finregress fmb — Fama–MacBeth regression+

+This command is part of StataNow.

Description

finregress fmb performs Fama and MacBeth (1973) regression, which explores cross-sectional behavior of average returns. This is a two-step regression method. The first step consists of time-series regressions of a set of dependent variables on a set of independent variables (also called factors). These regressions produce coefficients called factor exposures. The second step regresses the cross-section of dependent variables at a point in time against the factor exposures to characterize how the dependent variables vary with factor exposure. The second step is repeated in the cross-section for each time period, and the results are combined over time to produce one number. This combined result is the price of risk, which characterizes how the cross-section of the dependent variables is related to the factor exposures.

Quick start

Find price of risk for assets r1, r2, and r3 using market rate rmkt

```
finregress fmb r1 r2 r3 = rmkt
```

Same as above, but use excess returns net of the risk-free rate (rf) for dependent variables

```
finregress fmb r1 r2 r3 = rmkt, rfrate(rf)
```

Same as above, but adjust the independent variable rmkt for the risk-free rate

```
finregress fmb r1 r2 r3 = rmkt, rfrate(rf) adjust
```

Three-factor Fama–MacBeth regression, adjusting all dependent variables and independent variable rmkt but not the independent variables x1 and x2

```
finregress fmb r1 r2 r3 = rmkt x1 x2, rfrate(rf) adjust(rmkt)
```

Store the period-by-period price of risk estimates in variables with prefix lambda

```
finregress fmb r1 r2 r3 = rmkt x1 x2, rfrate(rf) adjust(rmkt) generate(lambda*)
```

Menu

Statistics > Financial statistics > Fama–MacBeth regression

Syntax

Basic syntax

finregress fmb *depvars* = [*indepvars*] [*if*] [*in*] [, *options*]

Syntax with a risk-free rate

finregress fmb *depvars* = [*indepvars*] [*if*] [*in*] , rfrate(*varname*) [*options*]

options	Description
Model	
rfrate(*varname*)	specify risk-free rate variable and adjust all *depvars*
adjust	adjust all independent variables by risk-free rate variable
adjust(*varlist*)	adjust independent variables in *varlist* by risk-free rate variable
noconstant	omit the constants in the first-stage regressions
small	report small-sample t statistics
dfk	use small-sample degrees-of-freedom adjustment
gls	use generalized least squares in cross-sectional stage
generate(*newvars* | *stub**)	generate variables containing price-of-risk estimates
shanken	apply Shanken correction
Reporting	
level(#)	set confidence level; default is level(95)
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
coeflegend	display legend instead of statistics

You must tsset your data before using finregress fmb; see [TS] **tsset**.

indepvars may contain factor variables; see **[U] 11.4.3 Factor variables**.

depvars, *indepvars*, *varname*, and *varlist* may contain time-series operators; see **[U] 11.4.4 Time-series varlists**.

collect is allowed; see **[U] 11.1.10 Prefix commands**.

coeflegend does not appear in the dialog box.

See **[U] 20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

Model

rfrate(*varname*) specifies the risk-free rate to be used in the first-stage time-series regressions. If rfrate() is not specified, no adjustment is made, and the risk-free rate is implicitly set to 0 in all periods. When rfrate() is specified, it indicates that you wish to transform all dependent variables by subtracting the same risk-free rate, the variable in rfrate().

adjust and adjust(*varlist*) specify which independent variables are to be adjusted by subtracting the risk-free rate in rfrate(). adjust adjusts all independent variables. adjust(*varlist*) adjusts only the independent variables specified in *varlist*.

noconstant omits a constant from the first-stage regressions.

`small` specifies that small-sample t statistics be computed for the tests of coefficients instead of large-sample normal z statistics.

`dfk` specifies that a small-sample degrees-of-freedom adjustment (Campbell, Lo, and MacKinlay 1997) be made when estimating the variance of the price-of-risk coefficient estimate. Specifically, $1/T(T-1)$ is used instead of the large-sample divisor $1/T^2$, where T is the number of time periods.

`gls` uses generalized least squares instead of ordinary least squares when computing the price-of-risk coefficient estimate.

`generate(`*newvars* | *stub*`*)` generates new variables containing the period-by-period estimates of the price-of-risk coefficient. The price-of-risk coefficient reported by the Fama–MacBeth regression is the average of these period-by-period estimates. *newvars* must have the same number of variables as are in *depvars*.

`shanken` applies the Shanken (1992) errors-in-variables correction to the variance of the price-of-risk coefficient estimate.

Reporting

`level(#)`; see [R] **Estimation options**.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(`*style*`)`, `cformat(%`*fmt*`)`, `pformat(%`*fmt*`)`, `sformat(%`*fmt*`)`, and `nolstretch`; see [R] **Estimation options**.

The following option is available with `finregress fmb` but is not shown in the dialog box:

`coeflegend`; see [R] **Estimation options**.

Remarks and examples

Remarks are presented under the following headings:

Introduction
The Fama–MacBeth regression

Introduction

`finregress fmb` fits cross-sectional asset pricing models using the method of Fama and MacBeth (1973). In a cross-sectional asset pricing model, we are interested in learning how exposure of asset returns to factors is related to the average performance of those returns. Intuitively, assets that are "more risky" ought to bring higher average returns (otherwise, no one would hold the riskier assets), and the cross-sectional approach provides a convenient way to measure the strength of this relationship.

To motivate the formal theory, consider the following example. Suppose there are K assets, $i = 1, 2, \ldots, K$, observed over T periods, $t = 1, 2, \ldots, T$. Each asset has a mean return, μ_i and variance σ_i^2. Different assets have different average returns, so μ_i varies across assets i. Why would an investor hold both a high-returning asset and a low-returning asset? The simplest story might be that assets that have high returns also tend to have high variance and that if investors dislike variance, then they will be willing to hold a mix of high-return and low-return assets if the low-return assets also have low variance. This story leads to a testable implication: that there should be an upward-sloping relationship between mean and variance,

$$\mu_i = \gamma + \lambda\sigma_i^2 + u_i$$

By computing the collection of (μ_i, σ_i^2), we can test the theory by regressing the cross-section of mean returns on the cross-section of return variances. A positive value of λ indicates that as variance rises, return also rises; that is, there is higher average return as a reward for higher risk.

The cross-sectional method performs an exercise in a similar spirit. However, instead of using the variance of an asset as its measure of risk, the cross-sectional method measures risk by covariance of a return with a factor, such as market return, or economic conditions. The intuition is that assets that covary strongly with the factor are more risky. In a first step, compute the factor exposures by a time-series regression,

$$r_{it} = \alpha_i + \beta_i x_t + e_{it}$$

where the dependent variable r_{it} is the return of asset i in period t, α_i is the average return in excess of market risk, and the independent variable x_t is a common, observed factor that is proposed to be driving the asset returns. This regression is run, once per asset, to compute a cross-section of factor exposures β_i. There will be K different values for β_i, one for each asset. The value β_i for asset i is called that asset's factor exposure. A β_i of zero indicates that the asset's return does not vary with that factor and so is said to not be exposed to variation in that factor. Factor exposure greater than zero indicates that the asset in question has higher returns in periods when that factor itself is relatively high. Negative factor exposure indicates that the asset in question is a "hedge" in that its returns tend to be high precisely in periods that the returns of the factor in question are low.

In the second stage, we test whether assets that have high factor exposures also have higher average returns:

$$\mu_i = \gamma + \lambda\beta_i + u_i \tag{1}$$

In this regression, the independent variables are themselves regression coefficients from the first stage. The parameter of interest is the coefficient λ, called the price of risk, which captures by how much expected return rises with a one-unit increase in factor exposure.

The Fama–MacBeth regression

The Fama–MacBeth regression is a method for estimating the parameters of the cross-sectional model. Actually running a regression on (1) would produce correct point estimates, but standard formulas would understate the variance of $\widehat{\lambda}$ because of the regression being run on averages and the independent variables β_i being themselves estimated quantities. The Fama–MacBeth regression addresses the first problem by running (1) in each period.

Regression (1) places the $K \times 1$ vector of average returns on the left-hand side. The Fama–MacBeth regression runs T regressions, one per time period, with returns r_{it} on the left-hand side:

$$r_{it} = \gamma_t + \lambda_t\beta_i + u_i$$

In this regression, the left-hand side is the K asset returns at time t, and the right-hand side is the K first-stage estimates of β_i, one for each asset. Thus, the parameter λ_t is the slope coefficient relating the cross-section of returns at time t to the factor exposures. The parameter γ_t is the intercept term, or the zero-beta return, in time period t. Performing this regression for each t, we obtain a time series for $\{\widehat{\lambda}_t\}$. The interpretation is that $\{\lambda_t\}$ is a time series of the price of risk, period by period.

In the second step, the λ_t are averaged over time to obtain a single estimate of the price of risk over the time-period sample.

$$\widehat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\lambda}_t$$

The Fama–MacBeth regression estimates the variance of $\hat{\lambda}$ via the sample variance of the $\hat{\lambda}_t$,

$$\hat{\sigma}^2_{\hat{\lambda}} = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\lambda}_t - \hat{\lambda})^2$$

The estimates of the average zero-beta excess return and its variance, $\hat{\gamma}$ and $\hat{\sigma}^2_{\hat{\gamma}}$, are obtained similarly. The method extends to multiple factors.

For a recent survey of financial regressions that includes a discussion of the Fama–MacBeth regression, see Goyal (2012, 13). For textbook treatments of financial regressions that include the Fama–MacBeth regression, see Cochrane (2005, 246) and Campbell (2018, 64).

▷ Example 1: Single-factor Fama–MacBeth regression

We use simulated data on fictional stock returns and describe the first few variables.

```
. use https://www.stata-press.com/data/r19/finex
(Fictional stock price data)

. describe datestr-wgt

Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
datestr         str11   %11s                  String date
datem           int     %tm                   Monthly date
sp500           double  %10.0g                S&P 500
vol             float   %9.0g                 Volatility index
fedfunds        float   %9.0g                 Federal funds rate
acme            float   %9.0g                 Aciron Medical, Inc.
bat             float   %9.0g                 Boron Advanced Technologies
iron            float   %9.0g                 Industrial Operations Network
dune            float   %9.0g                 Digital Urban Network Enterprise
tyr             float   %9.0g                 Tyndale Resources Group
glo             float   %9.0g                 Green Logistics, Inc.
spa             float   %9.0g                 Space Rocket MFG
wgt             float   %9.0g                 Widget Gadgets
```

This dataset contains fictional stock prices for 25 assets, monthly, from 1955 to 2019.

To prepare the data for analysis, we 1) use `finreturns` to generate monthly returns on the 25 assets; 2) use `finreturns` to generate the return on the market (which is used as a factor); and 3) transform the interest rate so that it is in the same units. The units in this example are log monthly returns.

To create log monthly returns for the 25 assets, we use `finreturns` with the `log()` option. The `multiply(100)` option scales the result so that the units are approximately percentages. A value of 0.50 indicates that the asset increased in value by 0.50 log points or about 0.5%.

```
. quietly finreturns acme-tks, log(lnr_) multiply(100)
```

The result is a set of variables `lnr_acme`, `lnr_bat`, etc, all the way up to `lnr_tks`. We treat the market index, `sp500`, in an identical manner.

```
. quietly finreturns sp500, log(lnr_mkt) multiply(100)
```

The result is a `lnr_mkt` variable that holds the log monthly return on the market. With this preprocessing complete, we can fit a Fama–MacBeth regression.

Recall that the object of interest is the relationship between the cross-section of average returns and the cross-section of time-series coefficients β. As a prelude, we run `finsummarize` to see the average log return for a few assets and their slope coefficients with respect to the market factor.

```
. finsummarize lnr_acme-lnr_tyr, benchmark(lnr_mkt) statistics(beta)

Financial summary statistics

Number of obs = 779
Sample: 1955m2 thru 2019m12

-----------------------------------------------
             |     Mean   Std. dev.       beta
-------------+---------------------------------
        ACME |   0.4437     1.1813      0.1282
         BAT |   0.7534     5.4833      1.5338
        IRON |   0.7979     6.6804      1.8767
        DUNE |   0.8331     6.7105      1.8845
         TYR |   0.6214     3.9969      1.0878
-----------------------------------------------
Note: Variable lnr_mkt used as benchmark
      asset.
```

The `finsummarize` table shows the mean, standard deviation, and slope coefficient (`beta`) for each asset. The goal is to regress mean returns on beta to determine whether, as an asset becomes riskier (higher beta), it also generates increased return, on average, and whether there is a reward for taking on more risk, on average.

`finregress fmb` determines this by directly running that regression. We specify all 25 log return variables as dependent variables and the market factor `lnr_mkt` as the independent variable.

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt

Fama-MacBeth regression

Sample: 1955m2 thru 2019m12                      Number of obs     = 779
Method: OLS                                      Number of depvars =  25

--------------------------------------------------------------------------------
depvar_means | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+------------------------------------------------------------------
beta         |
     lnr_mkt |   .1875542   .1264035     1.48   0.138     -.060192    .4353005
-------------+------------------------------------------------------------------
       _cons |   .4561836   .0193535    23.57   0.000     .4182514    .4941158
--------------------------------------------------------------------------------
Dependent variables: lnr_acme lnr_bat lnr_iron lnr_dune lnr_tyr lnr_glo
                     lnr_spa lnr_wgt lnr_bar lnr_yum lnr_aaa ... lnr_tks
```

The output table provides two numbers. These numbers are the coefficient on a regression of mean log returns (the dependent variable in the second stage) on beta risk (the independent variable in the second stage) and an intercept term. The slope coefficient expresses the expected rise in mean log returns as beta risk rises by one point. The coefficient on `lnr_mkt` is 0.19. This slope indicates that as beta risk rises by one unit, expected average log return rises by 0.19 points.

Although we ran the regression using log returns, the interpretation of the results is often made with respect to simple returns. For small returns, such as daily and monthly returns, we can use the slope coefficient estimated by using log returns to approximate the slope coefficient we would have obtained by using simple returns. So, in our example, we can say that as beta risk rises by one unit, we expect the average simple return to rise by 0.19 percentage points.

More generally, when we can interpret the results as simple returns, we can use the following intuition: A return on a stock with a beta of 1 will be up 1 percentage point every time the market return is also up 1 percentage point and, similarly, will be down 1 percentage point when the market return is down 1 percentage point. A return on a stock with a beta of 2 is more volatile: it will be up 2 percentage points for every time the market return is up 1 percentage point and will be down 2 percentage points whenever the market return is down 1 percentage point. The second stock reacts twice as strongly with respect to the market return as does the first stock. The coefficient estimate of 0.19 indicates that the reward for taking on this increased volatility is a higher average return of 0.19 percentage points per month.

The second number in the coefficient table is the intercept in the regression of average log return on beta risk. Thus, it can be interpreted as the zero-beta return; that is, it is an estimate of the log return on an asset with zero beta risk (an asset that does not react to the market at all). This estimate of 0.46 means that a zero-beta asset is expected to have an average log return of 0.46 or approximately a simple return of 0.46% per month.

◁

▷ Example 2: Multifactor Fama–MacBeth regression

In the presence of independent variables, they are included in the first-stage time-series regressions. For each of the K assets, there will be M independent variables and thus M coefficients, β_{ij}, $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, M$.

The second-stage regressions become a multivariate regression of returns against the estimated first-stage coefficients, performed period by period, and the resulting single number per coefficient is the price of risk with respect to exposure to that variable.

In our dataset, we also have a fictional volatility factor, `vol`, that captures volatility each time period. Adding this factor as an independent variable to the Fama–MacBeth regression produces the following:

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt vol
Fama-MacBeth regression

Sample: 1955m2 thru 2019m12                         Number of obs     = 779
Method: OLS                                         Number of depvars =  25

------------------------------------------------------------------------------
depvar_means | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
beta         |
     lnr_mkt |   .0239579   .1315547     0.18   0.855    -.2338845    .2818003
         vol |   1.384656   .3671927     3.77   0.000     .6649715    2.104341
-------------+----------------------------------------------------------------
       _cons |   .5841546    .046591    12.54   0.000     .4928379    .6754714
------------------------------------------------------------------------------
Dependent variables: lnr_acme lnr_bat lnr_iron lnr_dune lnr_tyr lnr_glo
                     lnr_spa lnr_wgt lnr_bar lnr_yum lnr_aaa ... lnr_tks
```

The coefficient on `lnr_mkt` is essentially zero, indicating that average log returns are not expected to change as market beta risk increases. The coefficient on `vol`, the price of volatility risk, is sharply positive. This indicates that as volatility risk rises, average log returns are expected to rise. Indeed, the estimated coefficient implies that investors require an average increased log return of 1.38 for each unit of volatility risk they bear, or about 1.38% additional simple return per month. That is, assets that are highly exposed to volatility also tend to have high returns. This can be explained as compensation. If investors dislike exposure to volatility, then assets that covary highly with volatility must generate higher returns to entice investors to hold them.

◁

▷ Example 3: Dependent and independent variable adjustment

So far, we have used unadjusted returns. However, it is common to instead use returns in excess of a risk-free rate, r_t^f, as the object of interest. Thus if the dependent variable is the return on asset i, r_{it}, and the independent variables are the market return r_t^m and an additional factor x_t, then it is common to consider

$$(r_{it} - r_t^f) = \alpha + \beta_{i1}(r_t^m - r_t^f) + \beta_{i2}x_t + e_{it}$$

where the dependent variable r_{it} and some independent variables like r_t^m are expressed as excess returns.

Let's use the federal funds interest rate as a risk-free rate. It is reported at an annual rate, so we additionally adjust it to be a monthly rate:

```
. generate double rf = 100 * log(1 + fedfunds/100)/12
```

In `finregress`, the dependent variables are expressed net of a risk-free rate if the `rfrate()` option is specified. Thus, typing

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt, rfrate(rf)
```

would subtract `rf` from each dependent variable prior to fitting the first-stage time-series regressions.

To also adjust independent variables, we could specify them in the `adjust()` option. Thus, typing

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt vol, rfrate(rf) adjust(lnr_mkt)
```

would adjust both the dependent variables and the independent variable `lnr_mkt` by subtracting `rf` prior to estimation. Because `vol` does not appear in `adjust()`, no adjustment is made. Similarly, typing

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt vol, rfrate(rf) adjust(lnr_mkt vol)
```

would adjust all dependent variables and both independent variables `lnr_mkt` and `vol`. This behavior could also be obtained by typing

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt vol, rfrate(rf) adjust
```

Let's run the two-factor model above but use excess log returns for the dependent variables and for the market factor.

```
. finregress fmb lnr_acme-lnr_tks = lnr_mkt vol, rfrate(rf) adjust(lnr_mkt)

Fama-MacBeth regression

Sample: 1955m2 thru 2019m12                     Number of obs     = 779
Method: OLS                                     Number of depvars =  25

------------------------------------------------------------------------------
 depvar_means | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+----------------------------------------------------------------
beta          |
      lnr_mkt |   .1298006    .127215     1.02   0.308    -.1195363    .3791375
          vol |   1.383778   .3676425     3.76   0.000     .6632122    2.104344
--------------+----------------------------------------------------------------
        _cons |   .0907121   .0199399     4.55   0.000     .0516305    .1297936
------------------------------------------------------------------------------
Dependent variables: lnr_acme lnr_bat lnr_iron lnr_dune lnr_tyr lnr_glo
                     lnr_spa lnr_wgt lnr_bar lnr_yum lnr_aaa ... lnr_tks
Notes: Dependent variables adjusted for risk-free rate rf.
       Independent variable lnr_mkt adjusted for risk-free rate rf.
```

The result is that the market factor's price of risk rises to 0.13. The first-stage coefficient on lnr_mkt is now the coefficient on a regression of excess log returns on excess log market return. The associated price of the risk coefficient, shown in the table with a value of 0.13, indicates that as exposure to excess log market return rises by 1, then the excess log return on an asset is expected to rise by 0.13.

◁

Stored results

finregress fmb stores the following in e():

Scalars

e(N)	number of observations
e(k_indepvars)	number of independent variables, not including the constant
e(hastscons)	1 if a constant was included in first stage, 0 otherwise
e(k_dv)	number of dependent variables
e(dfk)	1 if dfk was specified, 0 otherwise
e(shanken)	1 if shanken was specified, 0 otherwise
e(df_m)	model degrees of freedom for second-stage regressions (small only)
e(df_r)	residual degrees of freedom for second-stage regressions (small only)
e(tmin)	minimum time
e(tmax)	maximum time

Macros

e(cmd)	finregress
e(subcmd)	fmb
e(cmdline)	command as typed
e(depvar)	names of dependent variables
e(indepvars_unadj)	names of unadjusted independent variables
e(indepvars_adj)	names of adjusted independent variables
e(rfrate)	name of risk-free rate variable
e(small)	small, if specified
e(model)	type of model
e(title)	title in estimation output
e(tmaxs)	formatted maximum time
e(tmins)	formatted minimum time
e(tvar)	time variable
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved

Matrices

e(b)	coefficient vector
e(V)	variance matrix
e(Sigma_e)	variance matrix of first-stage residuals
e(beta)	first-stage regression coefficients

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

Methods and formulas are presented under the following headings:

Fama–MacBeth regression
Details of options

Fama–MacBeth regression

There are K returns, $i = 1, 2, \dots, K$, and there are T time periods, $t = 1, 2, \dots, T$. The $K \times 1$ vector of returns in period t is $\mathbf{r}_t$. There are M independent variables, and the $M \times 1$ vector of independent variables in period t is $\mathbf{x}_t$.

In the first stage, a typical time period is characterized by the model

$$\mathbf{r}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_t + \mathbf{e}_t$$

where $\mathbf{a}$ is a $K \times 1$ vector of intercepts, $\mathbf{B}$ is a $K \times M$ matrix of coefficients, and $\mathbf{e}_t$ is a $K \times 1$ vector of time-series errors. The matrix $\widehat{\mathbf{B}}$ is an estimate of $\mathbf{B}$ obtained by ordinary least squares, and it contains the point estimates of the time-series regression.

In the second stage, the cross-sectional regression is run,

$$\mathbf{r}_t = \mathbf{1}_K\gamma_t + \widehat{\mathbf{B}}\boldsymbol{\lambda}_t + \mathbf{u}_t$$

where $\mathbf{1}_K$ is a vector of 1s, γ_t is an intercept term for period t, $\boldsymbol{\lambda}_t$ is an $M \times 1$ vector of coefficients for period t, and $\mathbf{u}_t$ are cross-sectional errors in period t. The cross-sectional regression is run T times, one for each t, with the result being a time series of coefficients $\{\widehat{\boldsymbol{\lambda}}_t\}_{t=1}^T$ and a time series of constants $\{\widehat{\gamma}_t\}_{t=1}^T$. The final Fama–MacBeth estimates are

$$\widehat{\boldsymbol{\lambda}} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\lambda}}_t \qquad \hat{\gamma} = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_t$$

with their corresponding variance estimates

$$\widehat{\mathbf{V}}_{\widehat{\boldsymbol{\lambda}}} = \frac{1}{T^2}\sum_{t=1}^{T}(\widehat{\boldsymbol{\lambda}}_t - \widehat{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\lambda}}_t - \widehat{\boldsymbol{\lambda}})' \qquad \hat{\sigma}^2_{\hat{\gamma}} = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\gamma}_t - \hat{\gamma})^2$$

The variance calculation is based on Cochrane (2005, 246).

Details of options

When `noconstant` is specified, the first-stage (time-series) regressions are run without an intercept. An intercept is still included in the second-stage (cross-sectional) regressions.

When `gls` is specified, the cross-sectional regressions are fit with generalized least squares instead of ordinary least squares.

When `dfk` is specified, a small-sample degrees-of-freedom adjustment is applied to the second-stage regressions.

When `shanken` is specified, a correction is made in the second-stage regressions to account for the fact that the coefficient matrix $\mathbf{B}$ was estimated in the first stage. The Shanken correction, described in Shanken (1992) and (Goyal 2012, 14) is

$$\widetilde{\mathbf{V}}_{\widehat{\boldsymbol{\lambda}}} = \frac{1}{T}\left\{(1+c)(T\widehat{\mathbf{V}}_{\widehat{\boldsymbol{\lambda}}} - \boldsymbol{\Sigma}_{\mathbf{x}}) + \boldsymbol{\Sigma}_{\mathbf{x}}\right\}$$

where $\boldsymbol{\Sigma}_{\mathbf{x}}$ is the covariance matrix of the independent variables and $c = \widehat{\boldsymbol{\lambda}}'\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\widehat{\boldsymbol{\lambda}}$.

References

Campbell, J. Y. 2018. *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton, NJ: Princeton University Press.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j.ctt7skm5.

Cochrane, J. H. 2005. *Asset Pricing*. Rev. ed. Princeton, NJ: Princeton University Press.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636. https://doi.org/10.1086/260061.

Goyal, A. 2012. Empirical cross-sectional asset pricing: A survey. *Financial Markets and Portfolio Management* 26: 3–38. https://doi.org/10.1007/s11408-011-0177-7.

Shanken, J. 1992. On the estimation of beta-pricing models. *Review of Financial Studies* 5: 1–33. https://doi.org/10.1093/rfs/5.1.1.

Also see

[FIN] **finregress fmb postestimation** — Postestimation tools for finregress fmb+

[FIN] **finregress capm** — Capital asset pricing model (CAPM)+

[FIN] **finreturns** — Generate financial returns+

[U] 20 Estimation and postestimation commands