The Fama MacBeth '73 type of analysis

The paper by Fama and MacBeth [1973] is important in empirical finance, much because of its methodological innovation.
Look at the original analyisis, designed to test the CAPM.

CAPM estimation

r_{jt} is the return on stock j at time t.
r_{mt} is the return on a stock market index m at time t.
r_{ft} is the risk free interest rate over the same period.
Define the *excess return* as the return in excess of the risk free return.

$$er_{jt} = r_{jt} - r_{ft}$$

$$er_{mt} = r_{mt} - r_{ft}$$

The CAPM specifies

$$E[r_{jt}] = r_{ft} + (r_{mt} - r_{ft})\beta_{jm},$$

where β_{jm} can be treated as a constant.
This can be rewritten as

$$E[r_{jt}] - r_{ft} = (r_{mt} - r_{ft})\beta_{jm}$$

or, in excess return form

$$E[er_{jt}] = E[er_{mt}]\beta_{jm}$$

CAPM estimation ctd

Consider now estimating the crossectional relation

$$(r_{jt} - r_{ft}) = a_t + b_t \beta_{j\hat{m}} + u_{jt} \; j = 1, 2, \ldots, N$$

or in excess return form

$$er_{jt} = a_t + b_t \beta_{j\hat{m}} + u_{jt} \; j = 1, 2, \ldots, N$$

Comparing this to the CAPM prediction

$$er_{jt} = er_{mt} \beta_{jm}$$

we see that the prediction of the CAPM is:

$$E[a_t] = 0$$

$$E[b_t] = (E[r_m] - r_f) > 0$$

CAPM estimation ctd

To test this,

$$E[a_t] = 0$$

$$E[b_t] = (E[r_m] - r_f) > 0$$

average estimated a_t, b_t:
Test whether

$$\frac{1}{T}\sum_{t=1}^{T} a_t \to 0$$

$$\frac{1}{T}\sum_{t=1}^{T} b_t > 0$$

To do these tests we need an estimate of $\beta_{j\hat{m}}$. The "usual" approach is to use time series data to estimate $\beta_{j\hat{m}}$ from the "market model"

$$r_{jt} = \alpha_j + \beta_{jm} r_{mt} + \varepsilon_{jt}$$

on data *before* the crossection.

Exercise

Gather the returns of 10 size based portfolios from Ken French website. Using the data from 1926-2013, do a Fama-MacBeth analysis, i.e.
Estimate

$$er_{it} = a_t + b_t \beta_{it} + e_{it}$$

and test whether $a_t = 0$ and $b_t > 0$.
In doing this use the previous five years to estimate betas using the market model.

Exercise – Solution

The following computer code will do the trick

```
source ("../data/read_pricing_factors.R")
source ("../data/read_size_portfolios.R")
eR <- (FFSize10EW - RF)/100.0
eRm <- RMRF/100.0
n  <- length(eRm)
B <- NULL
for (n2 in 61:n) {
    n1 <- n2-60
    regr <- lm(eR[n1:(n2-1),]~eRm[n1:(n2-1)])
    betai <- regr$coefficients[2,]
    eRi <-  eR[n2,]
    attributes(betai) <- NULL
    attributes(eRi)   <- NULL
    regr <- lm(eRi ~ betai)
    b <- regr$coefficients
    B <- rbind(B,b)
}
head(B)
colMeans(B)
t.test(B[,1])
t.test(B[,2])
```

Let us go over this in come detail.
The data has the form

```
> head(eR)
              Lo10    Dec2    Dec3    Dec4    Dec5    Dec6    Dec7
1926(7)  -0.0141 -0.0183  0.0135  0.0124  0.0083  0.0178  0.0127
1926(8)   0.0478  0.0239  0.0370  0.0328  0.0281  0.0434  0.0126
1926(9)  -0.0048 -0.0111 -0.0232 -0.0097 -0.0065  0.0021 -0.0184
1926(10) -0.0443 -0.0327 -0.0306 -0.0522 -0.0343 -0.0315 -0.0381
1926(11) -0.0150  0.0010  0.0010  0.0289  0.0304  0.0365  0.0338
1926(12) -0.0327  0.0581  0.0421  0.0211  0.0066  0.0095  0.0207
              Dec9    Hi10
1926(7)   0.0306  0.0307
1926(8)   0.0070  0.0341
1926(9)  -0.0124  0.0044
1926(10) -0.0408 -0.0275
1926(11)  0.0337  0.0239
1926(12)  0.0304  0.0273
> head(eRm)
 1926(7)  1926(8)  1926(9) 1926(10) 1926(11) 1926(12)
  0.0295   0.0263   0.0038  -0.0324   0.0254   0.0262
```

Let us look at the first round of the loop,

```
> n2 <- 61
> n1 <- 1
> regr <- lm(eR[n1:(n2-1),]~eRm[n1:(n2-1)])
```

This runs 10 different regressions on 10 size sorted portfolios

	Dependent variable: eRi			
	(1)	(2)	(3)	...
eRm[n1:(n2 - 1)]	1.287^{***}	1.312^{***}	1.192^{***}	···
	(0.141)	(0.095)	(0.080)	···
Constant	0.001	−0.009	−0.007	...
	(0.010)	(0.006)	(0.005)	...
Observations	60	60	60	...
Adjusted R^2	0.583	0.762	0.791	...
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

We now pull the vector of beta coefficients

```
> betai <- regr$coefficients[2,]
```

Lo10	Dec2	Dec3	Dec4	Dec5	Dec6	Dec7	Dec8	Dec9	Hi10
1.287	1.312	1.192	1.049	1.100	1.129	1.114	1.085	1.100	0.964

This is the explanatory variable in a regression in *next-period* returns

```
> eRi <-  eR[n2,]
> attributes(betai) <- NULL
> attributes(eRi)   <- NULL
> regr <- lm(eRi ~ betai)
```

(the attributes part is to allow the coefficients to be used as an explanatory variable.

The results of this regression is

	Dependent variable:
	eRi
betai	0.041
	(0.033)
Constant	-0.116^{**}
	(0.038)
Observations	10
R^2	0.162
Adjusted R^2	0.057
Residual Std. Error	0.011 (df = 8)
F Statistic	1.544 (df = 1; 8)
Note:	*p<0.1; **p<0.05; ***p<0.01

The results of the loop in the Fama-Macbeth analysis is then doing this over and over again, moving a “window” of time over which we estimate the beta coefficients using the market model, and using this beta coefficient to predict the return.

```
> head(B)
  (Intercept)        betai
b -0.11625046  0.041478451
b  0.19939307 -0.192615845
b -0.29409924 -0.009725232
b  0.07420183  0.021346338
b -0.13722297  0.040692233
b  0.22861050 -0.372036413
> colMeans(B)
 (Intercept)        betai
-0.006526818  0.013277653
```

```
> t.test(B[,1])

One Sample t-test

data:  B[, 1]
t = -1.4332, df = 989, p-value = 0.1521
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.015463375  0.002409738
sample estimates:
   mean of x
-0.006526818
```

Regarding the test for the market risk premium we need to specify the alternative differently, since we are explicitly testing whether it is positive.

```
> t.test(B[,2],alternative=c("greater"))

One Sample t-test

data:  B[, 2]
t = 2.7907, df = 989, p-value = 0.002681
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 0.005444296         Inf
sample estimates:
 mean of x
0.01327765
```

Summarizing the results

	constant	beta
average	-0.007	0.013
p.value	0.152	0.002

Econometric issues in Fama MacBeth

The tests across time standard tests, assuming iid.
However, econometric issues in this type of analysis.
Best known: Errors in Variables, since betas are estimated
Solution used by Fama and MacBeth [1973]: Group stocks into portfolios, reducing estimation error in betas.
A recent overview of econometrics of Panel data in finance, including Fama Macbeth: **?**

Replicating Chen Roll and Ross

As a more involved example of using the Fama and Macbeth type of methodology, let us look at the replication of a well known empirical study. In 1986 Chen, Roll and Ross published a paper where they did a Fama MacBeth type of analysis of US stock market crossections, asking whether a number of explanatory variables were risk factors.

We will do a similar analysis updating their data set till today.

Specifically, they use the following explanatory variables

- US Inflation
- US Treasury bill rate (short term)
- US industrial production
- US Long term treasury rates
- Low-Grade bonds (Baa)
- Stock market return
- US Consumption (per capita)
- Oil Prices

They investigate to what degree these alternative "pricing factors" can explain the crossection of asset returns.
The factors they use are (slightly simplified)

- β – Stock market beta
- *dIP* – change in (log) Industrial Production
- *Infl* – Inflation (change in log cpi)
- *dInfl* – first difference of Inflation (not log)
- *Term* – Term Premium (TermSpread)
- *Qual* – Risk premium (Quality Spread)
- *dCons* – change in (log) Consumption
- *dOil* – change in (log) Oil prices

Construct these variables.
We will use as stock market return data for 49 different industry portfolios (ew) provided by Ken French, and use returns starting in 1970.

We will first do the FM analysis for each of the variables. For example, for the CAPM beta we analyze

$$er_{it} = a + b_\beta \widehat{\beta}_{it}^m + e_{it},$$

where the betas are estimated using a MM type regression on data before t, for example five years.

$$er_{i\tau} = \alpha_i + \beta_{it}^m er_{m\tau} + \varepsilon_{i\tau}$$

using observations $\tau = t - 61, \cdots, t - 1$.

For each of the non-beta variables, we will also do an analysis adding the variable to beta and investigate whether it adds explanatory power to the CAPM.
For example, for Industrial Production one will estimate

$$er_{it} = a + b_{\beta}\widehat{\beta}_{it}^{m} + b_{ip}\widehat{\beta}_{it}^{ip} + e_{it},$$

where the betas are estimated using a MM type regression on data before t, for example five years.

Reading the original paper it is not clear which version of the MM regression they do, are we looking at a joint regression

$$er_{i\tau} = \alpha_i + \beta_{it}^{m} er_{m\tau} + \beta_{it}^{ip} dIP_\tau + \varepsilon_{i\tau}$$

using observations $\tau = t - 61, \cdots, t - 1$.
or a "factor by factor" type of analysis?

$$er_{i\tau} = \alpha_i + \beta_{it}^{m} er_{m\tau} + \varepsilon_{i\tau}$$

$$er_{i\tau} = \alpha_i + \beta_{it}^{ip} dIP_\tau + \varepsilon_{i\tau}$$

We will be using this latter version.

Gathering the data

In R, gathering this data is actually relatively simple, as they can be downloaded from the St. Louis Fed data library FRED. Specifically, we will download

- CPIAUCSL (Cpi)
- POP (Population)
- DNDGRA3M086SBEA (Real Consumption)
- INDPRO (Industrial Production)
- OILPRICE
- BAA
- DTB3 (3 month t bills)
- DGS10 (10 year treasuries)

and use these to construct the data series.

R code for doing the data variable construction

```
library(stargazer)
library(zoo)
library(quantmod)

source ("../data/read_pricing_factors.R")
source ("../data/read_industry_portfolios.R")
eR <- (FF49IndusEW - RF)/100.0
head(eR)
eR <- window(eR,start=c(1970,1))

eRm <- RMRF/100.0
names(eRm) <- "eRm"
                                        # cpi urban
getSymbols("CPIAUCSL",src="FRED")
head(CPIAUCSL)
cpi <- CPIAUCSL
head(cpi)
Infl <- diff(log(cpi))
head(Infl)
Infl <- zooreg(coredata(Infl),order.by=as.yearmon(index(Infl)))
names(Infl)<-"Infl"
head(Infl)
                                        #plain difference, not log, inflation m
dInfl <- diff(Infl)
names(dInfl)<-"dInfl"
                                        # total US population
getSymbols("POP",src="FRED")
```

Let us first illustrate the R code for doing the simplest possible Fama Macbeth analysis, a single estimation of the CAPM

```
library(stargazer)
library(zoo)
library(quantmod)

source ("../data/read_pricing_factors.R")
source ("../data/read_industry_portfolios.R")

eR <- (FF49IndusEW - RF)/100.0
length(eR)
head(eR)
eRm <- RMRF/100.0
head(eRm)
eR <- window(eR,start=c(1970,1),end=c(2014,1))

data <- merge(eR,eRm,all=FALSE)

ER <- data[,1:49]
ERM <- data[,50]
head(ER)
head(ERM)

n  <- length(ERM)
B <- NULL
Rsqrs <- NULL
n2 <- 61
```

This results in the following output tables

	constant	beta
average	0.008	0.002
p.value	0.001	0.229

n	mean R2
468	0.094

As we see, the CAPM is not supported in this sample.

The next example shows the analysis of a model where we add the oil price to the market portfolio, and test whether the oil price is a priced risk factor.

Here we find the following results:
First, just oil, without the market portfolio.

	constant	oil
average	0.010	0.003
p.value	0.00002	0.688

n	mean R2
463	0.076

Then the two versions of the analysis adding oil to the market portfolio.

	constant	beta	oil
average	0.006	0.004	0.002
p.value	0.006	0.129	0.739

n	mean R2
463	0.155

Industrial Production

	constant	ind.prod
average	0.009	-0.002
p.value	0.0002	0.008

n	mean R2
468	0.055

	constant	beta	ip
average	0.005	0.005	-0.002
p.value	0.040	0.061	0.006

n	mean R2
468	0.136

Inflation

	constant	dInfl
average	0.011	-0.0003
p.value	0.00005	0.273

n	mean R2
468	0.070

	constant	beta	dInfl
average	0.007	0.004	0.00001
p.value	0.002	0.125	0.969

n	mean R2
468	0.151

Qual Spread

	constant	QualSpread
average	0.009	0.027
p.value	0.001	0.624

n	mean R2
468	0.054

	constant	beta	QualSpread
average	0.008	0.001	0.016
p.value	0.0005	0.416	0.770

n	mean R2
468	0.133

Term Spread

	constant	TermSpread
average	0.007	0.059
p.value	0.005	0.585

n	mean R2
468	0.059

	constant	beta	TermSpread
average	0.006	0.001	0.028
p.value	0.004	0.389	0.801

n	mean R2
468	0.146

Consumption

	constant	dConsum
average	0.009	0.0004
p.value	0.0005	0.524

n	mean R2
468	0.050

	constant	beta	Consum
average	0.007	0.003	-0.0003
p.value	0.003	0.187	0.643

n	mean R2
468	0.133

Chen Roll Ross approximation

To gather the above analysis into a single analysis, we look at The formulation that Chen Roll and Ross focus on,

$$R = a + b_{mp}MP + b_{dei}DEI + b_{ui}UI + b_{upr}UPR + b_{uts}UTS$$

where MP – montly change in industrial production
DEI – change in expected inflation
UI – unexpected inflation
UPR - risk premium (quality spread)
UST - term structure (term spread)
These are the risk premia associated with the various factors.

We will instead of their two inflation measures merely use one variable measuring inflation differences. We therefore estimate using the following data

dIndProd – change in (log) industrial production

dInfl – change in inflation

QualSpread – Quality Spread

TermSpread – Term Spread

Summary stats

Table:

Statistic	N	Mean	St. Dev.	Min	Max
eRm	619	0.005	0.045	−0.232	0.161
dIndProd	619	0.002	0.007	−0.043	0.030
dInfl	619	0.00000	0.003	−0.014	0.018
QualSpread	619	2.000	0.859	0.100	6.280
TermSpread	619	1.529	1.249	−1.910	4.390
dRealCons	619	0.001	0.007	−0.040	0.034
dOilPrice	619	0.006	0.075	−0.396	0.853

Correlations

	eRm	dIndProd	dInfl	QualSpread	TermSpread	dRe
eRm	1					
dIndProd	-0.002	1				
dInfl	-0.062	-0.014	1			
QualSpread	0.073	-0.318	-0.041	1		
TermSpread	0.069	0.022	-0.031	0.464	1	
dRealCons	0.163	0.153	-0.18	-0.057	0.03	
dOilPrice	0.012	0.034	0.298	-0.103	-0.069	-(

First we do the analysis without the market portfolio

	constant	dIndProd	dInfl	QualSpread	TermSpread
average	0.008	-0.002	-0.0001	-0.023	0.080
p.value	0.0004	0.007	0.829	0.776	0.511

n	mean R2
468	0.183

Then we add the market portfolio to this analysis

	constant	betai	dIndProd	dInfl	QualSpread	TermSp
average	0.006	0.002	-0.002	0.0001	-0.003	0.02
p.value	0.010	0.255	0.017	0.761	0.967	0.81

n	mean R2
468	0.235

Finally, add consumption as another potential explanatory variable

	constant	betai	dIndProd	dInfl	QualSpread	Te
average	0.005	0.002	-0.002	0.0003	-0.023	
p.value	0.016	0.259	0.052	0.321	0.765	

n	mean R2
468	0.254

and oil as a final alternative explanatory variable

	constant	betai	dIndProd	dInfl	QualSpread	Te
average	0.006	0.004	-0.002	0.0001	-0.044	
p.value	0.007	0.144	0.002	0.693	0.565	

n	mean R2
463	0.258

Interestingly, Oil seems to “destroy” dIndProd as an explanatory variable. Let us look at just those in isolation

	constant	beta	IP	oil
average	0.004	0.006	-0.002	0.001
p.value	0.069	0.037	0.008	0.813

n	mean R2
463	0.181

Nai fu Chen, Richard Roll, and Stephen Ross. Economic forces and the stock market. *Journal of Business*, 59:383–403, 1986.

Eugene F Fama and J MacBeth. Risk, return and equilibrium, empirical tests. *Journal of Political Economy*, 81:607–636, 1973.