# Stata Camp

## Day 2: Linear Regression Analysis

**Agenda:**

Learn how to create do-files

Learn how to draw line plots between variables

Learn how to run benchmark regression

Learn how to run group regression

Learn how to perform robustness checks for panel data

1. Create do-files and record your commands

Way 1: click the button “New do-file editor”

Way 2: select the command from the history and send the command to the do-file editor.

1. Line plot

Line plot allows you to review the relationship between two variables based on each individual

----------------------------------------------------

xtline invest mvalue

------------------------------------------------

1. Benchmark regression

Benchmark regression is the main regression that you show in your main conclusion.

First use “xttest0” and “hausman fe\_result re\_result” to choose suitable models

----------------------------------------------------

reg invest mvalue kstock, r

est store ols\_result

xtreg invest mvalue kstock,fe

est store fe\_result

xtreg invest mvalue kstock,re

est store re\_result

esttab ols\_result fe\_result re\_result using regression\_result.rtf, replace b(%12.3f) se(%12.3f) nogap compress s(N r2 r2\_a)star(\* 0.1 \*\* 0.05 \*\*\* 0.01)

----------------------------------------------------

Then output your regression table and explain the results in your paper.

1. Group regression

Group regression allows you to run regression in different groups.

1. Use the command “merge” to include regional data
2. Transfer the string variables into numeric variables (egen g\_region = group( region ) )
3. label define:

----------------------------------------------------

label define lg\_region 1 "Asia" 2 "Europe" 3 "North America"

----------------------------------------------------

1. combine the label with variables:

----------------------------------------------------

label values g\_region lg\_region

----------------------------------------------------

1. drop original variable “region”

----------------------------------------------------

drop region

----------------------------------------------------

1. start group regression

----------------------------------------------------

bysort g\_region: xtreg invest mvalue kstock i.year,fe

----------------------------------------------------

1. if you want to output the results, use the command outreg2 to help:

----------------------------------------------------

bysort g\_region:outreg2 using grouping\_result.doc, replace pvalue bdec(3) pdec(3) keep(invest mvalue kstock) addtext(Time fixed effect, YES, Individual fixed effect, YES): xtreg invest mvalue kstock i.year,fe

----------------------------------------------------

1. Robustness check

To check whether your results are stable, you should: replace the dependant variables and core explanatory variables respectively; and solve the endogenous problems.

What is endogenous problems?

It originates from the sample selection bias, omitted variables and simultaneous causality.

How to solve endogenous problems?

We usually use instrument variables (IV) to solve endogenous problems. Other methods such as DID(Difference in differences), RDD(regression discontinuity design), PSM(propensity score matching) are also applicable to solve endogenous problems.

2SLS and GMM

1. 2SLS: principal and command

Example: graduation rate and education quality

----------------------------------------------------

webuse mroz,clear

global convars "exper expersq"

ivregress 2sls lwage $convars (educ = motheduc fatheduc),first

----------------------------------------------------

1. System GMM(Generalized Method of Moment)

----------------------------------------------------

webuse grunfeld,clear

sort company year

xtabond2 invest l.invest kstock , gmm( invest , lag(1 4) collapse) gmm( kstock , lag(1 4) collapse) iv( mvalue ) robust two

----------------------------------------------------

AR（1）<0.1

AR（2）>0.1

Hansen p-value: 0.1-0.25