

Practical Guides To Panel Data Analysis

Hun Myoung Park
05/16/2010

1. Which effect? Group vs. Time? Fixed vs. Random?

Panel data models examine cross-sectional (group) and/or time-series (time) effects. These effects may be fixed and/or random. Fixed effects assume that individual group/time have different intercept in the regression equation, while random effects hypothesize individual group/time have different disturbance. When the type of effects (group versus time) and property of effects (fixed versus random) combined, there are several specific models: fixed group effect model (one-way), fixed time effect model (one-way), fixed group and time effect model (two-way), random group effect model (one-way), random time effect (one-way), and random group and time effect model (two-way).

As a result, **you have to select one specific model**. Presenting all possible models sounds quite absurd; strictly speaking, if one model is right, the other models are wrong. It must be ridiculous to present wrong models as well unless you want to go “data-fishing.”

2. Determine Your Model

In order to determine your model, conduct necessary tests: F test for the fixed effect model and Breusch-Pagan Lagrange Multiplier (LM) test. Hausman test is needed if you find both fixed and random effects. See the next section for the test. Once you finish F and LM tests, determine your model as shown in the following table.

Fixed effect (F test or Wald test)	Random effect (Breusch-Pagan LM test)	Your model
H_0 is not rejected (No fixed effect)	H_0 is not rejected (No random effect)	Data are poolable. Pooled OLS
H_0 is rejected (fixed effect)	H_0 is not rejected (No random effect)	Fixed effect model
H_0 is not rejected (No fixed effect)	H_0 is rejected (random effect)	Random effect model
H_0 is rejected (fixed effect)	H_0 is rejected (random effect)	(1) Fixed and random effect model ¹ or (2) choose one of the two depending on the result of Hausman test (recommended direction).

If you have a small number of observations in your data set, avoid two-way models especially when a group or time has many different values (e.g., 100 countries in group and 400 quarters in time). Why? The null hypothesis of fixed or random effect model is almost useless; almost all cases end up rejecting H_0 .

1 In theory, it is not allowed to impose both fixed and random effects for a group or time variable. For instance, it is not possible to have fixed group (say, urban versus rural) and random group effect simultaneously. Imposing two effects of a group or time is contradictory conceptually. However, you may fit a model with a fixed group effect and random time effect (or vice versa) using both least squares dummy variable (LSDV) model and a random effect model. This model is possible but least recommended largely due to the loss of parsimony and degrees of freedom.

3. F Test (Wald Test) for Fixed Effects

F test reported in the output of the fixed effect model is for overall goodness-of-fit, not for the test of the fixed effect. In order to test fixed effect, run `.test` command in Stata after fitting the least squares dummy variable model with `.regress` (not `.xtreg`). For example, if you have 3 dummies variables for 4 groups, you may run the command as, “`.test d1 d2 d3`”

4. Breusch-Pagan Lagrange Multiplier (LM) Test for Random Effect

`.xtreg` conducts Wald test to report the goodness-of-fit but this is not for random effect. Use `.xttest0` command to test if there is any random effect.

5. Hausman Test for Comparing Fixed and Random Effects

Hausman test compares the fixed and random effect models. If both fixed and random effects turn out significant, Hausman test will give you a good idea when choosing one between the two. The null hypothesis is that the (fixed or random) effect is not correlated with other regressors (independent variables). “A random effect model is better than the fixed effect model” and “A random effect model is consistent” are NOT CORRECT null hypotheses for the Hausman test.

If the null hypothesis is rejected, a random effect model will be suffering from the violation of the Gauss-Markov theorem and end up with biased and inconsistent estimates; by contrast, a fixed effect model still remains unbiased and consistent. To make it simple, **if the null hypothesis is rejected, use the fixed effect model; otherwise, go for the random effect model.**

6. Reporting R^2 (Stata Specific)

Some software packages do not report correct statistics. Stata (up to version 11) `.xtreg` returns wrong R^2 in the fixed effect model because the command fits the within model (running OLS on transformed data with the intercept suppressed) without adjusting R^2 . And other two R^2 (between and overall R^2) are almost meaningless. `.xtreg` also report incorrect (a bit different) R^2 in random effect models. Therefore, DO NOT REPORT any R^2 from the output of the fixed effect model that Stata produces unless Stata revises the command to report the correct R^2 . In order to get correct R^2 for the fixed effect model, use `.areg` command in Stata, SAS TSCSREG or PANEL produce, or LIMDEP.

7. Presenting the Results

You need to report parameter estimates and their standard errors. Of course, present proper goodness-of-fit measures (e.g., F, R^2) as well. In a random effect model, you also need to report estimated theta and variance component for error. In F-test and LM test, do not forget to report degrees of freedom.

If you fit a fixed effect model, you may report parameter estimate for constant term of each group or time. This is not always required, but it depends on your research questions. For instance, if you want to show which group (or year) has the largest initial status and this issue is essential in your research, individual intercepts in addition to parameter estimates of regressors should be reported. In this case, do not simply present estimated constant term and dummy parameters.

8. Interpretation

Evaluate the overall goodness-of-fit measures (e.g., F , R^2) of the model. If the model does not fit the data well. You must stop here and report no significant effects of independent variables.

Interpret parameter estimates (coefficients) in a substantive way. Do not simply report signs and magnitude of coefficients. Do not simply say “significant,” “negatively (or positively) related”, or “insignificant relationship between...” One recommended example is “For one unit increase in a IV, DV is expected to increase by OO units, holding all other variables constant.”

9. Comments on Functional Forms

Many students use log linear regression model to take advantage of getting growth rate. I guess this biased practice is pervasive in economics. In fact, it is a nonlinear model, however. You SHOULD report the original equation. Many studies have unstructured equations that may be estimated by simple OLS without log transformation. In this case, estimation and interpretation become much easier; growth rate needs to be computed after estimation though. Think about this functional form as an alternative. However, I am not saying that log linear regression is wrong though.

- $\ln_Y = b_0 + b_1 \ln_{X1} + b_2 \ln_{X2} + b_3 \ln_{X3}$ (log linear equation)
- $Y = (b_0) \cdot (X1^{b1}) \cdot (X2^{b2}) \cdot (X3^{b3})$ (original nonlinear equation)
- $Y = b_0 + b_1 X1 + b_2 X2 + b_3 X3$ (alternative equation)

Some References for Panel Data Analysis

Baltagi, Badi H. 2001. *Econometric Analysis of Panel Data*. Wiley, John & Sons.

Greene, William H. 2003. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

<http://www.indiana.edu/~statmath/stat/all/panel/panel.pdf>