

Fama-Macbeth回归

(大约可能是DNS污染或者被墙了，图片显示不出来了)

一、理论概述

Fama-Macbeth回归是1973年Fama和Macbeth为验证CAPM模型而提出的一种因子统计方法，该模型现如今被广泛用于计量经济学的面板数据分析，而在金融领域在用于多因子模型的回归检验，用于估计各类模型中的因子暴露和因子收益（风险溢价）。

Fama-Macbeth与传统的截面回归类似，本质上也是一个两阶段回归，不同的是它用巧妙的方法解决了截面相关性的问题，从而得出更加无偏，相合的估计。

时间序列回归

Fama-Macbeth模型与传统截面回归相同，第一步都是做时间序列回归。在因子分析框架中，时间序列回归是为了获得个股在因子上的暴露。如果模型中的因子是 portfolio returns（即使用投资组合收益率作为因子，例如Fama-French三因子模型中的SMB，HML和市场因子），那么可以通过时间序列回归（time-series regression）来分析 $E[R_i]$ 和 β_i 在截面上的关系。

令 f_t 为因子组合在t期的收益率， R_{it} 为个股i在t期的收益率，用 f_t 对每只股票的 R_{it} 回归，即可得到每只股票的全样本因子暴露 β_i 。

$$R_{it} = \alpha_i + \beta_i f_t + \epsilon_{it}, t = 1, 2, \dots, T, \forall i \quad (1)$$

也可滚动计算某个时间段的因子暴露 β_{it} ，体现个股随市场的变化设置时间段长度为period

$$R_{ik} = \alpha_i + \beta_{it} f_k + \epsilon_{ik}, k = t - \text{period}, 2, \dots, t, \forall i \quad (2)$$

截面回归

传统截面回归的第一步是通过时间序列回归得到个股暴露，这一步与Fama-Macbeth回归相同，而第二步回归体现了传统截面回归和Fama-Macbeth的最大不同。

传统截面回归：

在时序回归中回归式在时间序列上取均值，在 $E[\epsilon] = 0$ 的假设下可以得出：

$$E[R_i] = \alpha_i + \beta_i E[f] \quad (3)$$

上式正是个股的期望收益与因子暴露在截面上的关系，截距 α_i 为个股的错误定价。

那么便可通过截面回归找到因子的期望收益率 $E[f]$ ，方法是最小化个股定价错误 α_i 的平方和。对个股的收益在时序上取均值得到个股期望收益 $E[R_i]$ ，用全样本的个股因子暴露对个股期望收益做无截距回归。

$$E[R_i] = \beta_i \lambda + \alpha_i \quad (4)$$

回归残差 α_i 为个股的错误定价， λ 为因子的期望收益率。

截面回归最大的缺陷在于忽略了截面上的残差相关性，使得OLS给出的标准误存在巨大的低估。

Fama-Macbeth回归

与截面回归相同，Fama-Macbeth回归第一步是通过时间序列回归得到因子暴露值，不同的是，第二步中，Fama-Macbeth在每个t上都做了一次无截距截面回归：

$$R_{it} = \beta_i \lambda_t + \alpha_{it}, i = 1, 2, \dots, N, \forall t \quad (5)$$

上式中的 β_i 为全样本 β ，当然若使用滚动回归数据，也可以在不同截面的回归上使用对应时期的 $\beta_{i,t}$ 。

Fama-Macbeth回归相当于在每个t上做一次独立的截面回归，这T次回归的参数取均值作为回归的估计值：

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^T \hat{\lambda}_t, \hat{\alpha}_i = \frac{1}{T} \sum_{t=1}^T \hat{\alpha}_{it} \quad (6)$$

上述方法的巧妙之处在于它把T期的回归结果当作T个独立的样本。参数的 standard errors 刻画的是样本统计量在不同样本间是如何变化的。在传统的截面回归中，我们只进行一次回归，得到 λ 和 α_i 的一个样本估计。而在Fama-Macbeth截面回归中，把T期样本点独立处理，得到T个 λ 和 α_i 的样本估计。

若使用全样本因子暴露 β_i 进行估计，截面回归和Fama-Macbeth的估计结果相同，当使用滚动窗口进行估计时（Fama and MacBeth (1973)中作者使用了滚动窗口），截面回归和Fama-Macbeth回归会得到完全不同的估计结果。

Fama-Macbeth回归很好的解决了截面相关性的问题，但对于时间序列上的相关性仍然无力。

二、Stata实现

为简单说明Fama-Macbeth两阶段回归的主要步骤，以下用投资组合数据估计一个简单的CAPM模型。数据主要使用了[25 Portfolios Formed on Size and Book-to-Market]中的25个投资组合1926.7-2020.10期间的月度收益率(RP.csv)，和[Fama/French 3 Factors]中的无风险收益、市场超额收益数据(Mkt-RF.csv)。

数据说明：仓库中RP.csv中存储的是25个投资组合1926.7-2020.10期间的月度收益率，每行代表一个月份，每列代表一个投资组合；Mkt-RF.csv存储的是1926.7-2020.10期间的无风险收益、市场超额收益数据，每行代表一个月份，Mkt-RF和RF列代表市场超额收益率和无风险收益。

数据预处理：

变量	含义
port_num	投资组合编号，1~25
t	时期，如1936m7格式
rpe	超额收益，投资组合收益-无风险收益

第一阶段：

pass1 1930.1-1938.11: 25*48次时序回归（1930.1-1934.12->1933.12-1938.11）

估计 $\beta_{it}, i = 1, 2 \dots 25$ ，窗口为五年，每次向后移动一个月

```
bys port_num: asreg rp mktrf if (t>=ym(1930,1) & t<=ym(1938,12)) , wind(t 60)
rmse se newey(4)
```

```
. list in 46/50
```

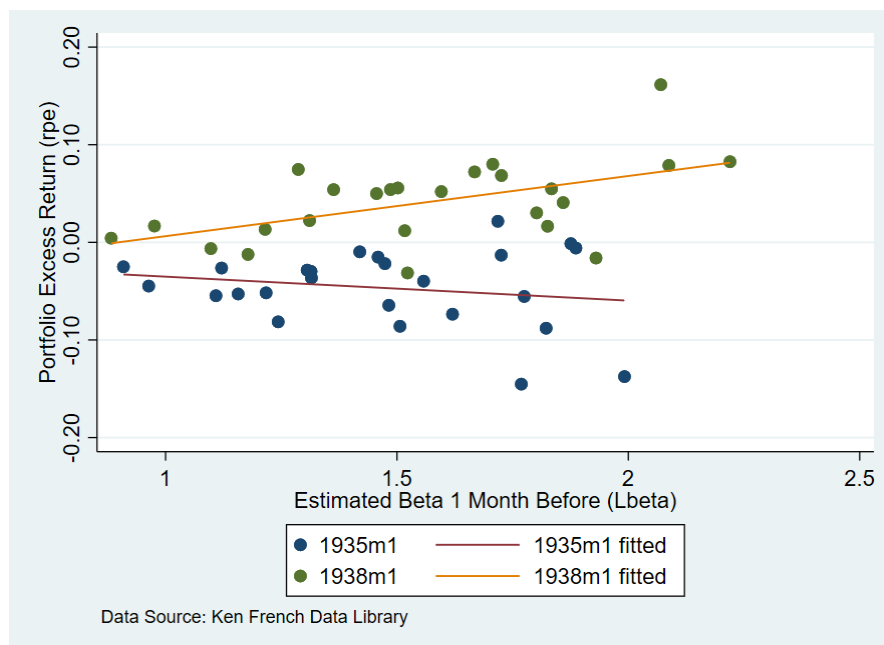
	port_num	t	mktrf	rpe	_rmse	_Nobs	_R2	_adjR2	_b_mktrf	_b_cons	_se_mk~f	_se_cons
46.	1	1938m9	0.01	-0.14	0.16	60	0.42	0.41	1.90	-0.00	0.33	0.01
47.	1	1938m10	0.08	0.12	0.16	60	0.42	0.41	1.91	-0.00	0.33	0.01
48.	1	1938m11	-0.02	-0.04	0.16	60	0.44	0.43	1.97	-0.00	0.38	0.01
49.	1	1938m12	0.04	0.04	0.15	60	0.47	0.46	1.97	0.00	0.38	0.02
50.	2	1934m12	0.00	-0.01	0.22	60	0.48	0.47	1.62	0.03	0.30	0.03

port_num	_b_mktrf	_se_mk~f	_R2	_rmse
1	1.89	0.32	0.48	0.21
2	1.82	0.30	0.52	0.19
3	1.87	0.20	0.72	0.12
4	1.75	0.17	0.70	0.12
5	1.95	0.22	0.65	0.15
6	1.39	0.15	0.66	0.11
7	1.51	0.13	0.77	0.09
8	1.55	0.16	0.79	0.09
9	1.62	0.15	0.80	0.09
10	1.74	0.15	0.75	0.10
11	1.25	0.08	0.81	0.06
12	1.20	0.06	0.90	0.04
13	1.36	0.09	0.92	0.04
14	1.47	0.09	0.86	0.06
15	1.79	0.10	0.89	0.07
16	0.96	0.05	0.92	0.03
17	1.13	0.07	0.94	0.03
18	1.32	0.06	0.93	0.04
19	1.50	0.09	0.91	0.05
20	1.98	0.15	0.86	0.08
21	0.90	0.03	0.97	0.02
22	1.09	0.03	0.98	0.02
23	1.26	0.06	0.95	0.03
24	1.50	0.06	0.92	0.05
25	1.69	0.18	0.78	0.10
Total	1.50	0.13	0.81	0.08

(_b_mktrf就是beta)

为了截面回归更方便，直接将自变量取滞后项(beta滞后一个月)

在做截面回归之前，先看一下rpe和beta估计值的关系



该图画出了 1935m1 和 1938m1 两个时间节点上投资组合超额收益率 rpe 和上一月 估计值 **Lbeta** 的关系，横轴是 Lbeta，纵轴是 rpe。

接下来使用xtfmb进行第二阶段估计，也可以用asreg fmb，还可以用statsby

```
. global regvar "rpe Lbeta"
```

```
.      *xtfmb
```

```
.      xtfmb $regvar
```

```
Fama-MacBeth (1973) Two-Step procedure      Number of obs      =      1200
                                                Num. time periods =      48
                                                F( 1, 47)          =      0.91
                                                Prob > F           =      0.3437
                                                avg. R-squared     =      0.2567
```

rpe	Fama-MacBeth					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Lbeta	.0122536	.0128116	0.96	0.344	-.01352	.0380272
_cons	-.0029518	.0131503	-0.22	0.823	-.0294067	.0235032