THE FAMA-MACBETH REGRESSION TO TEST CAPM, AND OTHER ASSET PRICING MODELS

[1]Sant Kumar, [2]Dr. Peeyush Bangur, [3]Dr. Nidhi Malhotra

[1]Research Scholar, [2]Associate Professor, [3]Assistant Professor
[1]Faculty of Social Sciences,
[1]Banasthali Vidyapith, Rajasthan, India

Abstract : In financial economics, asset pricing has been a critical topic. For asset pricing, Fama-MacBeth devised a unique regression procedure. Fama-MacBeth regression is a frequently used method in academia and practice for assessing CAPM, Fama-French three component model. This method has its own mathematics; nevertheless, the only way to run this method is through EViews, which is a very expensive application. As a result, the current work is an attempt to devise a strategy to run Fama-MacBeth regression in MS-Excel. There is no Fama-MacBeth regression add-in in MS-Excel.

IndexTerms – **Fama-MacBeth Regression, Cross-section, CAPM, Regression in Excel**

I. INTRODUCTION

The Fama-MacBeth econometric instrument (or type of regression) is used to estimate asset pricing models such as the Capital Asset Pricing Model, the Fama-French three component model, and the Carhart model. In 1973, Nobel winner American economist Eugene F. Fama and D. James MacBeth devised this regression technique. The Fama-MacBeth regression is a two-step regression approach that starts with time series regression and ends with cross-sectional regression. We do, in fact, deal with panel data when performing the Fama-MacBeth regression. In practice, panel data is made up of the returns of various stocks, portfolios, and assets over a specified period of time. There are as many regressions in the first stage of the Fama-MacBeth regression as there are assets. As the number of assets grows, so does the time it takes to conduct regressions. Fama-MacBeth regression is built into the popular statistical programme EViews to save time. While MS-Excel does not provide such an add-in, and thus the Fama-MacBeth regression cannot be run in Excel, this does not indicate that there is nothing in Excel that may be used for this purpose.

II. OBJECTIVES
  (a) Explaining the mathematics of Fama-MacBeth regression.
  (b) Running Fama-MacBeth regression in MS-Excel as a substitute of EViews.
  (c) A comparative study of the results Fama-MacBeth regression on MS-Excel and EViews.

III. DATA AND RESEARCH METHODLOGY
      The work is conceptual in nature, and neither primary nor secondary data is required. However, a hypothetical example has been presented to explain the Fama-MacBeth regression.

IV. THE FAMA-MACBETH REGRESSION: THE THEORETICAL FRAMEWORK
The Fama-MacBeth regression can be understood in two steps
Step 1: Time series regression
Suppose we have *n* portfolios or assets with *m* factors affecting the dependent variable. In equation form we can say that

$$y_{i,t} = \alpha_i + \beta_{1,t}^i x_{1,t} + \beta_{2,t}^i x_{2,t} + \beta_{3,t}^i x_{3,t} + \cdots \ldots \ldots \ldots .. + \beta_{m,t}^i x_{m,t} + \mu_{i,t}$$

where, $y_{i,t}$ shows the returns of the ith asset at time t such that i =1, 2, 3… *n* and t = 1, 2, 3,.., *k* and $\beta_{m,k}^i$ shows the mth coefficient of an explanatory variable $x_{m,t}$ at time t, and $\mu_{i,t}$ is the residual term. Thus, we have the following set of equations:

$$y_{1,t} = \alpha_1 + \beta_{1,t}^1 x_{1,t} + \beta_{2,t}^1 x_{2,t} + \beta_{3,t}^1 x_{3,t} + \cdots \ldots \ldots \ldots .. + \beta_{m,t}^1 x_{m,t} + \mu_{1,t}$$
$$y_{2,t} = \alpha_2 + \beta_{1,t}^2 x_{1,t} + \beta_{2,t}^2 x_{2,t} + \beta_{3,t}^2 x_{3,t} + \cdots \ldots \ldots \ldots .. + \beta_{m,t}^2 x_{m,t} + \mu_{2,t}$$
$$y_{3,t} = \alpha_3 + \beta_{1,t}^3 x_{1,t} + \beta_{2,t}^3 x_{2,t} + \beta_{3,t}^3 x_{3,t} + \cdots \ldots \ldots \ldots .. + \beta_{m,t}^3 x_{m,t} + \mu_{3,t}$$
$$\cdots$$
$$y_{n,t} = \alpha_n + \beta_{1,t}^n x_{1,t} + \beta_{2,t}^n x_{2,t} + \beta_{3,t}^n x_{3,t} + \cdots \ldots \ldots \ldots .. + \beta_{m,t}^n x_{m,t} + \mu_{n,t}$$

We run a simple or a multiple regression as the case may be on each equation to obtain the estimates $of \ \beta_{1,t}^i, \beta_{2,t}^i, \beta_{3,t}^i, \ldots \beta_{m,t}^i$.

Step 2: **Cross-Sectional Regression**

Under this step we run a cross-sectional regression where the dependent variable is the return of a portfolio/security at a time t. In other words, we can say that we run a simple or a multiple regression, as the case may be, on each equation, where, $\beta_{1,t}^i, \beta_{2,t}^i, \beta_{3,t}^i, \dots \beta_{m,t}^i$ are numbers taken from step 1. Since, t = 1, 2, 3… k, therefore, the number of cross-regression is k.

when t = 1: Cross sectional data set 1

Table 1: Iteration 1

Dependent variable	Independent variable				
$y_{1,1}$	$\beta_{1,1}^1$	$\beta_{2,1}^1$	$\beta_{3,1}^1$	…..	$\beta_{m,1}^1$
$y_{2,1}$	$\beta_{1,1}^2$	$\beta_{2,1}^2$	$\beta_{3,1}^2$	…..	$\beta_{m,1}^2$
$y_{3,1}$	$\beta_{1,1}^3$	$\beta_{2,1}^3$	$\beta_{3,2}^3$	…..	$\beta_{m,1}^3$
…..	…..	…..	…..	…..	…..
$y_{n,1}$	$\beta_{1,1}^n$	$\beta_{2,1}^n$	$\beta_{3,1}^n$	…..	$\beta_{m,1}^n$

Using the above data set, run following regression equation to estimate $\gamma_{1,1}, \gamma_{2,1}, \gamma_{3,1}, \dots, \gamma_{m,1}$.

$$y_1 = \gamma_1 + \gamma_{1,1}\beta_{1,1}^1 + \gamma_{2,1}\beta_{2,1}^1 + \gamma_{3,1}\beta_{3,1}^1 + \dots\dots\dots\dots + \gamma_{m,1}\beta_{m,1}^1 + e_1$$

When t = 2

Table 2: Iteration 2

Dependent variable	Independent variable				
$y_{1,2}$	$\beta_{1,2}^1$	$\beta_{2,2}^1$	$\beta_{3,2}^1$	…..	$\beta_{m,2}^1$
$y_{2,2}$	$\beta_{1,2}^2$	$\beta_{2,2}^2$	$\beta_{3,2}^2$	…..	$\beta_{m,2}^2$
$y_{3,2}$	$\beta_{1,2}^3$	$\beta_{2,2}^3$	$\beta_{3,2}^3$	…..	$\beta_{m,2}^3$
…..	…..	…..	…..	…..	…..
$y_{n,2}$	$\beta_{1,2}^n$	$\beta_{2,2}^n$	$\beta_{3,2}^n$	…..	$\beta_{m,2}^n$

Using the above data set, run following regression equation to estimate $\gamma_{1,2}, \gamma_{2,2}, \gamma_{3,2}, \dots, \gamma_{m,2}$.

$$y_2 = \gamma_2 + \gamma_{1,2}\beta_{1,2}^1 + \gamma_{2,2}\beta_{2,2}^1 + \gamma_{3,2}\beta_{3,2}^1 + \dots\dots\dots\dots + \gamma_{m,2}\beta_{m,2}^1 + e_2$$

When t = 3

Table 3: Iteration 3

Dependent variable	Independent variable				
$y_{1,3}$	$\beta_{1,3}^1$	$\beta_{2,3}^1$	$\beta_{3,3}^1$	…..	$\beta_{m,3}^1$
$y_{2,3}$	$\beta_{1,3}^2$	$\beta_{2,3}^2$	$\beta_{3,3}^2$	…..	$\beta_{m,3}^2$
$y_{3,3}$	$\beta_{1,3}^3$	$\beta_{2,3}^3$	$\beta_{3,3}^3$	…..	$\beta_{m,3}^3$
…..	…..	…..	…..	…..	…..
$y_{n,3}$	$\beta_{1,3}^n$	$\beta_{2,3}^n$	$\beta_{3,3}^n$	…..	$\beta_{m,3}^n$

Using the above data set, run following regression equation to estimate $\gamma_{1,3}, \gamma_{2,3}, \gamma_{3,3}, \dots, \gamma_{m,3}$.

$$y_3 = \gamma_3 + \gamma_{1,3}\beta_{1,3}^1 + \gamma_{2,3}\beta_{2,3}^1 + \gamma_{3,3}\beta_{3,3}^1 + \dots\dots\dots\dots + \gamma_{m,3}\beta_{m,3}^1 + e_3$$

When t = k

Table 4: Iteration k

Dependent variable	Independent variable				
$y_{1,k}$	$\beta_{1,k}^1$	$\beta_{2,k}^1$	$\beta_{3,k}^1$	…..	$\beta_{m,k}^1$
$y_{2,k}$	$\beta_{1,k}^2$	$\beta_{2,k}^2$	$\beta_{3,k}^2$	…..	$\beta_{m,k}^2$
$y_{3,k}$	$\beta_{1,k}^3$	$\beta_{2,k}^3$	$\beta_{3,k}^3$	…..	$\beta_{m,k}^3$
…..	…..	…..	…..	…..	…..
$y_{n,k}$	$\beta_{1,k}^n$	$\beta_{2,k}^n$	$\beta_{3,k}^n$	…..	$\beta_{m,k}^n$

Using the above data set, run following regression equation to estimate $\gamma_{1,k}, \gamma_{2,k}, \gamma_{3,k}, \dots, \gamma_{m,k}$.
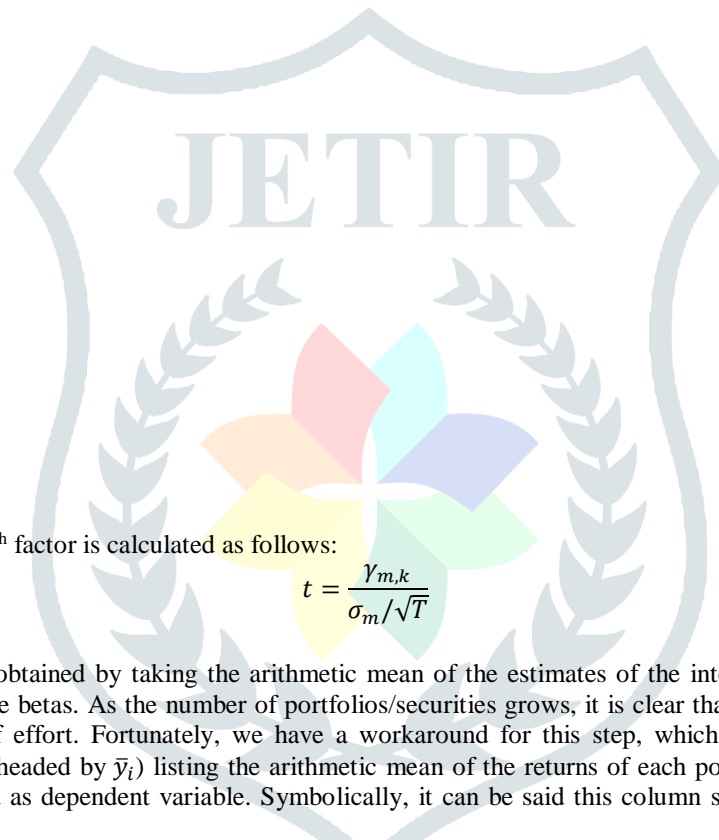
$$y_k = \gamma_k + \gamma_{1,1}\beta_1^1 + \gamma_{2,1}\beta_2^1 + \gamma_{3,1}\beta_3^1 + \dots\dots\dots\dots + \gamma_{m,1}\beta_m^1 + e_k$$

In this way, we get the following estimates

	Constant	Regression coefficients

Cross-sectional regression coefficients	γ_1	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	....	$\gamma_{m,1}$
	γ_2	$\gamma_{1,2}$	$\gamma_{2,2}$	$\gamma_{3,2}$	....	$\gamma_{m,2}$
	γ_3	$\gamma_{1,3}$	$\gamma_{2,3}$	$\gamma_{3,3}$	....	$\gamma_{m,3}$
		....	....	....	....	....
	γ_k	$\gamma_{1,k}$	$\gamma_{2,k}$	$\gamma_{3,k}$	....	$\gamma_{m,k}$
Total of each column	$\sum_{t=1}^{k}\gamma_k$	$\sum_{t=1}^{k}\gamma_{1,k}$	$\sum_{t=1}^{k}\gamma_{2,k}$	$\sum_{t=1}^{k}\gamma_{3,k}$	....	$\sum_{t=1}^{k}\gamma_{m,k}$
Fama-MacBeth regression estimates	$\frac{1}{k}\left(\sum_{t=1}^{k}\gamma_k\right)$	$\frac{1}{k}\left(\sum_{t=1}^{k}\gamma_{1,k}\right)$	$\frac{1}{k}\left(\sum_{t=1}^{k}\gamma_{2,k}\right)$	$\frac{1}{k}\left(\sum_{t=1}^{k}\gamma_{3,k}\right)$	....	$\frac{1}{k}\left(\sum_{t=1}^{k}\gamma_{m,k}\right)$

Table 5 :Time Series Regression

Fama-MacBeth *t*-statistic of mth factor is calculated as follows:

$$t = \frac{\gamma_{m,k}}{\sigma_m/\sqrt{T}}$$

Fama-MacBeth estimates are obtained by taking the arithmetic mean of the estimates of the intercept terms and the arithmetic means of the coefficients of the betas. As the number of portfolios/securities grows, it is clear that step 2 is time demanding and needs a significant amount of effort. Fortunately, we have a workaround for this step, which is outlined below: Under this alternative prepare a column (headed by $\bar{y}_i$) listing the arithmetic mean of the returns of each portfolio over time *t* from 1 to T. This column should be treated as dependent variable. Symbolically, it can be said this column should list the values calculated from the following formula:

$$\bar{y}_i = \left(\sum_{t=1}^{T} y_{i,t}\right) / T$$

where, $y_{i,t}$ = the return of a portfolio *i* at a time *t* from 1 to T.

Now we prepare *m* number of columns in addition to list the values of values of $\beta_1^i, \beta_2^i, \beta_3^i, \dots \beta_m^i$ calculated from first step. These columns should be treated as independent or explanatory variables. Now we run a simple or multiple regression as the case may warrant to estimate the following equation.

$$\bar{y}_i = \alpha + \gamma_1\beta_1^i + \gamma_2\beta_2^i + \gamma_3\beta_3^i + \cdots \dots \dots \dots + \gamma_m\beta_m^i + e_i$$

Fama-MacBeth t-static of m$^{\text{th}}$ factor is calculated as follows:

$$t = \frac{\gamma_m}{\sigma_m/\sqrt{T}}$$

V. THE RESULT AND DISCUSSION OF FAMA-MACBETH REGRESSION

Suppose that we have 3 companies, say, X, Y, and Z. The annual excess returns (i.e., returns in excess of risk-free return) are as follows:

Years	X returns	Y returns	Z returns
2011	12	21	18
2012	14	22	19
2013	10	24	25
2014	9	19	27
2015	17	19.34	10
2016	17.5	15.45	5
2017	20	13.23	3
2018	21	11	11
2019	14	14	11.1
2020	8	11	4

As a first step, we run a time series regression in MS-Excel by clicking the "**data analysis tab**". Since there are three stocks, therefore we run the regression for three times. Consequently, the following regression results are generated. This step is known as **time series regression.** Further, the regression coefficients collected from the regression results are tabulated in Table 1.

Table 6: Time series regression coefficients

Asset	Betas
X	$-0.582182396 = \beta_{1,t}^{1}$
Y	$-0.201370194 = \beta_{1,t}^{2}$
Z	$-0.392637487 = \beta_{1,t}^{3}$

Cross-sectional regression

Cross-section data is a data set where we have various subjects (three companies) at point of time (a particular year in the illustration). In this way, we have ten cross-sectional data sets tabulated in table 2.

Table 7: Cross sections

Cross-sectional data sets	Companies	Returns = Dependent variable	Betas = Independent variable
Cross section 1 (Year 2011)	X	12	-0.582182396
	Y	21	-0.201370194
	Z	18	-0.392637487
Cross section 2 (Year 2012)	X	14	-0.582182396
	Y	22	-0.201370194
	Z	19	-0.392637487
Cross section 3 (Year 2013)	X	10	-0.582182396
	Y	24	-0.201370194
	Z	25	-0.392637487
Cross section 4 (Year 2014)	X	9	-0.582182396
	Y	19	-0.201370194
	Z	27	-0.392637487
Cross section 5 (Year 2015)	X	17	-0.582182396
	Y	19.34	-0.201370194
	Z	10	-0.392637487
Cross section 6 (Year 2016)	X	17.5	-0.582182396
	Y	15.45	-0.201370194
	Z	5	-0.392637487
Cross section 7 (Year 2017)	X	20	-0.582182396
	Y	13.23	-0.201370194
	Z	3	-0.392637487
Cross section 8	X	21	-0.582182396

(Year 2018)	Y	11	-0.201370194
	Z	11	-0.392637487
Cross section 9 (Year 2019)	X	14	-0.582182396
	Y	14	-0.201370194
	Z	11.1	-0.392637487
Cross section 10 (Year 2020)	X	8	-0.582182396
	Y	11	-0.201370194
	Z	4	-0.392637487

Now, a simple OLS (Ordinary Least Squares) regression is run on each cross-sectional data set by clicking again the "**data analysis tab.**" This entire step is known as **cross-sectional regression.**

Table 3: Cross-sectional regression constants and coefficients

	Constants	Regression coefficients
Cross sectional regression results	26.261	23.622
	26.567	21
	34.055	36.7
	28.588	26.157
	17.881	6.2094
	10.575	-5.2923
	5.149	-17.67
	4.0535	-26.22
	13.042	0.023
	10.772	7.9214
Total	**176.9435**	**72.4505**
Fama-MacBeth regression estimates = Average of each column	**17.69435**	**7.245050**
Sample Standard deviation = Standard error	**10.55249**	**20.09478**

Thus, Fama-MacBeth regression equation is
$$Y = 17.69435 + 7.245050X$$
Fama-MacBeth's *t-statistic*

$$t = \frac{\gamma_m}{\sigma_m/\sqrt{k}} = \frac{7.245050 \times \sqrt{10}}{20.09478} = 1.14014$$

and

$$t = \frac{\gamma_m}{\sigma_m/\sqrt{k}} = \frac{17.69435 \times \sqrt{10}}{10.55249} = 5.30249$$

VI. CONCLUSION

. In the realm of financial economics, asset pricing holds significant importance. Fama-MacBeth introduced a distinctive regression approach for asset pricing evaluation. This Fama-MacBeth regression method is widely utilized in both academic research and practical applications, particularly for assessing models like CAPM and the Fama-French three-factor model. Despite its mathematical complexity, the primary challenge lies in its execution, which traditionally requires the use of EViews, a costly software package. Consequently, the present endeavor aims to develop a strategy for conducting Fama-MacBeth regression within the confines of MS-Excel. It's worth noting that MS-Excel lacks a dedicated add-in for Fama-MacBeth regression.

REFERENCES

[1] Fama, Eugene F.; MacBeth, James D. (1973). "Risk, Return, and Equilibrium: Empirical Tests". Journal of Political Economy. 81 (3): 607–636.

[2] Fama Macbeth Reg. (n.d.). Retrieved from https://goldinlocks.github.io/Fama-MacBeth-regression/

[3] Mark, N. C. (2020, October 26). Financial Econometrics Econ 40357 [Review of Financial Econometrics Econ 40357]. Https://Www3.Nd.edu/~Nmark/FinancialEconometrics/Lecture%20Slides/SlidesFall2020_16.Pdf.