

```

1  // =====
2  // STEP 1. INSTALL PACKAGE - Optional
3  // =====
4  // ssc install egenmore, replace
5  // ssc install asreg, replace
6  // ssc install tidy, replace
7
8  // =====
9  // STEP 2. CHANGE FILE LOADER
10 // =====
11 // *****
12 // 2.1 Raw data loader(read data from it)
13 // *****
14 cd D:\FF_3Factor\Data\Raw
15 // *****
16 // 2.2 Target data loader(save data to it)
17 // *****
18 global res_path D:\FF_3Factor\Data\Target
19
20 // =====
21 // STEP 3. DATA PRE PROCESS
22 // =====
23 // *****
24 // 3.1 generate rf
25 // *****
26 use RiskFreeRates.dta, clear // 无风险利率采用一年期定期存款利率
27 gen 交易月份=substr(交易日期, 1, 7) // 生成交易月份变量
28 collapse (mean) Rf=月度化无风险利率, by(交易月份) // 转化为月份数据
29 save Rf.dta, replace
30 // *****
31 // 3.2 generate rm
32 // *****
33 use TotalMarketReturn_month.dta, clear //
34 采用市值加权平均法计算, 考虑现金红利再投资的综合月度市场回报率
35 keep if 市场类型==21
36 save Rm.dta, replace
37 // *****
38 // 3.3 根据日交易数据判断当月是否 ST
39 // *****
40 use ReturenRate_day.dta, clear
41 gen 是否ST=inlist(交易状态, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15)
42 gen 是否PT=inlist(交易状态, 16)
43 gen 是否ST或PT=inlist(交易状态, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 16)
44 gen 交易月份=substr(交易日期, 1, 7)
45 collapse (max) 是否ST 是否PT 是否ST或PT, by(stkcd 交易月份) // 转化为月份数据
46 save ST_orNot.dta, replace
47 // *****
48 // 3.4
49 // GET BE 账面价值是t-1年12月底的账面所有者权益
50 // GET ME 市场价值是t-1年12底的市场总值
51 // *****
52 use BalanceSheet.dta, clear
53 keep stkcd year 资产总计 所有者权益合计
54 save BE.dta, replace
55
56 use ReturenRate_year.dta, clear
57 replace 年个股总市值=年个股总市值*1000 // 市值的单位是千元, 转化为元
58 replace 年个股流通市值=年个股流通市值*1000
59 keep stkcd year 年个股流通市值 年个股总市值
60 save ME.dta, replace
61 // =====
62 // STEP 4. RETURN DATA PROCESS
63 // =====
64 use ReturenRate_month.dta, clear
65 keep stkcd 证券代码 year 交易月份 月个股流通市值 月个股总市值 考虑现金红利再投资的月个股回报率
66 市场类型
67 merge m:1 交易月份 using Rf.dta, nogen keep(1 3)

```

```

67     merge m:1 交易月份 using Rm.dta, nogen keep(1 3) keepusing(
考虑现金红利再投资的综合月市场回报率流通市值加权平均法)
68     merge m:1 stkcd using CompanyFile.dta,  nogen keep(1 3) keepusing(上市日期 行业代码C)
69     merge 1:1 stkcd 交易月份 using ST_orNot.dta, nogen keep(1 3)
70     gen R_Rf=(考虑现金红利再投资的月个股回报率-Rf)*100 // 月超额收益率 百分比
71     drop if R_Rf==.
72     gen MKT=(考虑现金红利再投资的综合月市场回报率流通市值加权平均法-Rf)*100 // 市场因子 百分比
73     // *****
74     // 选取 t 年 5 月至 t+1 年 4 月作为组合
75     // *****
76     gen m = monthly(交易月份,"ym")
77     format m %tm
78     xtset stkcd m
79     gen group_year=year(dofm(m-4))
80     label var group_year 分组对应的年份
81
82     gen 上市月份= monthly(substr(上市日期, 1, 7),"YM") // 上市月份
83     format 上市月份 %tm
84
85     // 中国A股市场所有股票，包括沪深主板、中小板和创业板
86     // 市场类型数值含义： 1=上海A， 2=上海B， 4=深圳A， 8=深圳B， 16=创业板， 32=科创板。
87     keep if inlist(市场类型, 1, 4, 16)
88
89     // 选择2000-2019年数据
90     keep if year>=2000 & year<=2019
91
92     // 剔除IPO后前6个月的数据（包括上市月）
93     gen month_gap=m-上市月份
94     drop if month_gap<6
95
96     // 剔除ST、*ST、PT
97     drop if 是否ST或PT==1
98
99     // 剔除金融类股票
100    drop if regexm(行业代码C, "J")
101
102    save 收益率数据.dta, replace
103
104    // =====
105    // STEP 5. GENETRAT GROUP DATA
106    // =====
107    use 收益率数据.dta, clear
108    keep stkcd group_year
109    duplicates drop stkcd group_year, force
110    save temp.dta, replace
111
112    // 匹配市值和账面市值数据
113    use BE.dta, clear
114    merge m:1 stkcd year using ME.dta, nogen keep(1 3)
115
116    // 使用上一年末的数据来分组
117    replace year = year + 1
118
119    gen 市值=年个股流通市值
120    xtset stkcd year
121    replace 市值=L.市值 if 市值==.
122    gen 账面价值=所有者权益合计
123
124    // 删除账面价值为负的股票
125    drop if 账面价值 < 0
126    gen 账面市值比=账面价值/市值
127    keep stkcd year 市值 账面市值比
128    drop if 账面市值比==. | 市值==.
129    rename year group_year
130
131    // 剔除没有收益率数据的
132    merge 1:1 stkcd group_year using temp.dta, nogen keep(3)
133    save 分组数据.dta, replace

```

```

134
135 // =====
136 // STEP 6. GROUPING USE ME and BE/ME
137 // =====
138 use 分组数据.dta, clear
139
140 // 剔除市值最小的 30% 的股票 - Optional
141 // egen 市值10等分=xtile(市值), n(10) by(group_year)
142 // drop if 市值10等分<=3
143
144 // 规模的分组点为中位数
145 // 前50%为小规模组S
146 // 后50%为大规模组B
147 egen ME_group=xtile(市值), n(2) by(group_year)
148 gen 规模分组="S" if ME_group==1
149 replace 规模分组="B" if ME_group==2
150
151 // 账面市值比分组点为第30个和第70个百分位数
152 egen BM_group=xtile(账面市值比), n(10) by(group_year)
153 gen 账面市值比分组="L" if BM_group<=3
154 replace 账面市值比分组="M" if BM_group>3 & BM_group<=7
155 replace 账面市值比分组="H" if BM_group>7
156
157 // 分成25组
158 egen ME_group5=xtile(市值), n(5) by(group_year)
159 egen BM_group5=xtile(账面市值比), n(5) by(group_year)
160
161 keep stkcd group_year 规模分组 账面市值比分组 ME_group5 BM_group5 市值 账面市值比
162 save 分组结果.dta, replace
163
164
165 // =====
166 // STEP 7. CALCULATE AVERAGE RETURN OF 25 PORTIFOLIO
167 // =====
168 // 计算流通市值加权组合收益率
169 use 收益率数据.dta, clear
170 merge m:1 stkcd group_year using 分组结果.dta, nogen keep(3)
171 bys ME_group5 BM_group5 交易月份: egen 市值比例=pc(市值), prop
172 gen 加权组合收益率=R_RF*市值比例
173 collapse (sum) 加权组合收益率, by(ME_group5 BM_group5 交易月份) // 至此就没有个股了
174 save data.dta, replace
175
176 // 计算总市值比例
177 use 收益率数据.dta, clear
178 merge m:1 stkcd group_year using 分组结果.dta, nogen keep(3)
179 bys 交易月份: egen 月组合市值比例=pc(市值), prop
180 collapse (sum) 月组合市值比例, by(ME_group5 BM_group5 交易月份) // 至此就没有个股了
181 save data_size_pc.dta, replace
182
183 // 计算平均 Size 和 B.E
184 use 收益率数据.dta, clear
185 merge m:1 stkcd group_year using 分组结果.dta, nogen keep(3)
186 collapse (mean) 账面市值比 市值, by(ME_group5 BM_group5 交易月份) // 至此就没有个股了
187 save data_average_size_be.dta, replace
188
189 // 计算每月总 firm 量
190 use 收益率数据.dta, clear
191 merge m:1 stkcd group_year using 分组结果.dta, nogen keep(3)
192 gen n = 1
193 collapse (sum) n , by(ME_group5 BM_group5 交易月份) // 至此就没有个股了
194 save data_firm_num.dta, replace
195
196 merge 1:1 交易月份 ME_group5 BM_group5 using data.dta, nogen keep(3)
197 merge 1:1 交易月份 ME_group5 BM_group5 using data_size_pc.dta, nogen keep(3)
198 merge 1:1 交易月份 ME_group5 BM_group5 using data_average_size_be.dta, nogen keep(3)
199
200 collapse (mean) 加权组合收益率 月组合市值比例 账面市值比 n 市值 , by(ME_group5 BM_group5)
// 至此就没有个股了

```

```

201
202 // 计算 Big Small Newey-West t检验
203 use data.dta, clear
204 spread ME_group5 加权组合收益率
205
206 egen t=group(交易月份)
207 xtset BM_group5 t
208
209
210 * 滞后项根据  $q=4*(240/100)^{(2/9)}$  计算
211 *  $di\ 4*(240/100)^{(2/9)}$ 
212
213 * 验证第五组与第一组差异, 求newey t值
214 gen gap=加权组合收益率5-加权组合收益率1
215 forv i=1/5 {
216     qui newey gap if BM_group5==`i', lag(4)
217     est store r1_`i'
218 }
219 esttab r1_* , nogap replace keep(_cons) b(3) t(3) star(* 0.1 ** 0.05 *** 0.01)
220 esttab r1_* using $res_path/Newey-West_T检验1.rtf, nogap replace keep(_cons) b(3) t(3) star
(* 0.1 ** 0.05 *** 0.01)
221
222 // =====
223 // STEP 8. CONSTRUCT FACTOR
224 // =====
225
226 use 收益率数据.dta, clear
227 merge m:1 stkc d group_year using 分组结果.dta, nogen keep(3)
228
229 // 将市值和BM两个指标交叉2*3组
230 gen 组合=规模分组+账面市值比分组
231
232 // 计算各组合每一期的市值加权平均收益率
233 bys 组合 交易月份: egen p1=pc(市值), prop
234 bys 组合 交易月份: egen 加权月超额收益率=sum(p1*R_Rf)
235 collapse (mean) 加权月超额收益率 , by(组合 规模分组 账面市值比分组 交易月份 MKT)
236
237 // 生成变量便于计算SMB和HML
238 foreach i in SH SM SL BH BM BL {
239     bys 交易月份: egen `_i'=mean(加权月超额收益率) if 组合=="`i'"
240     bys 交易月份: egen `_i'=mean(_`i')
241     drop _`i'
242 }
243
244 keep 交易月份 MKT SH SM SL BH BM BL
245 duplicates drop 交易月份 MKT SH SM SL BH BM BL, force
246
247 // 计算SMB
248 gen SMB=(SH+SM+SL)/3 -(BH+BM+BL)/3
249
250 // 计算HML
251 gen HML=(SH+BH)/2- (SL+BL)/2
252
253 keep 交易月份 MKT SMB HML
254
255 save 三因子数据.dta, replace
256
257
258 // =====
259 // STEP 9. REG 25 PORTIFOLIO
260 // =====
261
262 use data, clear
263 merge m:1 交易月份 using 三因子数据.dta, nogen
264
265 // 仅市场因子
266 bys ME_group5 BM_group5: asreg 加权组合收益率 MKT , se rmse newey(4)
267 keep ME_group5 BM_group5 _Nobs _rmse _R2 _adjR2 _b_MKT _b_cons _se_MKT _se_cons

```

```

268     duplicates drop ME_group5 BM_group5 _Nobs _rmse _R2 _adjR2 _b_MKT _b_cons _se_MKT _se_cons,
force
269     // 计算t值和p值 标注星号
270     foreach i in cons MKT {
271         gen t_`i'=_b_`i'/_se_`i'
272         gen p_`i'=ttail(_Nobs, abs(t_`i'))*2
273         gen star_`i'="*" if p_`i'<0.1
274         replace star_`i'="*" if p_`i'<0.05
275         replace star_`i'="***" if p_`i'<0.01
276     }
277     export excel using $res_path/25分组回归结果_仅市场因子.xlsx, firstrow(var) replace
278
279     use data, clear
280     merge m:1 交易月份 using 三因子数据.dta, nogen
281     // 市值因子和BM因子
282     bys ME_group5 BM_group5: asreg 加权组合收益率 SMB HML, se rmse newey(4)
283     keep ME_group5 BM_group5 _Nobs _rmse _R2 _adjR2 _b_SMB _b_HML _b_cons _se_SMB _se_HML
_se_cons
284     duplicates drop ME_group5 BM_group5 _Nobs _rmse _R2 _adjR2 _b_SMB _b_HML _b_cons _se_SMB
_se_HML _se_cons, force
285     // 计算t值和p值 标注星号
286     foreach i in cons SMB HML {
287         gen t_`i'=_b_`i'/_se_`i'
288         gen p_`i'=ttail(_Nobs, abs(t_`i'))*2
289         gen star_`i'="*" if p_`i'<0.1
290         replace star_`i'="*" if p_`i'<0.05
291         replace star_`i'="***" if p_`i'<0.01
292     }
293
294     export excel using $res_path/25分组回归结果_市值因子和BM.xlsx, firstrow(var) replace
295
296     use data, clear
297     merge m:1 交易月份 using 三因子数据.dta, nogen
298     // 分组回归计算各组的回归系数, 使用Newey-West t统计量
299     bys ME_group5 BM_group5: asreg 加权组合收益率 MKT SMB HML, se rmse newey(4)
300
301     keep ME_group5 BM_group5 _Nobs _rmse _R2 _adjR2 _b_MKT _b_SMB _b_HML _b_cons _se_MKT _se_SMB
_se_HML _se_cons
302     duplicates drop ME_group5 BM_group5 _Nobs _rmse _R2 _adjR2 _b_MKT _b_SMB _b_HML _b_cons
_se_MKT _se_SMB _se_HML _se_cons, force
303     // 计算t值和p值 标注星号
304     foreach i in cons MKT SMB HML {
305         gen t_`i'=_b_`i'/_se_`i'
306         gen p_`i'=ttail(_Nobs, abs(t_`i'))*2
307         gen star_`i'="*" if p_`i'<0.1
308         replace star_`i'="*" if p_`i'<0.05
309         replace star_`i'="***" if p_`i'<0.01
310     }
311
312     export excel using $res_path/25分组回归结果_三因子.xlsx, firstrow(var) replace
313
314
315
316
317
318
319
320
321

```