

Instructions for *bigreg*

The command *bigreg* implements a pooled OLS regression with firm-specific slope coefficients in Stata. It can handle any number of firm-specific slope coefficients (e.g., $5,000 \times 3 = 15,000$ coefficients in a sample of 5,000 firms with 3 slope coefficients per firm¹). The methodology is from Appendix A of Byzalov and Basu (2023), “The misuse of regression-based *x-Scores* as dependent variables”.

To get started, download *bigreg.ado* and save it either in the Stata directory for user-installed ado files (e.g., C:\ado\plus per standard Stata directory definitions for Windows) or in the code directory for your current project. Before first use, install the third-party command *rangestat* by typing *ssc install rangestat*.

If you saved *bigreg.ado* in the code directory for your project, you should *cd* to this directory (e.g., *cd c:\MyCurrentProject\code*) in your code before calling *bigreg* so that Stata can find it.

The general syntax is

bigreg Y pooledX [if], firmspecX(varlist) firmID(varname) [cluster(varname)]

where

Y = the dependent variable for the regression;

pooledX = the list of *X* variables with pooled slope coefficients;

if = Stata sample selection condition (optional);

firmspecX(varlist) = defines the list of *X* variables with firm-specific slope coefficients, e.g., *firmspecX(D R D_R)* for firm-specific slopes on *D*, *R*, and $D \times R$ in the Basu model;²

firmID(varname) = defines the firm index for firm-specific coefficients, e.g., *firmID(gvkey)* in Compustat data;³

cluster(varname) = defines the cluster variable for the standard errors (optional); if it is omitted, then the standard errors are clustered by the *firmID* variable.⁴

¹ You should carefully weigh whether you have sufficient data for each firm to estimate 3 firm-specific slope coefficients (plus a firm-specific intercept).

² The firm-specific intercept (i.e., a firm fixed effect) is included automatically. Therefore, the variable list for *firmspecX(...)* should not include the constant.

³ “Firm” is shorthand for any relevant group of observations with group-specific slope coefficients (e.g., household, industry, state).

⁴ The cluster variable may differ from the *firmID* variable, e.g., *firmID(gvkey) cluster(SIC3)*. Only one-way clustering is supported (but you can modify *bigreg.ado* by replacing *reg* in stage 2 with two-way-clustered regression).

The command does not accept factor notation (#, ##, etc.) in *pooledX* or *firmspecX*. Therefore, you should prepare all relevant interactions (e.g., Basu model interactions of *D*, *R*, and $D \times R$ with the control variables) in your code before calling *bigreg*. The command also does not accept time-series operators (L., D., etc.).

Example: Basu model with firm-specific slope coefficients on *D*, *R*, and $D \times R$

```
bigreg Earn X D_X R_X D_R_X `PooledVars', firmspecX(D R D_R) firmID(gvkey)
```

where *X* is the test variable, *D* is the negative return dummy, *R* is stock return, *D_R* is $D \times R$, *D_X...D_R_X* are the interactions of *X* with *D*, *R*, and $D \times R$ (prepared earlier in the code), *PooledVars* is the list of control variables and their interactions with *D*, *R*, and $D \times R$,⁵ and the options *firmspecX(D R D_R) firmID(gvkey)* define firm-specific slope coefficients on *D*, *R*, and $D \times R$ (plus a firm-specific intercept) for each distinct *gvkey*. The standard errors are clustered by the *firmID* variable *gvkey*. To cluster the standard errors by three-digit SIC industry *SIC3*, add the option *cluster(SIC3)*.

⁵ For example, for a Basu model with control variables *Size*, *BM*, and *Lev*, define *pooledVars* as *local pooledVars = "Size BM Lev D_Size R_Size D_R_Size D_BM R_BM D_R_BM D_Lev R_Lev D_R_Lev"*, where interactions like *D_R_Size* have been prepared earlier in the code. Do not include stand-alone *D*, *R*, and *D_R* in *pooledVars*, because they are handled separately in *firmspecX(D R D_R)*.