set more off

clear

import excel "C:\Users\johnz\Desktop\三因子模型数据和代码\数据整理-三因子.xlsx", sheet("Sheet1") firstrow clear

//样本筛选

gen gap = date - ListDate

drop if gap<=900

drop if year < 2006

egen id=group(Stkcd)

sort id

by id: gen N=\_N

drop if N<=24

drop N

gen Ri\_Rf = Ri - Rf

gen Rm\_Rf = Rm - Rf

//对市值分类： 流通股市值作为ME因子的数据来源，根据Fama和French的做法，按照ME的中位数分为S和B两类。

//对账面市值比分类：将BE/ME从高到低排列后按照前30%为高账面市值比（H）、后30%为低账面市值比（L），其余的为中等账面市值比（M）

sort year id

by year: egen ME\_median=median(ME)

gen group\_ME="S" if ME<ME\_median

replace group\_ME="B" if ME>=ME\_median

gen group\_BE\_ME=""

forvalues i = 2006(1)2016{

xtile temp = BE\_ME if year==`i' , nquantile(10)

replace group\_BE\_ME="L" if temp<=3 & year==`i'

replace group\_BE\_ME="M" if temp>3 & temp<=7 & year==`i'

replace group\_BE\_ME="H" if temp>7 & year==`i'

drop temp

}

gen group\_combine=group\_ME+group\_BE\_ME

//计算SMB

sort date id

by date: egen S\_temp=mean(Ri) if group\_ME=="S"

by date: egen S=mean(S\_temp)

by date: egen B\_temp=mean(Ri) if group\_ME=="B"

by date: egen B=mean(B\_temp)

drop S\_temp B\_temp

gen SMB = S - B

//计算HML

by date: egen H\_temp=mean(Ri) if group\_BE\_ME=="H"

by date: egen H=mean(H\_temp)

by date: egen L\_temp=mean(Ri) if group\_BE\_ME=="L"

by date: egen L=mean(L\_temp)

drop H\_temp L\_temp

gen HML = H - L

//描述性统计

sum Ri Rm Rf Ri\_Rf Rm\_Rf SMB HML , d

corr Rm\_Rf SMB HML

//分成25组

gen ME\_group=.

forvalues i = 2006(1)2016{

xtile temp = ME if year==`i' , nquantile(5)

replace ME\_group=temp if year==`i'

drop temp

}

replace ME\_group=6-ME\_group

gen BM\_group=.

forvalues i = 2006(1)2016{

xtile temp = BE\_ME if year==`i' , nquantile(5)

replace BM\_group=temp if year==`i'

drop temp

}

replace BM\_group=6-BM\_group

gen combine=ME\_group\*10+BM\_group

//分组描述性统计

logout , save(desc1) excel replace: tabstat Ri, by(combine) stat(mean sd max min sk k) c(s) f(%9.4f)

logout , save(desc2) excel replace: tabstat Ri if year<=2009, by(combine) stat(mean sd max min sk k) c(s) f(%9.4f)

logout , save(desc2) excel replace: tabstat Ri if year>2009, by(combine) stat(mean sd max min sk k) c(s) f(%9.4f)

save data.dta, replace

//单因子模型

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf Rm\_Rf, r

save \_b1.dta, replace

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf Rm\_Rf if year<=2009, r

save \_b2.dta, replace

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf Rm\_Rf if year>2009, r

save \_b3.dta, replace

//二因子模型

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf SMB HML, r

save \_2b1.dta, replace

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf SMB HML if year<=2009, r

save \_2b2.dta, replace

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf SMB HML if year>2009, r

save \_2b3.dta, replace

//三因子模型

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf Rm\_Rf SMB HML, r

save \_3b1.dta, replace

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf Rm\_Rf SMB HML if year<=2009, r

save \_3b2.dta, replace

use data, clear

statsby \_b \_se r2=e(r2) se=e(rmse) , by(combine) clear: reg Ri\_Rf Rm\_Rf SMB HML if year>2009, r

save \_3b3.dta, replace