stkcd publidate)

175 .
176 . label variable mef "Managerial earnings forecast"

177 . label variable mef_pos "Positive Managerial earnings forecast"

178 . label variable mef_neg "Negative Managerial earnings forecast"

179 . label variable mef_neu "Neutral Managerial earnings forecast"

180 .
181 . rename publidate date

182 . compress
    variable mef was double now byte
    variable mef_pos was double now byte
    variable mef_neg was double now byte
    variable mef_neu was double now byte
    variable meg_extpos was double now byte
    variable meg_extneg was double now byte
    variable date was str12 now str10
    (230,956 bytes saved)

183 . save "$data\Mef.dta",replace
    file datasets\daily\data\Mef.dta saved

184 .
185 . *****
186 . // 1.4 Analyst Report
187 . import excel "$rawdata\分析师预测指标文件213714289\AF_Forecast.xlsx", sheet("sheet1") f
    > firstrow case(lower) clear
    (4 vars, 124,161 obs)

```

```

188 . labone, nrow(1) // ssc install labone

189 . drop in 1/2
(2 observations deleted)

190 . destring _all,replace
    stkcd: all characters numeric; replaced as long
    rptdt: contains nonnumeric characters; no replace
    fenddt: contains nonnumeric characters; no replace
    reportid: all characters numeric; replaced as long

191 .
192 . // keep stkcd rptdt
193 . // duplicates drop
194 . rename rptdt date

195 . collapse (count) reportid, by(stkcd date)

196 . rename reportid AnaReport

197 .
198 . compress
    variable AnaReport was long now byte
    variable date was str12 now str10
    (150,225 bytes saved)

199 . save "$data\AnalystReports.dta",replace
    file datasets\daily\data\AnalystReports.dta saved

200 .
201 . *****
202 . // 1.5 Media Coverage
203 . import excel "$rawdata\新闻资讯信息表213959190\ER_NewsInfo.xlsx", sheet("sheet1") firs
    > trow case(lower) clear
    (5 vars, 1,000,002 obs)

204 . labone, nrow(1) // ssc install labone

205 . drop in 1/2
(2 observations deleted)

206 .
207 . replace symbol = "600018" if symbol == "600018,600018"
(57 real changes made)

208 . destring _all,replace
    newsid: all characters numeric; replaced as long
    declaredate: contains nonnumeric characters; no replace
    classifyid: all characters numeric; replaced as long
    classify: contains nonnumeric characters; no replace
    symbol: all characters numeric; replaced as long
    (656379 missing values generated)

209 . save "$data\MediaCoverageRaw1.dta",replace
    file datasets\daily\data\MediaCoverageRaw1.dta saved

210 .
211 . import excel "$rawdata\新闻资讯信息表213959190\ER_NewsInfo1.xlsx", sheet("sheet1") fir
    > strow case(lower) clear
    (5 vars, 252,687 obs)

```

```

212 . labone, nrow(1) // ssc install labone

213 . drop in 1/2
      (2 observations deleted)

214 . replace symbol = "600018" if symbol == "600018,600018"
      (14 real changes made)

215 . destring _all,replace
      newsid: all characters numeric; replaced as long
      declaredate: contains nonnumeric characters; no replace
      classifyid: all characters numeric; replaced as long
      classify: contains nonnumeric characters; no replace
      symbol: all characters numeric; replaced as long
      (177273 missing values generated)

216 . save "$data\MediaCoverageRaw2.dta",replace
      file datasets\daily\data\MediaCoverageRaw2.dta saved

217 .
218 . use "$data\MediaCoverageRaw1.dta",clear

219 . append using "$data\MediaCoverageRaw2.dta"

220 . // erase "$data\MediaCoverageRaw1.dta"
221 . // erase "$data\MediaCoverageRaw2.dta"
222 .
223 . rename symbol stkcd

224 . rename declaredate date

225 .
226 . replace date = substr(date,1,10)
      (1,252,685 real changes made)

227 .
228 . // keep stkcd date
229 . // duplicates drop
230 . // gen media = 1
231 . collapse (count) newsid, by(stkcd date)

232 . rename newsid media

233 . label variable media "Meida Coverage Today"

234 .
235 . compress
      variable media was long now int
      variable date was str19 now str10
      (2,267,375 bytes saved)

236 . save "$data\MediaCoverage.dta",replace
      file datasets\daily\data\MediaCoverage.dta saved

237 .
238 . *****
239 . // 1.6 Abnormal Volume
240 . // Prepare Market Volume
241 . import excel "$rawdata\日市场规模统计142342589\TRD_MKStructD.xlsx", sheet("sheet1") fir
      > strow case(lower) clear
      (14 vars, 17,695 obs)

```

```

242 . labone, nrow(1) // ssc install labone

243 . drop in 1/2
(2 observations deleted)

244 . destring _all,replace
sgndate: contains nonnumeric characters; no replace
datasgncode: all characters numeric; replaced as byte
datasgn: contains nonnumeric characters; no replace
listedconum: all characters numeric; replaced as int
(379 missing values generated)
stocknum: all characters numeric; replaced as int
totalshares: all characters numeric; replaced as double
(4452 missing values generated)
circulatedshares: all characters numeric; replaced as double
(2590 missing values generated)
marketvalue: all characters numeric; replaced as double
circulatedmktvalue: all characters numeric; replaced as double
volume: all characters numeric; replaced as double
(2188 missing values generated)
amount: all characters numeric; replaced as double
avgpe: all characters numeric; replaced as double
(15010 missing values generated)
turnoverrate1: all characters numeric; replaced as double
(14668 missing values generated)
turnoverrate2: all characters numeric; replaced as double
(14345 missing values generated)

245 .
246 . keep if datasgncode == 1 | datasgncode == 5
(14,971 observations deleted)

247 . keep sgndate datasgncode datasgn turnover2

248 . rename sgndate date

249 .
250 . compress
variable date was str12 now str10
variable datasgn was str16 now str9
(24,498 bytes saved)

251 . save "$data\MarketVolume.dta",replace
file datasets\daily\data\MarketVolume.dta saved

252 .
253 . // Prepare Stock Outstanding Shares
254 . import excel "$rawdata\股票历史日行情信息表(后复权)164159702\TRD_BwardQuotation.xlsx",
> sheet("sheet1") firstrow case(lower) clear
(4 vars, 1,000,002 obs)

255 . labone, nrow(1) // ssc install labone

256 . drop in 1/2
(2 observations deleted)

257 . destring _all,replace
tradingdate: contains nonnumeric characters; no replace
symbol: all characters numeric; replaced as long
circulatedshare: all characters numeric; replaced as double
acirculatedshare: all characters numeric; replaced as double

```

```

258 .
259 . rename symbol stkcd

260 . rename tradingdate date

261 . save "$data\SharesOutstanding1.dta",replace
    file datasets\daily\data\SharesOutstanding1.dta saved

262 .
263 . import excel "$rawdata\股票历史日行情信息表(后复权)164159702\TRD_BwardQuotation1.xlsx",
    > sheet("sheet1") firstrow case(lower) clear
    (4 vars, 1,000,002 obs)

264 . labone, nrow(1) // ssc install labone

265 . drop in 1/2
    (2 observations deleted)

266 . destring _all,replace
    tradingdate: contains nonnumeric characters; no replace
    symbol: all characters numeric; replaced as long
    circulatedshare: all characters numeric; replaced as double
    acirculatedshare: all characters numeric; replaced as double
    (6501 missing values generated)

267 .
268 . rename symbol stkcd

269 . rename tradingdate date

270 . save "$data\SharesOutstanding2.dta",replace
    file datasets\daily\data\SharesOutstanding2.dta saved

271 .
272 . import excel "$rawdata\股票历史日行情信息表(后复权)164159702\TRD_BwardQuotation2.xlsx",
    > sheet("sheet1") firstrow case(lower) clear
    (4 vars, 211,139 obs)

273 . labone, nrow(1) // ssc install labone

274 . drop in 1/2
    (2 observations deleted)

275 . destring _all,replace
    tradingdate: contains nonnumeric characters; no replace
    symbol: all characters numeric; replaced as long
    circulatedshare: all characters numeric; replaced as double
    acirculatedshare: all characters numeric; replaced as double
    (85422 missing values generated)

276 .
277 . rename symbol stkcd

278 . rename tradingdate date

279 . save "$data\SharesOutstanding3.dta",replace
    file datasets\daily\data\SharesOutstanding3.dta saved

280 .
281 . use "$data\SharesOutstanding1.dta",clear

```

```

282 . append using "$data\SharesOutstanding2.dta"
283 . append using "$data\SharesOutstanding3.dta"
284 . save "$data\SharesOutstanding.dta",replace
    file datasets\daily\data\SharesOutstanding.dta saved
285 .
286 . // erase "$data\SharesOutstanding1.dta"
287 . // erase "$data\SharesOutstanding2.dta"
288 . // erase "$data\SharesOutstanding3.dta"
289 .
290 . // stock volume
291 . use stkcd trddt dnrtrtd using "$data\2224dret.dta",clear

292 . rename trddt date

293 .
294 . // gen valid trading days
295 . preserve

296 . keep if date >= "2023-01-01"
    (1,171,038 observations deleted)

297 . keep date

298 . duplicates drop

    Duplicates in terms of all variables

    (2,038,115 observations deleted)

299 . gen trd_dt = _n

300 . label variable trd_dt "交易日序号"

301 . save "$data\TradingDays.dta",replace
    file datasets\daily\data\TradingDays.dta saved

302 . restore

303 .
304 . merge 1:1 stkcd date using "$data\SharesOutstanding.dta",nogen
    (variable date was str10, now str12 to accommodate using data's values)

```

Result	Number of obs
Not matched	1,343,672
from master	1,171,038
from using	172,634
Matched	2,038,503

```

305 . drop if circulatedshare == .
    (1,171,038 observations deleted)

306 .
307 . gen datasgncode = 1 if stkcd >= 600000 & stkcd < 700000
    (1,272,169 missing values generated)

```

```
308 . replace datasgncode = 5 if stkcd < 200000 | (stkcd >= 300000 & stkcd < 400000)
      (1,180,246 real changes made)
```

```
309 . drop if dnshttrd == .
      (172,634 observations deleted)
```

```
310 . merge m:1 datasgncode date using "$data\MarketVolume.dta",nogen
```

Result	Number of obs
Not matched	<b>86,912</b>
from master	<b>84,966</b>
from using	<b>1,946</b>
Matched	<b>1,953,537</b>

```
311 . drop if stkcd == .
      (1,946 observations deleted)
```

```
312 . sort stkcd date
```

```
313 .
```

```
314 . gen turnover = dnshttrd / acirculatedshare
      (84,966 missing values generated)
```

```
315 . label variable turnover "Stock Turnover (Free-Float Shares)"
```

```
316 . rename turnoverrate2 market_turnover
```

```
317 . label variable market_turnover "Market Turnover (Free-Float Shares)"
```

```
318 .
```

```
319 . merge m:1 date using "$data\TradingDays.dta",nogen keep(3)
```

Result	Number of obs
Not matched	<b>0</b>
Matched	<b>2,038,503</b>

```
320 . xtset stkcd trd_dt
```

```
Panel variable: stkcd (unbalanced)
Time variable: trd_dt, 1 to 388, but with gaps
Delta: 1 unit
```

```
321 . bys stkcd: asreg turnover market_turnover, wind(trd_dt -31 -1) min(20)
```

```
322 . gen AbnVol = turnover - (_b_cons + _b_market_turnover * market_turnover)
      (194,828 missing values generated)
```

```
323 . label variable AbnVol "Daily abnormal trading volume, reg wind(-30,-1)"
```

```
324 .
```

```
325 . * the second method of calculation, to align with the AbnSpread measure
```

```
326 . rangestat (mean) turnover, interval(trd_dt -20 -1) by(stkcd)
```

```
327 . gen AbnVol20to1 = turnover - turnover_mean
      (90,187 missing values generated)
```



```

328 . label variable AbnVol20to1 "Daily abnormal trading volume, demeaned(-20,-1)"

329 . drop if date < "2023-09-01"
      (834,713 observations deleted)

330 .
331 . keep stkcd date market_turnover turnover AbnVol*

332 .
333 . compress
      variable date was str12 now str10
      (2,407,580 bytes saved)

334 . save "$data\AbnVol.dta",replace
      file datasets\daily\data\AbnVol.dta saved

335 .
336 .
337 . *****
338 . // 1.7 Abnormal Return
339 . // Prepare Market Return
340 . import excel "$rawdata\综合日市场回报率文件144257496\TRD_Cndalym.xlsx", sheet("sheet1")
      > firstrow case(lower) clear
      (11 vars, 2,724 obs)

341 . labone, nrow(1) // ssc install labone

342 . drop in 1/2
      (2 observations deleted)

343 . destring _all,replace
      markettype: all characters numeric; replaced as byte
      trddt: contains nonnumeric characters; no replace
      cnshrtrdtl: all characters numeric; replaced as double
      cnvaltrdtl: all characters numeric; replaced as double
      cdretwdeq: all characters numeric; replaced as double
      cdretmdeq: all characters numeric; replaced as double
      cdretwdos: all characters numeric; replaced as double
      cdretmdos: all characters numeric; replaced as double
      cdretwdtl: all characters numeric; replaced as double
      cdretmdtl: all characters numeric; replaced as double
      cdnstkcal: all characters numeric; replaced as int

344 .
345 . rename trddt date

346 . // markettype: 33 SH, 20 SZ
347 . compress
      variable date was str12 now str10
      (5,444 bytes saved)

348 . save "$data\MarketReturn.dta",replace
      file datasets\daily\data\MarketReturn.dta saved

349 .
350 . // stock return
351 . use stkcd trddt hiprc loprc clsprc dnaltrd dretwd precloseprice using "$data\2224dret.
      > dta",clear

352 . rename trddt date

```

```

353 .
354 . // calculate bid-ask spread
355 . keep if date >= "2023-01-01"
      (1,171,038 observations deleted)

356 . merge m:1 date using "$data\TradingDays.dta",nogen keep(3)

      Result                                Number of obs
      -----                                -
      Not matched                            0
      Matched                               2,038,503

357 . xtset stkcd trd_dt

      Panel variable: stkcd (unbalanced)
      Time variable: trd_dt, 1 to 388, but with gaps
      Delta: 1 unit

358 . gen l1_hiprc = l.hiprc
      (5,939 missing values generated)

359 . gen l1_loprc = l.loprc
      (5,939 missing values generated)

360 .
361 . egen two_high = rowmax(hiprc l1_hiprc)

362 . egen two_low = rowmin(loprc l1_loprc)

363 .
364 . gen overnight_inc = max(0, loprc - precloseprice)

365 . gen overnight_dec = max(0, precloseprice - hiprc)

366 .
367 . gen adj_high = hiprc- overnight_inc + overnight_dec

368 . gen adj_low = lopr - overnight_inc + overnight_dec

369 .
370 . gen adj_two_high = max(l1_hiprc, adj_high)

371 . gen adj_two_low = min(l1_loprc, adj_low )

372 .
373 . gen beta = (log(adj_high / adj_low ))^2 + (log(hiprc / loprc))^2

374 . gen gamm = (log(adj_two_high / adj_two_low ))^2

375 . gen alph = (sqrt(2 * beta) - sqrt(beta)) / (3 - 2 * sqrt(2)) - sqrt(gamm / (3 - 2 * sqrt(2)))

376 . gen spread = 2*(exp(alph) - 1) / (1 + exp(alph))

377 . label variable spread "Bid-ask Spread, Corwin and Schultz (2012)"

378 .
379 . // calculate Amihud (2002) illiquidity measure
380 . // with trading value scaled by 1,000,000 to avoid extremely small magnitudes

```

```

381 . gen amihud = abs(dretwd) / (dnvaltrd / 1000000)
382 . label variable amihud "Amihud (2002) illiquidity measure"
383 .
384 . keep stkcd date trd_dt hiprc loprc clsprc dretwd precloseprice spread amihud
385 .
386 . gen markettype = 33 if stkcd >= 600000 & stkcd < 700000
    (1,172,218 missing values generated)
387 . replace markettype = 20 if stkcd < 200000 | (stkcd >= 300000 & stkcd < 400000)
    (1,087,252 real changes made)
388 .
389 . merge m:1 date markettype using "$data\MarketReturn.dta",nogen

```

Result	Number of obs
Not matched	<b>86,912</b>
from master	<b>84,966</b>
from using	<b>1,946</b>
Matched	<b>1,953,537</b>

```

390 .
391 . xtset stkcd trd_dt

Panel variable: stkcd (unbalanced)
Time variable: trd_dt, 1 to 388, but with gaps
Delta: 1 unit

392 . bys stkcd: asreg dretwd cdretwdos, wind(trd_dt -31 -1) min(20)
393 .
394 . gen AbnRet = dretwd - (_b_cons + _b_cdretwdo * cdretwdos)
    (191,164 missing values generated)
395 . label variable AbnRet "Daily abnormal trading return, reg wind(-30,-1)"
396 . drop _Nobs _R2 _adjR2 _b_cdretwdos _b_cons
397 .
398 . xtset stkcd trd_dt

Panel variable: stkcd (unbalanced)
Time variable: trd_dt, 1 to 388, but with gaps
Delta: 1 unit

399 . gen m_adj_ret = dretwd - cdretwdos
    (86,912 missing values generated)
400 . bys stkcd (trd_dt): gen sum_ret = sum(m_adj_ret)
401 . gen CAR60_madj = l.sum_ret - l61.sum_ret
    (332,028 missing values generated)
402 . label variable CAR60_madj "Cumulative Market Adjusted Return, -60 to -1"

```

```

403 .
404 . xtset stkcd trd_dt

      Panel variable: stkcd (unbalanced)
      Time variable: trd_dt, 1 to 388, but with gaps
      Delta: 1 unit

405 . bys stkcd (trd_dt): gen sum_ar = sum(AbnRet)

406 . gen CAR60_md1 = l.sum_ar - l61.sum_ar
      (332,028 missing values generated)

407 . label variable CAR60_md1 "Cumulative Abnormal(Market Model) Return, -60 to -1"

408 .
409 . drop if date < "2023-09-01"
      (836,657 observations deleted)

410 . keep stkcd date dretwd spread amihud AbnRet CAR60_madj CAR60_md1 cdretwdos

411 .
412 . // measure in percent
413 . global to_percent "dretwd amihud AbnRet CAR60_madj CAR60_md1 cdretwdos"

414 . foreach var in $to_percent{
      2.         replace `var' = `var' * 100
      3. }
      (1,170,990 real changes made)
      (1,170,990 real changes made)
      (1,147,511 real changes made)
      (1,141,381 real changes made)
      (1,141,377 real changes made)
      (1,149,693 real changes made)

415 .
416 . drop if stkcd == .
      (2 observations deleted)

417 . compress
      (0 bytes saved)

418 . save "$data\AbnRet.dta",replace
      file datasets\daily\data\AbnRet.dta saved

419 .
420 . *****
421 . // 1.8 Standardized unexpected earnings
422 . import excel "$rawdata\利润表144637035\FS_Comins.xlsx", sheet("sheet1") firstrow case(1
      > over) clear
      (8 vars, 109,922 obs)

423 . labone, nrow(1) // ssc install labone

424 . drop in 1/2
      (2 observations deleted)

425 . destring _all,replace
      stkcd: all characters numeric; replaced as long
      shortname: contains nonnumeric characters; no replace
      accper: contains nonnumeric characters; no replace
      typrep: contains nonnumeric characters; no replace
      b001300000: all characters numeric; replaced as double
      b001000000: all characters numeric; replaced as double
      b002000000: all characters numeric; replaced as double
      b006000000: all characters numeric; replaced as double
      (171 missing values generated)

```

```

426 .
427 . gen quarter = month(date(accper,"YMD")) / 3

428 . drop if quarter < 1
      (19,962 observations deleted)

429 . sort stkcd accper

430 .
431 . global earnings "b001300000 b001000000 b002000000 b006000000"

432 . foreach var in $earnings{
      2.         replace `var' = `var' - `var'[_n-1] if stkcd == stkcd[_n-1] & (quarter == qu
> arter[_n-1] + 1) & quarter > 1
      3. }
      (60,143 real changes made)
      (60,143 real changes made)
      (60,143 real changes made)
      (60,077 real changes made, 272 to missing)

433 .
434 . sort stkcd quarter accper

435 . foreach var in $earnings{
      2.         gen d`var' = `var' - `var'[_n-1] if stkcd == stkcd[_n-1] & (quarter == quart
> er[_n-1])
      3. }
      (21,829 missing values generated)
      (21,829 missing values generated)
      (21,829 missing values generated)
      (22,383 missing values generated)

436 . label variable db001300000 "Unexpected营业利润"

437 . label variable db001000000 "Unexpected利润总额"

438 . label variable db002000000 "Unexpected净利润"

439 . label variable db006000000 "Unexpected综合收益"

440 .
441 . rename accper qtr_end

442 . merge 1:1 stkcd qtr_end using "$data\EA.dta",nogen

```

Result	Number of obs
Not matched	<b>78,844</b>
from master	<b>73,712</b>
from using	<b>5,132</b>
Matched	<b>16,246</b>

```

443 .
444 . keep if EA == 1
      (73,712 observations deleted)

445 . keep stkcd qtr_end db001300000 db001000000 db002000000 db006000000 reptyp date EA

```

```

446 .
447 . gen trddt = date
      (5,222 missing values generated)

448 . drop if date == ""
      (5,222 observations deleted)

449 .
450 . // Annual and Q1 reports may be released on the same trading day;
451 . // drop duplicates by firm-date and retain the annual report
452 . sort stkcd qtr_end

453 . duplicates drop stkcd trddt,force

      Duplicates in terms of stkcd trddt

      (2,613 observations deleted)

454 .
455 . merge 1:1 stkcd trddt using "$data\2224dret.dta", nogen keepus(stkcd trddt dsmvtl1)

```

Result	Number of obs
Not matched	<b>3,203,004</b>
from master	<b>3,503</b>
from using	<b>3,199,501</b>
Matched	<b>10,040</b>

```

456 . sort stkcd trddt

457 .
458 . // Note: Market value is reported in thousand RMB
459 . gen SUE_operearn = db001300000 / (dsmvtl1[_n-60] * 1000) if stkcd[_n-60] == stkcd
      (3,200,111 missing values generated)

460 . gen SUE_totalearn = db001000000 / (dsmvtl1[_n-60] * 1000) if stkcd[_n-60] == stkcd
      (3,200,111 missing values generated)

461 . gen SUE_netearn = db002000000 / (dsmvtl1[_n-60] * 1000) if stkcd[_n-60] == stkcd
      (3,200,111 missing values generated)

462 . gen SUE_comearn = db006000000 / (dsmvtl1[_n-60] * 1000) if stkcd[_n-60] == stkcd
      (3,200,229 missing values generated)

463 .
464 . drop if EA == .
      (3,199,501 observations deleted)

465 .
466 . //check
467 . label variable SUE_operearn "SUE (Operating Income)"

468 . label variable SUE_totalearn "SUE (Total Profit)"

469 . label variable SUE_netearn "SUE (Net Income)"

470 . label variable SUE_comearn "SUE (Comprehensive Income)"

```

```

471 .
472 . keep stkcd qtr_end date SUE_operearn SUE_totalearn SUE_netearn SUE_comearn

473 . compress
    variable qtr_end was str18 now str10
    (108,344 bytes saved)

474 . save "$data\SUE.dta",replace
    file datasets\daily\data\SUE.dta saved

475 .
476 . *****
477 . // 1.9 other liquidity measures
478 . import delimited "$rawdata\个股Amihud指标表(日)132457775\LIQ_AMIHU_D.csv", clear
    (encoding automatically selected: UTF-8)
    (3 vars, 1,000,000 obs)

479 . rename trddt date

480 . // markettype: 33 SH, 20 SZ
481 . compress
    (0 bytes saved)

482 . save "$data\Amihud.dta",replace
    file datasets\daily\data\Amihud.dta saved

483 .
484 . import delimited "$rawdata\个股Amihud指标表(日)132457775\LIQ_AMIHU_D1.csv", clear
    (encoding automatically selected: UTF-8)
    (3 vars, 349,828 obs)

485 . rename trddt date

486 . // markettype: 33 SH, 20 SZ
487 . compress
    (0 bytes saved)

488 . append using "$data\Amihud.dta"

489 . label variable illiq "Amihud from CSMAR"

490 . save "$data\Amihud.dta",replace
    file datasets\daily\data\Amihud.dta saved

491 .
492 . * jump ratio
493 . import delimited "$rawdata\个股跳跃指标表(日)131915531\HF_StockJump.csv", clear
    (encoding automatically selected: UTF-8)
    (4 vars, 1,000,000 obs)

494 . rename trddt date

495 . compress
    (0 bytes saved)

496 . save "$data\jump.dta",replace
    file datasets\daily\data\jump.dta saved

497 .
498 . forvalues i = 1(1)3 {
    2. import delimited "$rawdata\个股跳跃指标表(日)131915531\HF_StockJump`i'.csv", clea
    > r
    3. rename trddt date
    4. append using "$data\jump.dta"
    5. compress
    6. save "$data\jump.dta",replace
    7. }
    (encoding automatically selected: UTF-8)
    (4 vars, 1,000,000 obs)
    (0 bytes saved)
    file datasets\daily\data\jump.dta saved
    (encoding automatically selected: UTF-8)

```

```
(4 vars, 1,000,000 obs)
(0 bytes saved)
file datasets\daily\data\jump.dta saved
(encoding automatically selected: UTF-8)
(4 vars, 928,377 obs)
(0 bytes saved)
file datasets\daily\data\jump.dta saved
```

```
499 .
500 . collapse (max) isjump , by(stkcd date)
```

```
501 . save "$data\jump.dta",replace
file datasets\daily\data\jump.dta saved
```

```
502 .
503 . * bid-ask spread
504 . import delimited "$rawdata\个股买卖价差表(日)132559606\HF_Spread.csv", clear
(encoding automatically selected: UTF-8)
(6 vars, 1,000,000 obs)
```

```
505 . rename trddt date
```

```
506 . compress
(0 bytes saved)
```

```
507 . save "$data\spread_csmar.dta",replace
file datasets\daily\data\spread_csmar.dta saved
```

```
508 .
509 . import delimited "$rawdata\个股买卖价差表(日)132559606\HF_Spread1.csv", clear
(encoding automatically selected: UTF-8)
(6 vars, 340,569 obs)
```

```
510 . rename trddt date
```

```
511 . compress
(0 bytes saved)
```

```
512 . append using "$data\spread_csmar.dta"
```

```
513 .
514 . merge m:1 date using "$data\TradingDays.dta",nogen keep(3)
```

Result	Number of obs
Not matched	0
Matched	1,021,382

```
515 . xtset stkcd trd_dt
```

```
Panel variable: stkcd (unbalanced)
Time variable: trd_dt, 200 to 388, but with gaps
Delta: 1 unit
```

```
516 . rangestat (mean) esp_amount, interval(trd_dt -20 -1) by(stkcd)
```

```
517 . gen AbnSpread20to1 = esp_amount - esp_amount_mean
(5,491 missing values generated)
```



```

518 . label variable AbnSpread20to1 "Daily abnormal bid-ask spread, demeaned(-20,-1)"

519 . save "$data\spread_csmar.dta",replace
    file datasets\daily\data\spread_csmar.dta saved

520 .
521 . * PIN
522 . import delimited "$rawdata\个股知情交易概率指标表(日)102142108\HF_VPIN.csv", clear
    (encoding automatically selected: UTF-8)
    (4 vars, 1,000,000 obs)

523 . rename trddt date

524 . compress
    (0 bytes saved)

525 . save "$data\pin.dta",replace
    file datasets\daily\data\pin.dta saved

526 .
527 . forvalues i = 1(1)2 {
    2.   import delimited "$rawdata\个股知情交易概率指标表(日)102142108\HF_VPIN`i'.csv", c
    >   lear
    3.   rename trddt date
    4.   append using "$data\pin.dta"
    5.   compress
    6.   save "$data\pin.dta",replace
    7.   }
    (encoding automatically selected: UTF-8)
    (4 vars, 1,000,000 obs)
    (0 bytes saved)
    file datasets\daily\data\pin.dta saved
    (encoding automatically selected: UTF-8)
    (4 vars, 705,596 obs)
    (0 bytes saved)
    file datasets\daily\data\pin.dta saved

528 .
529 . sort stkcd date n

530 . bys stkcd date: keep if _n == 1
    (1,350,084 observations deleted)

531 . keep stkcd date vpin

532 . save "$data\pin.dta",replace
    file datasets\daily\data\pin.dta saved

533 .
534 . *****
535 . // 1.10 Periodical report
536 . use "$data\EA", replace

537 . gen date1 = date(date,"YMD")
    (5,222 missing values generated)

538 . drop date

539 . rename date1 date

```

```

540 . gen quarter_report = (reptyp < 4)

541 . preserve

542 . keep if quarter_report == 1
    (5,365 observations deleted)

543 . keep stkcd date quarter_report

544 . duplicates drop

    Duplicates in terms of all variables

    (2 observations deleted)

545 . save "$data\quarter_report.dta",replace
    file datasets\daily\data\quarter_report.dta saved

546 . restore

547 . keep if reptyp == 4
    (16,013 observations deleted)

548 . gen annual_report = 1

549 . keep stkcd date annual_report

550 . duplicates drop

    Duplicates in terms of all variables

    (0 observations are duplicates)

551 . save "$data\annual_report.dta",replace
    file datasets\daily\data\annual_report.dta saved

552 .
553 . *****
554 . // 1.11 Hudongyi: investor interaction platform
555 . import excel "$rawdata\投资者互动表(日)171702232\CMS_InvestorInteraction.xlsx", sheet("
    > sheet1") firstrow clear
    (8 vars, 1,000,002 obs)

556 . labone, nrow (1 2)

557 . drop in 1
    (1 observation deleted)

558 . drop in 1
    (1 observation deleted)

559 . gen date = date(substr(QuestionDate,1,10),"YMD")

560 . destring Rank, replace
    Rank: all characters numeric; replaced as int

561 . collapse (count) Rank, by(date Symbol)

562 . rename Rank ques_hudong

```

```

563 . save "$data\hungdong1.dta",replace
    file datasets\daily\data\hungdong1.dta saved

564 .
565 . import excel "$rawdata\投资者互动表(日)171702232\CMS_InvestorInteraction1.xlsx", sheet(
    > "sheet1") firstrow clear
    (8 vars, 1,000,002 obs)

566 . labone, nrow (1 2)

567 . drop in 1
    (1 observation deleted)

568 . drop in 1
    (1 observation deleted)

569 . gen date = date(substr(QuestionDate,1,10),"YMD")

570 . destring Rank, replace
    Rank: all characters numeric; replaced as int

571 . collapse (count) Rank, by(date Symbol)

572 . rename Rank ques_hudong

573 . save "$data\hungdong2.dta",replace
    file datasets\daily\data\hungdong2.dta saved

574 .
575 . import excel "$rawdata\投资者互动表(日)171702232\CMS_InvestorInteraction2.xlsx", sheet(
    > "sheet1") firstrow clear
    (8 vars, 95,547 obs)

576 . labone, nrow (1 2)

577 . drop in 1
    (1 observation deleted)

578 . drop in 1
    (1 observation deleted)

579 . gen date = date(substr(QuestionDate,1,10),"YMD")

580 . destring Rank, replace
    Rank: all characters numeric; replaced as byte

581 . collapse (count) Rank, by(date Symbol)

582 . rename Rank ques_hudong

583 . save "$data\hungdong3.dta",replace
    file datasets\daily\data\hungdong3.dta saved

584 .
585 . append using "$data\hungdong2.dta"

586 . append using "$data\hungdong1.dta"

587 . collapse (sum) ques_hudong, by(date Symbol)

```

```

588 . rename Symbol stkcd
589 . deststring stkcd, replace
      stkcd: all characters numeric; replaced as long
590 . save "$data\hudong_questions.dta",replace
      file datasets\daily\data\hudong_questions.dta saved
591 .
592 . // erase "$data\hungdong1.dta"
593 . // erase "$data\hungdong2.dta"
594 . // erase "$data\hungdong3.dta"
595 .
596 . *****
597 . *****
598 . **# 2. Merge Data
599 . use stkcd trddt dsmvtll using "$data\2224dret.dta",clear
600 . rename trddt date
601 . replace dsmvtll = log(dsmvtll)
      (3,209,541 real changes made)
602 . merge 1:1 stkcd date using "$data\AbnRet.dta", keep(3) nogen

```

Result	Number of obs
Not matched	0
Matched	1,203,790

```

603 .
604 . gen tradingday = 1
605 . merge 1:1 stkcd date using "$data\Events_new.dta",nogen

```

Result	Number of obs
Not matched	1,088,275
from master	1,048,233
from using	40,042
Matched	155,557

```

606 .
607 . global events "event_ea event_general event_financing event_merge event_transaction eve
> nt_meeting event_pledge event_litigation event_director event_emergency event_misconduct
> t event_transition event_donate event_other event_financing_file event_merge_file event
> _transaction_file event_meeting_file event_pledge_file event_litigation_file event_dire
> ctor_file event_emergency_file event_misconduct_file event_transition_file event_donate
> _file event_other_file event_general_file n_ann"
608 .
609 .
610 . merge 1:1 stkcd date using "$data\Mef.dta",nogen

```

Result	Number of obs
Not matched	1,238,667
from master	1,238,625
from using	42
Matched	5,207

611 . merge 1:1 stkcd date using "\$data\AnalystReports.dta",nogen

Result	Number of obs
Not matched	1,225,333
from master	1,219,581
from using	5,752
Matched	24,293

612 . merge 1:1 stkcd date using "\$data\MediaCoverage.dta",nogen

Result	Number of obs
Not matched	1,068,475
from master	1,055,988
from using	12,487
Matched	193,638

613 . merge 1:1 stkcd date using "\$data\Amihud.dta",nogen

Result	Number of obs
Not matched	590,671
from master	251,478
from using	339,193
Matched	1,010,635

614 . merge 1:1 stkcd date using "\$data\jump.dta",nogen

Result	Number of obs
Not matched	292,331
from master	292,089
from using	242
Matched	1,309,217

615 . merge 1:1 stkcd date using "\$data\spread\_csma.dta",nogen

Result	Number of obs
Not matched	580,166
from master	580,166
from using	0
Matched	1,021,382

616 . merge 1:1 stkcd date using "\$data\pin.dta",nogen

Result	Number of obs
Not matched	256,920
from master	251,478
from using	5,442
Matched	1,350,070



624 .  
 625 . merge 1:1 stkcd date using "\$data\AbnVol.dta",nogen

Result	Number of obs
Not matched	427,053
from master	427,053
from using	0
Matched	1,203,790

626 . merge 1:1 stkcd date using "\$data\AbnRet.dta",nogen

Result	Number of obs
Not matched	427,053
from master	427,053
from using	0
Matched	1,203,790

627 . merge 1:1 stkcd date using "\$data\SUE.dta",nogen

Result	Number of obs
Not matched	1,617,334
from master	1,617,317
from using	17
Matched	13,526

628 .  
 629 .  
 630 . gen date2 = date(substr(date,1,10),"YMD")

631 . drop date

632 . rename date2 date

633 . merge 1:1 stkcd date using "\$data\annual\_report.dta",nogen

Result	Number of obs
Not matched	1,625,521
from master	1,625,508
from using	13
Matched	5,352

634 . merge 1:1 stkcd date using "\$data\quarter\_report.dta",nogen

Result	Number of obs
Not matched	1,625,260
from master	1,620,061
from using	5,199
Matched	10,812

635 . merge 1:1 stkcd date using "\$data\hudong\_questions.dta",nogen

Result	Number of obs
Not matched	2,239,952
from master	1,339,657
from using	900,295
Matched	296,415

636 . merge 1:1 stkcd date using "\$data\SVI.dta",nogen

Result	Number of obs
Not matched	2,262,071
from master	2,035,037
from using	227,034
Matched	501,330

637 .

638 . replace tradingday = 0 if tradingday == .  
(1,559,611 real changes made)

639 . keep stkcd date tradingday \$events \$coverage Abn\* spread amihud CAR60\_madj CAR60\_md1 qt  
> r\_end SUE\_operearn SUE\_totalearn SUE\_netearn SUE\_comearn dretwd cdretwdos dsmvtll qsp\_e  
> qual esp\_equal qsp\_amount esp\_amount isjump illiq vpin ques\_hudong SVI annual\_report po  
> s\_bf neg\_bf quarter\_report

640 . order stkcd date tradingday \$events \$coverage Abn\* spread amihud CAR60\_madj CAR60\_md1 q  
> tr\_end SUE\_operearn SUE\_totalearn SUE\_netearn SUE\_comearn dretwd cdretwdos dsmvtll qsp\_  
> equal esp\_equal qsp\_amount esp\_amount isjump illiq vpin ques\_hudong SVI annual\_report p  
> os\_bf neg\_bf quarter\_report

641 .

642 . sort stkcd date

643 .

644 . compress  
variable date was float now int  
variable tradingday was float now byte  
variable annual\_report was float now byte  
variable quarter\_report was float now byte  
variable ques\_hudong was double now int  
(46,977,817 bytes saved)

645 . save "\$path\daily.dta",replace  
file datasets\daily\daily.dta saved

646 .

end of do-file

647 .