中央财经大学·中国金融发展研究院

名称：投资学第二次作业

组名：GGM Never Die

成员：孟舒晨、郭好格、高雅萌

日期：10/17/2019

Question 1: Underpricing with offer price and underwriters in different time periods.

As for the first question，we will pursue whether this result we get in class holds for other periods. Thus, we collect all IPOs from the "IPOsFrom1970To2006.xls" spreadsheet, and then generate a new variable "period", to distinguish different period, the codes are shown up below.

```
options ls=72 nodate;
libname worklib '/folders/myfolders/sasuser.v94';
* Import IPO data from Excel;
proc import datafile="/folders/myfolders/sasuser.v94/IPOsFrom1970To2006.xls"
      dbms=xls
      out=worklib.ipo REPLACE;
      sheet="IPO";
run;
```

Since 'offer_date' is not entered as a date variable, but simply as a numeric variable YYYYMMDD, we use the method mentioned in class to create a SAS date variable. We also recognize variables that are missing (the ones equal to -9999), setting them to null. And then we generate some variables we need in the following analyze.

```
data d;
	set worklib.ipo;
	if high=-9999 then high=.;
	if low=-9999 then low=.;
	if amend_high=-9999 then amend_high=.;
	if amend_low=-9999 then amend_low=.;
	offer_date_sas=mdy(int(mod(offer_date,10000)/100),mod(offer_date,100),int(offer_date/10
000));
	year_ipo=year(offer_date_sas);
	dollar_width=high-low;
	perc_width=(high-low)/low;
	expected_price=(high+low)/2;
	delta_offer=(offer_price-expected_price)/expected_price;
	ret=100*(close_price-offer_price)/offer_price;
	if offer_price<low then
		subsample="Offer below range";
	else
		if offer_price>high then
			subsample="Offer over range";
		else
			subsample="Offer in range";
	label
		delta_offer="% change in offer price"
	    offer_date_sas="Offering Date";
	format
	    offer_date_sas date9.;
	if 19720101<=offer_date<=19821231 then period="period1";
	if 19830101<=offer_date<=19870930 then period="period2";
	if 19871001<=offer_date<=19911231 then period="period3";
	if 19920101<=offer_date<=19951231 then period="period4";
	if 19960101<=offer_date<=19991231 then period="period5";
	if 20000101<=offer_date<=20061231 then period="period6";
run;
```

And then, We restrict the sample to particular period, take period1 as an example, period1 includes data from January 1972 to December 1982. There is one extra variable we creates: ‘market_power’. It is defined as mentioned in class, and as the underwriter’s power changing through time, so we calculate the underwriters’ market power separately during different period.

```
* Compute market power based on the number of offerings;
* underwritten by each underwriter in the period;
data d1;
	set d;
	where period="period1";
run;
proc sort data=d1;
	by first_uwr;
run;
proc means data=d1 noprint;
	output out=d1_uwr_all
		n(permno)=num_issues_all;
run;
proc means data=d1 noprint;
	output out=d1_uwr
		n(permno)=num_issues;
	by first_uwr;
run;
data d1_uwr;
	if _n_=1 then set d1_uwr_all;
	set d1_uwr;
	market_power=num_issues/num_issues_all;
	label market_power="Market share";
run;
proc sql;
	create table b as select
	d1.*,
	d1_uwr.market_power
	from d1 left join d1_uwr
	on d1.first_uwr=d1_uwr.first_uwr
	order by offer_date_sas;
quit;
```

Finally, we generate some summary statistics. Since we are putting such an emphasis on the price adjustment between the road show and the offering date, we generate a simple table showing the averages of the variables from sub samples based on the price adjustment: whether the final price was below the initial range; within the initial range; or above the initial range.

```
* Generate table with basic statistics;
proc sort data=b;
      by subsample;
run;
proc means data=b noprint;
      output out=b_stats_1
            n(permno)=num_issues
            mean(dollar_width)=dollar_width
            mean(perc_width)=perc_width
            mean(expected_price)=expected_price
            mean(offer_price)=offer_price
            mean(ret)=ret
            median(dollar_width)=mdollar_width
            median(perc_width)=mperc_width
            median(expected_price)=mexpected_price
            median(offer_price)=moffer_price;
run;
```

```
data b_stats_1;
      set b_stats_1;
      subsample="All IPOs";
run;
proc means data=b noprint;
      output out=b_stats_2
            n(permno)=num_issues
            mean(dollar_width)=dollar_width
            mean(perc_width)=perc_width
            mean(expected_price)=expected_price
            mean(offer_price)=offer_price
            mean(ret)=ret
            median(dollar_width)=mdollar_width
            median(perc_width)=mperc_width
            median(expected_price)=mexpected_price
            median(offer_price)=moffer_price;
      by subsample;
run;
```

```
data b_stats;
    set b_stats_1 b_stats_2;
run;
proc print data=b_stats label noobs;
    title 'Basic statistics: IPOs 1972.01-1982.12';
    var subsample num_issues dollar_width perc_width expected_price offer_price ret;
    label
        num_issues="Number of issues"
        dollar_width="Dollar width"
        perc_width="Percent width"
        expected_price="Expected offer"
        offer_price="Actual offer"
        ret="Initial returns";
run;
proc print data=b_stats label noobs;
    var subsample num_issues mdollar_width mperc_width mexpected_price moffer_price;
    label
        num_issues="Number of issues"
        mdollar_width="Median dollar width"
        mperc_width="Median percent width"
        mexpected_price="Median expected offer"
        moffer_price="Median ctual offer";
run;
```

The results of basic statistics are shown as below.

Basic statistics: IPOs 1972.01-1982.12

subsample	Number of issues	Dollar width	Percent width	Expected offer	Actual offer	Initial returns
All IPOs	598	1.50340	0.15632	9.9850	10.279264214	14.8445
Offer be	89	1.72143	0.14414	12.2500	9.6853932584	3.7533
Offer in	192	1.34424	0.16211	8.5784	8.4946614583	12.4526
Offer ov	317	1.96212	0.14868	13.3220	11.526892744	19.4071

Basic statistics: IPOs 1972.01-1982.12

subsample	Number of issues	Median dollar width	Median percent width	Median expected offer	Median ctual offer
All IPOs	598	2	0.15385	10.25	10
Offer be	89	2	0.17647	12.00	9.5
Offer in	192	1	0.12500	8.00	8
Offer ov	317	2	0.17391	13.50	11

Then we run regression, and get its results. We do the same thing for period2 to period6, so we will not repeat the codes here, we just show the coefficient results, and

analyze it changes.

```
* Run regression and analyze its results for different period;
proc reg data=b;
title '1972.01-1982.12 Regression results';
model ret=delta_offer market_power/acov spec dw dwprob;
      output out=reg_data
            residual=r
            predicted=p;
run;
quit;
```

参数估计

变量	标签	自由度	参数估计	标准误差	t 值	Pr > |t|	异方差性相容		
							标准误差	t 值	Pr > |t|
Intercept	Intercept	1	12.92191	2.24781	5.75	<.0001	2.23801	5.77	<.0001
delta_offer	% change in offer price	1	40.00084	9.57497	4.18	<.0001	6.41507	6.24	<.0001
market_power	Market share	1	8.79741	99.54212	0.09	0.9296	69.66120	0.13	0.8996

1972.01-1982.12 Regression results

参数估计

变量	标签	自由度	参数估计	标准误差	t 值	Pr > |t|	异方差性相容		
							标准误差	t 值	Pr > |t|
Intercept	Intercept	1	15.41478	0.77549	19.88	<.0001	0.93760	16.44	<.0001
delta_offer	% change in offer price	1	46.56971	3.40814	13.66	<.0001	3.25295	14.32	<.0001
market_power	Market share	1	-180.04562	39.37231	-4.57	<.0001	36.99606	-4.87	<.0001

1983.01-1987.09 Regression results

参数估计									
							异方差性相容		
变量	标签	自由度	参数估计	标准误差	t 值	Pr > |t|	标准误差	t 值	Pr > |t|
Intercept	Intercept	1	12.23651	0.87459	13.99	<.0001	0.96564	12.67	<.0001
delta_offer	% change in offer price	1	39.87832	3.95674	10.08	<.0001	3.70120	10.77	<.0001
market_power	Market share	1	-11.92458	20.65458	-0.58	0.5639	20.51798	-0.58	0.5614

1987.10-1991.12 Regression results

参数估计									
							异方差性相容		
变量	标签	自由度	参数估计	标准误差	t 值	Pr > |t|	标准误差	t 值	Pr > |t|
Intercept	Intercept	1	14.17298	0.71287	19.88	<.0001	0.75054	18.88	<.0001
delta_offer	% change in offer price	1	56.32255	2.44214	23.06	<.0001	3.15001	17.88	<.0001
market_power	Market share	1	49.63171	26.54961	1.87	0.0618	26.31745	1.89	0.0595

1992.01-1995.12 Regression results

参数估计									
							异方差性相容		
变量	标签	自由度	参数估计	标准误差	t 值	Pr > |t|	标准误差	t 值	Pr > |t|
Intercept	Intercept	1	16.92359	1.85346	9.13	<.0001	1.74352	9.71	<.0001
delta_offer	% change in offer price	1	112.93126	4.61620	24.46	<.0001	15.01167	7.52	<.0001
market_power	Market share	1	444.10835	68.37685	6.50	<.0001	89.57692	4.96	<.0001

1996.01-1999.12 Regression results

参数估计									
							异方差性相容		
变量	标签	自由度	参数估计	标准误差	t 值	Pr > |t|	标准误差	t 值	Pr > |t|
Intercept	Intercept	1	25.18412	1.99336	12.63	<.0001	2.11193	11.92	<.0001
delta_offer	% change in offer price	1	129.46311	4.38060	29.55	<.0001	9.79180	13.22	<.0001
market_power	Market share	1	24.98108	42.38744	0.59	0.5558	45.44059	0.55	0.5826

2000.01-2006.12 Regression results

As we can see, if we just test the period given by the paper, we will think the coefficient on 'delta_offer' is positive and very significant (t-stat=13.66), and underwriters with a large market share are associated lower levels of underpricing. However, things are different after testing all periods, we can easily find out that these results don not hold for other periods.

We cannot point out whether the underpricing level goes up or goes down thru time, as the coefficient changes. But it is clear that offering price affects underpricing for all the periods. The effect of delta_offer is always positive, and we can see t-stats and p values are significant. Maybe we can say the effect is getting bigger thru time, because the coefficient of period6 and period5 are much bigger than period2 or period3.

As for underwriter quality, we cannot say that it affects underpricing for all the periods. The t-stats of period1, period3, period4, period6 are not significant at all.

Question 2: Underpricing and asymmetric information

Question a), import new table from excel;

To import a new table from excel into SAS, we need first to assure the file is an Excel file from the 1997-2003 version and use the command PROC IMPORT. Here, we set up a permanent datafile into worklib, "founding_year".

```
proc import datafile='F:\Investment\Data\age19752019.xls'
    out=worklib.founding_year;
run;
```

Then, we incorporate the variable founding_year into our previous table, which already includes ipos' all kinds of information, such as delta_offer, market_power, ret, etc.

```
*incorporate founding_year into d2;
proc sort data=ipo_f;
    by permno;
proc sort data=d2;
    by permno;
proc sql;
    create table d7 as select
    d2.*,
    ipo_f.founding_year
    from d2 left join ipo_f
    on d2.permno=ipo_f.permno
    order by permno,offer_date;
quit;
```

And then, we generate a new variable “age”, which is the difference between year of ipo-offering and year of its founding.

```
*generate a new variable age;
data d7;
    set d7;
    age=year_ipo-founding_year;
run;
```

Question b), merge the other new variable, size;

In this step, we merge the new variable, size, into our previous table based on their same variable, permno.

```
data ipo_size;
    set worklib.Ipo_total_assets;
    by permno;
    rename data6=size;
run;
proc sort data=d7;
    by permno;
proc sql;
    create table d7 as select
    d7.*,
    ipo_size.size
    from d7 left join ipo_size
    on d7.permno=ipo_size.permno
    order by permno;
quit;
```

question c), generate two additional variables, lage and lsize;

we generate two additional new variables, lage and lsize, by using a function ’log’.

```
data d7;
    set d7;
    lage=log(1+age);
    lsize=log(1+size);
run;
```

Get some summary statistics of age, lage, size, lsize;

We use PROC MEANS to get some summary statistics of four variables, age, lage, size, lsize at one time.

```
proc means data=d7 noprint;
    var age lage size lsize;
    output out=d_stats
    n(age)=n_age
    mean(age)=mean_age
    median(age)=meadian_age
    min(age)=min_age
    max(age)=max_age
    n(lage)=n_lage
    mean(lage)=mean_lage
    median(lage)=meadian_lage
    min(lage)=min_lage
    max(lage)=max_lage
    n(size)=n_size
    mean(size)=mean_size
    median(size)=meadian_size
    min(size)=min_size
    max(size)=max_size
    n(lsize)=n_lsize
    mean(lsize)=mean_lsize
    median(lsize)=meadian_lsize
    min(lsize)=min_lsize
    max(lsize)=max_lsize;
run;
```

The results of summary statistics are shown as follows.

From these results, we can find mean, median, min and max of those logarithms are all smaller than their respective original data. Also, their standard deviations must be smaller than their respective original data, which will make the variables more stationary.

Table 1. summary statistics about age, lage, size, lsize.

variables	number	mean	meadia	min	max
age	1408	13.46804	6	0	87
lage	1408	2.0987734	1.9459101	0	4.4773368
size	1216	122.09127	21.4345	0.064	29190.297
lsize	1216	3.1641169	3.1105999	0.0620354	10.281626

question d) and e), run regresiions and analyze their results;

Model 1: $ret_i = \beta_0 + \beta_1 delta\_offer_i + \beta_2 market\_power_i + \beta_3 age_i + \varepsilon_i$

```
proc reg data=d7;
    title 'regression results7';
    model ret=delta_offer market_power age;
run;
```

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	14.61870	0.74653	19.58	<.0001
delta_offer	%change in offer price	1	44.83963	3.01352	14.88	<.0001
market_power	market share	1	-89.14172	35.31923	-2.52	0.0117
age		1	-0.11988	0.02397	-5.00	<.0001

Figure 1.the regression results of model 1.

According to the table above, we can see that the coefficient of age is less than zero(-0.11988), and p-value is less than 0.001, which means the coefficient is significant. The sign of coefficient means the variable age is negatively related with the variable ret, which proves our assumption is right: the less age means more asymmetry, and more asymmetry means more underpricing.

Model 2: $ret_i = \beta_0 + \beta_1 delta\_offer_i + \beta_2 market\_power_i + \beta_3 lage_i + \varepsilon_i$

```
proc reg data=d7;
    title 'regression results8';
    model ret=delta_offer market_power lage;
run;
```

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	19.02897	1.03996	18.30	<.0001
delta_offer	%change in offer price	1	45.14526	2.98312	15.13	<.0001
market_power	market share	1	-64.16421	35.26703	-1.82	0.0691
lage		1	-3.04463	0.41296	-7.37	<.0001

Figure 2.the regression results of model 2.

From the two above tables we can see that the parameters and p-value of intercept and delta_offer almost stay the same. But the parameter of lage is about 30 times over the parameter of age, and both are significant. We can see that from form of function log(1+x)

narrow down the size of age and keep the monotonicity and sign of x. And the parameter"-3.00463"means when the value of lage increases 1%, the value of ret decreases about 0.03. The log function could change the form of model (from non-linear function to linear function) and reduce the heteroscedasticity of sample data.

Model 3: $ret_i = \beta_0 + \beta_1 delta\_offer_i + \beta_2 market\_power_i + \beta_3 size_i + \varepsilon_i$

```
proc reg data=d7;
    title 'regression results9';
    model ret=delta_offer market_power size;
run;
```

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	15.12525	0.85441	17.70	<.0001
delta_offer	%change in offer price	1	47.59625	3.66611	12.98	<.0001
market_power	market share	1	-170.03767	43.03021	-3.95	<.0001
size	Assets - Total (MM$)	1	-0.00006662	0.00044375	-0.15	0.8807

Figure 3.the regression results of model 3.

According to the table above, we can see that the coefficient of size is less than zero(-0.00006662), and p-value is 0.8807, which means the coefficient is strongly insignificant, which means it can't prove our assumption is right: the less size means more asymmetry, and more asymmetry means more underpricing.

Model 4: $ret_i = \beta_0 + \beta_1 delta\_offer_i + \beta_2 market\_power_i + \beta_3 lsize_i + \varepsilon_i$

```
proc reg data=d7;
    title 'regression results10';
    model ret=delta_offer market_power lsize;
run;
```

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	21.20816	1.36946	15.49	<.0001
delta_offer	%change in offer price	1	48.12263	3.61967	13.29	<.0001
market_power	market share	1	-64.37138	46.39481	-1.39	0.1656
lsize		1	-2.42809	0.43052	-5.64	<.0001

Figure 4.the regression results of model 4.

From the two above tables we can see that the parameters and p-value of intercept and

delta_offer almost stay the same. But the parameter of lage is about 10000 times over the parameter of age, and is significant while the parameter of size is not. The change of p-value may be because the relationship between ret and size is not linear and the size of firms in different industries varies a lot. Log function can solve these problems.

Model 5:

$$ret_i = \beta_0 + \beta_1 delta\_offer_i + \beta_2 market\_power_i + \beta_3 lage_i + \beta_4 lsize_i + \varepsilon_i$$

```
proc reg data=d7;
    title 'regression results11';
    model ret=delta_offer market_power lage lsize;
run;
```

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	21.35766	1.42382	15.00	<.0001
delta_offer	%change in offer price	1	46.79694	3.26975	14.31	<.0001
market_power	market share	1	-33.67299	41.98231	-0.80	0.4227
lage		1	-2.51023	0.50315	-4.99	<.0001
lsize		1	-1.15252	0.42372	-2.72	0.0066

Figure 5. basic model including lage and lsize

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	21.35766	1.42382	15.00	<.0001
delta_offer	%change in offer price	1	46.79694	3.26975	14.31	<.0001
market_power	market share	1	-33.67299	41.98231	-0.80	0.4227
lage		1	-2.51023	0.50315	-4.99	<.0001
lsize		1	-1.15252	0.42372	-2.72	0.0066

Figure 6.the regression results of model 5.

From the above two tables, the most obvious difference is the P-value of market_power increases a lot and turn into insignificant. Because the size of firms and the existence period of firms is related with its market share. In general, the lager the firms are, the lager its market share are, so is the age. So we guess because of multicollinearity, the p-value of market_power changes so much.