

Basic Regressions and Panel Data in Stata

Ben Shepherd
Principal, Developing Trade Consultants

Basic regressions

- } Stata's regress command runs a simple OLS regression
 - } *Regress depvar indepvar1 indepvar2, options*

- } Always use the option robust to ensure that the covariance estimator can handle heteroskedasticity of unknown form.

- } Usually apply the cluster option and specify an appropriate level of clustering to account for correlation within groups
 - } Rule of thumb: apply cluster to the most aggregated level of variables in the model.
 - } Example: In a model with data by city, state, and country, cluster by country.

Fixed effects regressions

- } A variety of commands are available for estimating fixed effects regressions. We will focus on two only: regress with dummy variables, and xtreg.

- } The fixed effects model is sometimes called the Least Squares Dummy Variable (LSDV) model because the fixed effects can just be entered as dummies in a standard regression.
 - } *Regress depvar indepvar1 indepvar2 ... dum1 dum2, options*
 - } Specify dum* to include all dummy variables with the same stem.
 - } Stata automatically excludes one dummy if a constant is retained in the model

Fixed effects regressions

- } To create dummy variables based on categories of another variable, use the `tabulate` command with the `gen()` option.

- } For example:
 - } *Quietly tabulate country, gen(ctry_dum_)*
 - } Will produce `ctry_dum_1`, `ctry_dum_2`, etc. automatically.
 - } Then *regress depvar indepvar1 indepvar2 ... ctry_dum_*, robust cluster()*

Fixed effects regressions

- } Stata's `xtreg` command is purpose built for panel data regressions.
- } Use the `fe` option to specify fixed effects
- } Make sure to set the panel dimension before using the `xtreg` command, using `xtset`
- } For example:
 - } `xtset countries` sets up the panel dimension as countries.
 - } `xtreg depvar indepvar1 indepvar2 ..., fe` runs a regression with fixed effects by firm.
- } Hint: `xtset` cannot work with string variables, so use (e.g.) `egen countries = group(country)` to convert string categories to numbers.

Fixed effects regressions

- } As with regress, always specify the robust option with xtreg.
- } Xtreg will automatically correct for clustering at the level of the panel variable (firms in the previous example).
- } With the same clustering specification, results should be identical between regress with dummy variables and xtreg, fe.
- } Note that xtreg can only include fixed effects in one dimension. For additional dimensions, enter the dummies manually.

Random effects regressions

- } By specifying the `re` option, `xtreg` can also estimate random effects models.
 - } `xtreg depvar indepvar1 indepvar2 ..., robust re`

- } As for the fixed effects model, you need to specify `xtset` first
 - } `xtset countries`
 - } `xtreg depvar indepvar1 indepvar2 ..., robust re`
 - } Runs a regression with random effects by country

- } Fixed and random effects can be included in the same model by including dummy variables.

- } An alternative that can also be used for multiple dimensions of random effects is `xtmixed` (outside our scope).

Testing for fixed versus random effects

- } The fixed effects model always gives consistent estimates whether the data generating process is fixed or random effects, but random effects is more efficient in the latter case.
- } The random effects model only gives consistent estimates if the data generating process is random effects.
- } Intuitively, if random effects estimates are very close to fixed effects estimates, then using random effects is probably an appropriate simplification. If the estimates are very different, then fixed effects should be used.

Testing for fixed versus random effects

- } The Hausman test exploits this intuition.

- } To run it:
 - } Xtreg ... , fe
 - } Estimates store fixed
 - } Xtreg ..., re
 - } Estimates store random
 - } Hausman fixed random

- } If the test statistic is large, reject the null hypothesis that random effects is an appropriate simplification.

- } Caution: the Hausman test has poor properties empirically and often fails to give sensible results.

- } Note: You can only run the Hausman test on fixed and random effects estimates that do not include the robust option.