

## FAMA-MACBETH TWO-STEP REGRESSION

**Package Name:** fama-macbeth

**Author:** IHS EViews

**Date:** June 30, 2014

**Description:** This add-in performs Fama-MacBeth regression on a set of portfolio or asset returns and factors and returns summary results including the output of a simple cross-sectional average regression. It can be accessed from both dialogs and the command line.

**Add-Ins:** fama-macbeth

## Contents

1	Background	1
2	GUI	3
3	Command line	6
3.1	Parameters	6
3.2	Returns	8
3.3	Examples	8

## 1 Background

Theories of asset pricing frequently use “risk factors” to explain asset returns. These factors can range from macroeconomic (for example, consumer inflation or the unemployment rate) to financial (firm size, etc). The Fama-MacBeth two-step regression is a practical way of testing how these factors describe portfolio or asset returns. The goal is to find the premium from exposure to these factors. In the first step, each portfolio’s return is regressed against one or more factor time series to determine how exposed it is to each one (the “factor exposures”). In the second step, the cross-section of portfolio returns is regressed against the factor exposures, at each time step, to give a time series of risk premia coefficients for each factor. The insight of Fama-MacBeth is to then average these coefficients, once for each factor, to give the premium expected for a unit exposure to each risk factor over time.

In equation form, for  $n$  portfolio or asset returns and  $m$  factors, in the first step the factor exposure  $\beta$ s are obtained by calculating  $n$  regressions, each one on  $m$  factors (each equation in the following represents a regression):

$$\begin{aligned}
 R_{1,t} &= \alpha_1 + \beta_{1,F_1} F_{1,t} + \beta_{1,F_2} F_{2,t} + \dots + \beta_{1,F_m} F_{m,t} + \epsilon_{1,t} \\
 R_{2,t} &= \alpha_2 + \beta_{2,F_1} F_{1,t} + \beta_{2,F_2} F_{2,t} + \dots + \beta_{2,F_m} F_{m,t} + \epsilon_{2,t} \\
 &\vdots \\
 R_{n,t} &= \alpha_n + \beta_{n,F_1} F_{1,t} + \beta_{n,F_2} F_{2,t} + \dots + \beta_{n,F_m} F_{m,t} + \epsilon_{n,t}, \tag{1}
 \end{aligned}$$

where  $R_{i,t}$  is the return of portfolio or asset  $i$  ( $n$  total) at time  $t$ ,  $F_{j,t}$  is the factor  $j$  ( $m$  total) at time  $t$ ,  $\beta_{i,F_m}$  are the factor exposures, or loadings, that describe how returns are exposed to the factors, and  $t$  goes from 1 through  $T$ . Notice that each regression uses the same factors  $F$ , because the purpose is to determine the exposure of each portfolio’s return to a given set of factors.

The second step is to compute  $T$  cross-sectional regressions of the returns on the  $m$  estimates of the  $\beta$ s (call then  $\hat{\beta}$ ) calculated from the first step. Notice that each regression uses the same  $\beta$ s from the first step, because now the goal is the exposure of the  $n$  returns to the  $m$  factor loadings over time (e.g., does a larger factor exposure mean a higher return?):

$$\begin{aligned} R_{i,1} &= \gamma_{1,0} + \gamma_{1,1}\hat{\beta}_{i,F_1} + \gamma_{1,2}\hat{\beta}_{i,F_2} + \dots + \gamma_{1,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,1} \\ R_{i,2} &= \gamma_{2,0} + \gamma_{2,1}\hat{\beta}_{i,F_1} + \gamma_{2,2}\hat{\beta}_{i,F_2} + \dots + \gamma_{2,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,2} \\ &\vdots \\ R_{i,T} &= \gamma_{T,0} + \gamma_{T,1}\hat{\beta}_{i,F_1} + \gamma_{T,2}\hat{\beta}_{i,F_2} + \dots + \gamma_{T,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,T}, \end{aligned} \quad (2)$$

where the returns  $R$  are the same as those in Eqs. 1,  $\gamma$  are regression coefficients that are later used to calculate the risk premium for each factor, and in each regression  $i$  goes from 1 through  $n$ .

In the end there are  $m + 1$  series  $\gamma$  (including the constant in the second step) for every factor, each of length  $T$ . If the  $\varepsilon$  are assumed to be i.i.d, calculate the risk premium  $\gamma_m$  for factor  $F_m$  by averaging the  $m$ th  $\gamma$  over  $T$ , and also get standard deviations and t-stats. For example, t-stats for the  $m$ th risk premium are:

$$\frac{\gamma_m}{\sigma_{\gamma_m}/\sqrt{T}} \quad (3)$$

An interesting comparison is the t-stats calculated using the Fama-MacBeth regression and t-stats from a single cross-sectional regression of returns averaged over time. In other words, the second step with  $T$  regressions is replaced by a single regression of  $n$  portfolio returns, averaged over time, against  $m$  factor exposures with lengths  $n$ :

$$E(R_i) = \gamma_0 + \gamma_1\hat{\beta}_{i,F_1} + \gamma_2\hat{\beta}_{i,F_2} + \dots + \gamma_m\hat{\beta}_{i,F_m} + \varepsilon_i, \quad (4)$$

where  $E(R_i)$  is the average return over time of each portfolio or asset return, the factor exposures  $\hat{\beta}$  are the same as in Eqs. 2, and  $i$  goes from 1 through  $n$ . While the coefficients  $\gamma$  will be the same in both cases (as long as  $\beta$  is constant over time), the standard errors and t-stats won't. This is because in Eq. 4 the factor premia  $\gamma$  aren't estimated with the same time-averaging techniques of the Fama-MacBeth method. In fact, this method is slightly complicated by the fact that the  $\varepsilon$  in Eq. 4 are almost certainly heteroskedastic and autocorrelated. To help correct this problem Newey-West (HAC) standard errors are used in this part of the add-in.

## 2 GUI

This section will illustrate use of the dialogs by working through a simple example included in the Fama-MacBeth regression add-in folder. The monthly data come from 25 equal-weighted portfolios formed on size and book-to-market (pr??) and various market factors (mkt\_fr, rf, smb, hml<sup>1</sup>). In that folder, open fm\_example.wf1. The workfile should look like Fig. 1.

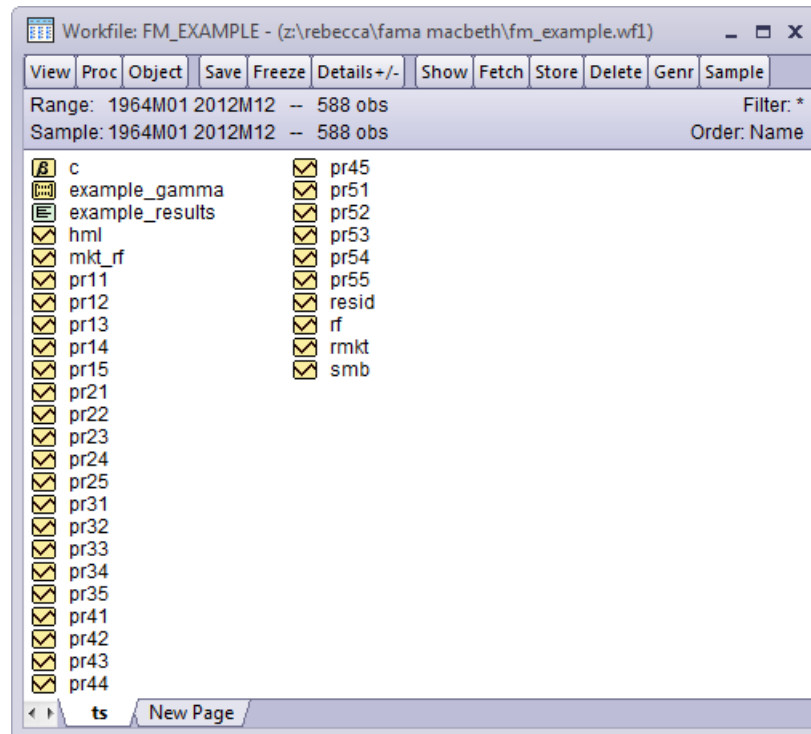


Figure 1: The contents of fm\_example.wf1

If desired set the size of the sample, keeping in mind that NAs in the dataset are not supported. Go to the add-in menu and select “Fama-MacBeth regression” (Fig. 2).

Enter the names of the objects in the workfile that go into the Fama-MacBeth regression. Some examples are in Fig. 3 and Sec. 3.3.

<sup>1</sup>Respectively, excess return, risk-free rate (1m T-bills), average return on small portfolios minus average return on big portfolios, average return on value portfolios minus average return on growth portfolios. All data from Ken French’s data library.

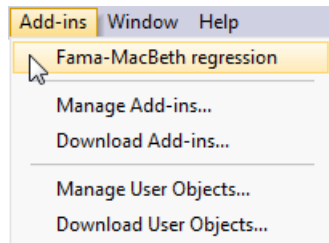


Figure 2: The first step.

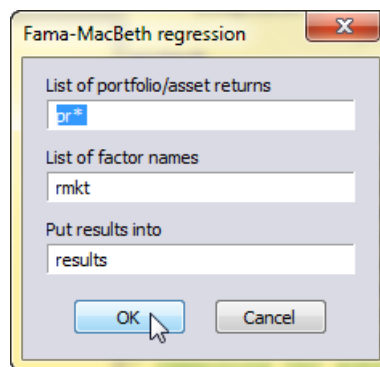


Figure 3: Entering parameters.

The results are collected in a spool object with the default name `results`<sup>2</sup>. If you choose a name for a spool object that already exists, another window will be displayed giving the choice to append to the pre-existing spool object, delete the pre-existing spool object and replace it with a new object of the same name, or create a new spool object. In Fig. 4 the third option has been chosen.

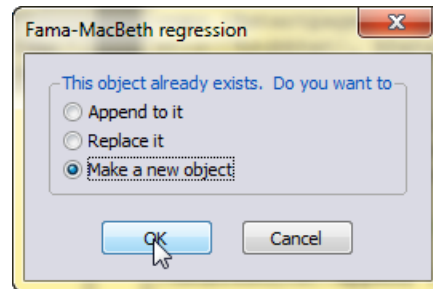


Figure 4: Choices.

When you choose to create a new object it needs to be named. Another window will be displayed (Fig. 5) allowing you to enter the name of your new spool object. In this example, the name of the new spool is `collected`. If you choose a name for an object that already exists, you will be asked again if you want to append to or delete that object or create a new object.

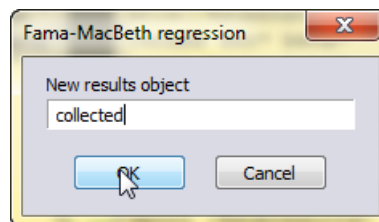


Figure 5: Pick your name.

EViews will quickly calculate the two-stage Fama-MacBeth regression and display a spool object containing the results. See Fig. 6 for an example. The first object in the spool is a table of summary results for each coefficient in the final cross-sectional stage of the regression. The second object is a equation object from a regression of the time series average of every portfolio/asset return against the  $\beta$ s calculated from the first stage of the regression. The comments for each object in the spool give information about the returns and factors used to create it. You may wish to

---

<sup>2</sup>You guessed it.

compare, for example, the t-stats from the two regressions. If you choose to append additional results to a pre-existing spool object, the tree pane will show “Gamma Summary 2”, “CS Average Regression 2”, etc.

### 3 Command line

**fmb**(*prompt*, *action=append*, *spoolname=results??*) **returns factors**

Compute a Fama-MacBeth regression.

As with the GUI the command line gives the options of appending to or replacing a preexisting spool of results, or making a new spool. In order to minimize additional user interaction the functionality for the command line is slightly different than that for the GUI.

The summary output of FMB regression (Gamma Summary 1) and the output of a simple regression of cross-sectional average returns (CS Average Regression 1) are combined in the default spool object **results** (note that there may be multiple objects in the spool, depending on how many regressions are run and the append/replace/new option chosen). The output of the first and second steps of the Fama-MacBeth regressions are collected in the matrix objects **bhat** and **gamma**. There may be more than one **bhat** and **gamma** object each (numbered **bhat01**, **gamma02**, etc) depending on how many regressions are run.

#### 3.1 Parameters

**prompt** :

Include this to show dialogs. Note that typing **fmb** in the command window with no parameters also shows the dialogs.

**action** : append, replace, new (optional)

Append attaches the results of the Fama-MacBeth regression to the end of a preexisting spool object *spoolname* (if this option is chosen and the object doesn't already exist, it is created). Replace deletes the preexisting spool object *spoolname*, replaces it with a new spool object of the same name, and attaches results to it (if this option is chosen and the object doesn't already exist, it is created). New creates a new spool object with the name *spoolname* and

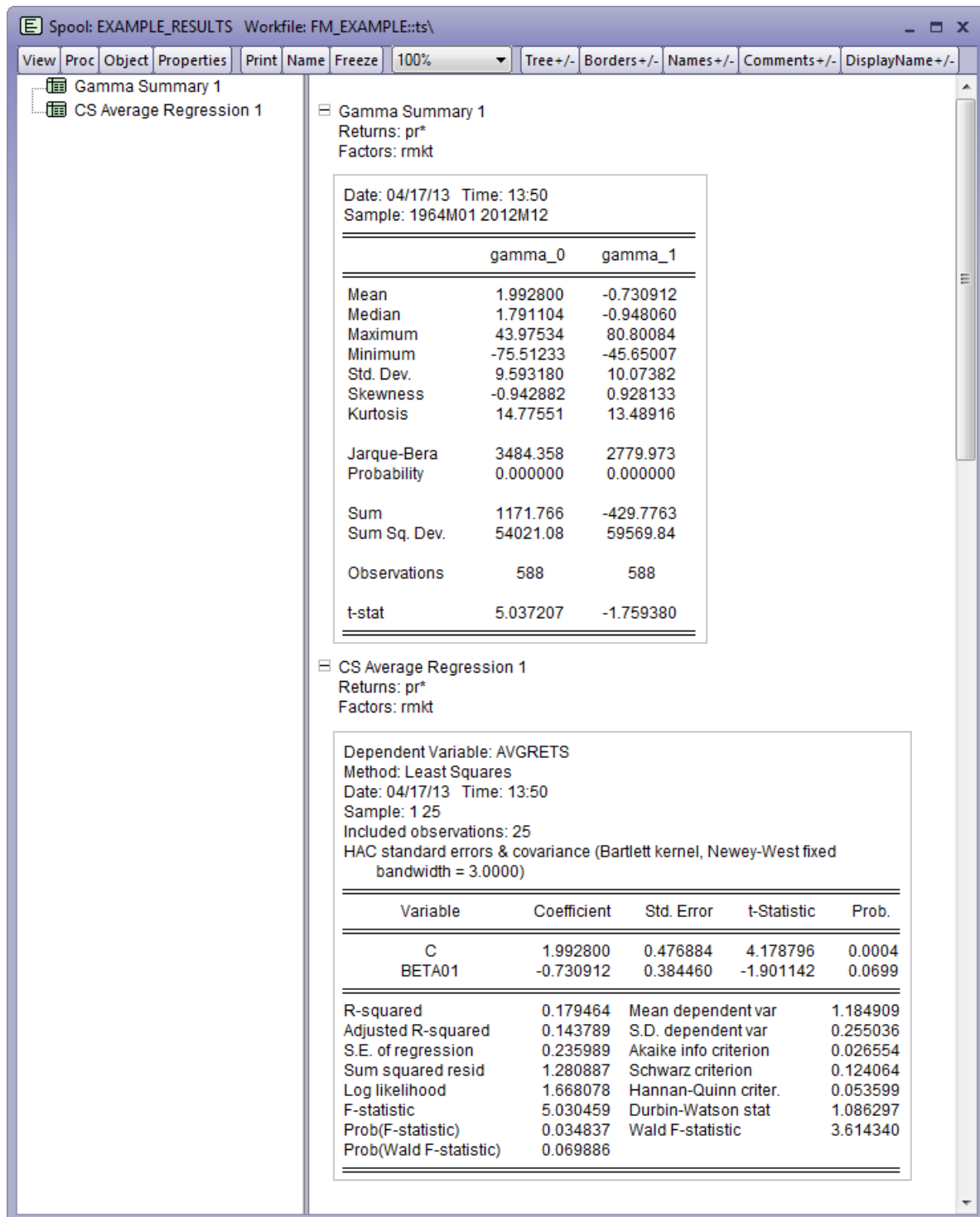


Figure 6: Example results.



attaches results to it (if this option is chosen and the object already exists, the name is incremented as *spoolname01*, *spoolname02*, etc). Default is append.

**spoolname** : name (optional)

Name of the spool object to be appended to, replaced, or created. Default is **result??**, where ?? are wildcard identifiers if multiple Fama-MacBeth regressions are being run (for example, 01, 02, etc).

**returns** : group or series wildcard

The collection of portfolio or asset returns that form the dependent variables of the Fama-MacBeth regression.

**factors** : series, space delimited list of multiple series, group, or series wildcard

The collection of factors that form the independent variables of the Fama-MacBeth regression.

## 3.2 Returns

**bhat** : Matrix containing the  $\beta$  coefficients from the first step of the Fama-MacBeth regression. Each row contains the coefficients from a regression over time of portfolio/asset returns on factors.

**gamma** : Matrix containing the factor risk premia from the second step of the Fama-MacBeth regression. Each row contains the coefficients from a cross-sectional regression of portfolio/asset returns on  $\beta$ s calculated from the first step of the regression.

**results** : Spool object containing the summary output of the the Fama-MacBeth regression (summary statistics by column of the matrix gamma) and, for comparison, the result of a regression of the time series average of each portfolio/asset against the  $\beta$ s. The name can be chosen with the *spoolname* option.

## 3.3 Examples

The first example command below (in the example workfile `fm_example.wf1`, see the series `pr11`, `pr22`, etc, and `rmkt`, `smb`, `hml`) reproduces the results in the `example_beta`, `example_gamma` and `example_results` objects.

```
fmb pr* rmkt
fmb rets smb hml (for previously defined group rets pr*)
```

```
fmb rets facs (for previously defined group rets pr* and group facs  
smb hml)
```

The following example command will create a new spool object named `collected` to hold the results of the Fama-MacBeth regression.

```
fmb(action=new, spoolname=collected) pr* rmkt
```