Panel data analysis: random effects and tests
Day 3.

David Pupovac

Previous classes

- In the previous classes we have discussed:

- Standard errors corrections
  - Clustered standard errors
  - Panel corrected standard errors

- Fixed effects model:
  - Intercept (dummy) estimators – we spoke about three variants
$$y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t}$$
  - Within estimator
$$y_{i,t} - \bar{y}_i = (x_{i,t} - \bar{x}_i)\beta + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

- Between effect model:
$$\bar{y}_i = \alpha + \bar{x}_i\beta + \nu_i + \bar{\varepsilon}_i$$

Random effects model

- Today we will focus on random effects.
- What is the place of random effects with respect to other panel data estimators?
  - (Pooled) OLS is a weighted average of between and within variation
  - Fixed effect ignores between variation and focuses predominantly on within variation (i.e. the within estimator uses complete demeaning)
  - Between estimator ignores all within variation
  - **The random effects estimator is somewhere in between**.

Random effects model

- The basic form of random effects can be defined as:

$$y_{i,t} = \alpha + x_{i,t}\beta + u_{i,t}$$

where: $u_{i,t} = \nu_i + \varepsilon_{i,t}$

$$\boldsymbol{y_{i,t} = \alpha + x_{i,t}\beta + \nu_i + \varepsilon_{i,t}}$$

- The first major difference in this expression, in comparison to fixed effects, is that instead unique intercept term for each unit α_i, in random effects we have a common intercept term α.

- The second major difference (which is due to the introduction of common intercept term α) is the error term ν_i. **Namely, the random effect approach is to estimate a "composite" error term that consists of a "random intercept" component for each unit, ν_i, and a random error term, $\varepsilon_{i,t}$.**

Random effect error terms ν_i and $\varepsilon_{i,t}$

- The ν_i is an error term representing the extent to which the intercept of the *i*th cross-sectional unit differs from the overall intercept. Thus, ν_i the unit-specific error term; it differs between units, but for any particular unit, its value is constant.

- Therefore unlike with fixed effects, we assume that the unobserved individual effects are <u>stochastic</u>. The unit means are assumed to have some (typically normal) distribution rather than being some fixed constant.

- On the other hand, $\varepsilon_{i,t}$ is the "usual" error term unique to each observation, with the usual properties

- Thus, to some extent, in random effect we are partitioning the variance on common and unique part.

Difference between random effect and fixed effects

- The core difference between fixed and random effect models lies in the role of dummy variables. If dummies are considered as a part of the intercept, this is a fixed effect model. So we are assuming <u>different intercepts but same slopes.</u>

- In a random effect model, the dummies act as an error term. A random effect model estimates variance components for groups (or time units) and error, assuming <u>the same intercept and slopes</u>. The difference among groups (or time units) lies in their variance of the error term, not in their intercepts.

Two way random effect: error terms ν_i, $\varepsilon_{i,t}$ and τ_t

- Sometimes a third error term is included, representing the extent to which the t^{th} time unit's intercept differs from the overall intercept. This idea is analogues to the two way fixed effects model

- Therefore, the residual, $u_{i,t}$, is often assumed to consist of three components:

$$u_{i,t} = \tau_t + \nu_i + \varepsilon_{i,t}$$

- The variance of $y_{i,t}$ conditional on $x_{i,t}$ is:

$$\sigma_y^2 = \sigma_\tau^2 + \sigma_\varepsilon^2 + \sigma_\nu^2$$

- For now we will assume that $\sigma_\tau^2 = 0$

Random Effects- correlation of x and v_i

- **As v_i is a part of error, it should not be correlated to any regressor x; otherwise, a core OLS assumption is violated – exogeneity.** Thus, the unobserved individual heterogeneity, is assumed to be uncorrelated with (or independent of) all other included and excluded variables.

- The crucial distinction between fixed and random effects is whether the unobserved individual effect embodies elements that are correlated with the regressors in the model, not whether these effects are stochastic or not.

- The fixed effects model allows the unobserved individual effects to be correlated with the included variables. If the individual effects are strictly uncorrelated with the regressors, then it might be appropriate to model the individual specific constant terms as randomly distributed across cross-sectional units. The cost is the possibility of inconsistent estimates, should the assumption turn out to be inappropriate.

Properties of random effects

- None of the desirable properties of the random effects model rely on T going to infinity. Indeed, in typical application T is likely to be quite small.

- The estimator of between effects (which is used in estimation of random effects) is equal to an average of N units, each based on the T observations for unit i. That is, its variance is of order $1/T$ or smaller. Because T is small, this variance may be relatively large. But, each term provides some information about the parameter. The average over the N cross-sectional units has a variance of order $1/(NT)$, which will go to zero if N increases, even if we regard T as fixed.

- The conclusion to draw is that nothing in this treatment relies on T growing large. Although it can be shown that some consistency results will follow from T increasing.

Properties of random effect

- In principal, fixed effects and random effects are isomorphic. Namely, as T grows large, random effects converges to fixed effects.

- Random effects is just a linear combination of fixed effects estimator and the between units estimator. The weight on the between unit estimators is

$$\lambda = \frac{\sigma_{\varepsilon}^{2}}{\sigma_{\varepsilon}^{2} + T\sigma_{\nu}^{2}}$$

- As T grows large λ goes to zero, between estimator becomes less relevant, and random effect estimator converges to its fixed effect counterpart (we will get back to this in a moment).

Random effect - estimation

- A random effects estimator is essentially a generalized least squares (GLS) estimator. It produces a matrix-weighted average of the between and within results

- A random effect model is estimated by generalized least squares (GLS) when the variance structure among groups is known (which, from practitioner's point of view, is never).

- The <u>feasible</u> generalized least squares (FGLS) method is used to estimate the variance structure when the variance structure among groups is not known.

There are various estimation methods for FGLS including the maximum likelihood method and simulation (Baltagi and Cheng 1994).

Random effects - estimation

- GLS is based on transform equation, where we subtract from each equation element a corresponding term multiplied by some constant, θ. The constant θ measures the weight given to the between-group variation (see formula below).

$$y_{i,t} - \theta\bar{y}_i = (1-\theta)\alpha + (x_{i,t} - \theta\bar{x}_i)\beta + \{(1-\theta)\nu_i + (\varepsilon_{i,t} - \theta\bar{\varepsilon}_i)\}$$

- When $\theta = 0$, random effects is equal to pooled OLS
- When $\theta = 1$, random effects is equal to fixed effects

$$\theta = 1 - \sqrt{\lambda}$$
$$= 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\nu^2}}$$

where $\hat{\sigma}_\epsilon^2$ is variance estimate of idiosyncratic component $\varepsilon_{i,t}$, and $\hat{\sigma}_\nu^2$ is the variance of unit component ν_i

- **So, if $T = 0$ we get OLS, and if $T \longrightarrow \infty$ we get fixed effects**

Random effects - estimation

- In essence random effects is a two step estimation.

- In the **first step** we need estimate of θ – i.e. we need to estimate within variation σ_ε^2 and between variation σ_ν^2. There are many ways to do this Typically we do this by fixed and between effects estimators. Noting that:

$$y_{i,t} - \bar{y}_i = (x_{i,t} - \bar{x}_i)\beta_w + (\nu_{i,t} - \bar{\nu}_i)$$
$$\bar{y}_i = \alpha + \bar{x}_i\beta_b + \bar{\varepsilon}_i$$

we can use the within- and between-group residuals to estimate σ_ε^2 and σ_ν^2

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(y_{i,t} - \bar{y}_i\right) - \hat{\beta}_w\left(x_{i,t} - \bar{x}_i\right)\right]^2}{N(T-1) - K}$$

$$\hat{\sigma}_\nu^2 = \frac{\sum_{i=1}^{N}(\bar{y}_i - \beta_b\bar{x}_i)^2}{N - K - 1} - \hat{\sigma}_\varepsilon^2$$

- In the **second step** the data are GLS transformed using the estimates. We then use OLS to estimate our model on the transformed data

Random effects estimation - MLE and GLS

- We could also fit this random-effects model with the maximum likelihood estimator.

- The MLE and GLS yield essentially the same results, except when total sample size is small (200 or less) and the data are unbalanced.

- If the assumption of normality does not hold, MLE will be more wrong than GLS.

- Typically you should check both MLE and GLS estimates of the model – they are likely to produce different results – if they are significantly different go with MLE

Two way random effects model

- As underlined earlier, similarly to the distinction between one-way and two-way fixed effects (based on cross-sectional and temporal units), we can also employ either one-way or two-way random effects models in estimation.

(We will return to the additional estimators soon)

Random effect vs. fixed effects

- There are trade offs in choosing between the fixed effects and random effects models. In general:

  - If the data exhaust the population (say observations on all firms, all EU countries), then the fixed effects approach, which produces results conditional on the units in the data set, is reasonable.

  - If the data are a drawing of observations from a large population and we wish to draw inferences regarding other members of that population, the fixed effects model is no longer reasonable; in this context, use of the random effects model has the advantage that it saves a lot of degrees of freedom.

- **However, keep in mind that while fixed effects will tend to attenuate the coefficient on x, random effects will tend to inflate it**

Random effect vs. fixed effects

- A drawback of fixed-effects models is that they require the estimation of a coefficient of the unit dummy variable - <u>this can substantially reduce the model's power and increase the standard errors of the estimates</u>. Thus, the estimate of β in the fixed-effects model may produce estimates that are overly sensitive to the random error:

  – If there are few observations per unit, or if x varies little within each unit relative to the amount of variation in y. Estimates of the within-unit effects of x may then diverge considerably from the true effect due to chance alone

- Furthermore, the within-unit effects may diverge from the true effects in the same direction, leading the estimate of β to be quite different from the true β.

- **Random-effects models enable estimation of $\boldsymbol{\beta}$ with lower sample-to-sample variability by partially pooling information across units.** By estimating the variance parameter σ_v^2, the random-effects estimator forms a compromise between the fixed-effects and pooled models. Groups with outlying unit effects will have their respective α_j shrunk back toward the mean, μ_α

Random effect vs. fixed effects

- If data are multilevel, or effectively so, you probably are good with random effects.

- Furthermore, if you have a small T, there is little alternative to random effects.

- While random effects makes sense with panel data, (i.e., where $N >> T$); with longer T, however, fixed effects is more appropriate.

- Using fixed effects you are obscuring some of the between information. This is not the case with random effects. However, it raises the question of how to interpret the effect (i.e. is the effect cross-sectional or temporal in nature?)

- In addition, random effects model also makes sense if your model includes explanatory variables that don't change over time for a unit (i.e. they are constants or near constants). In this case fixed effects will wipe out the effect of the variables

Random effect vs. fixed effects

- If predictor variables vary greatly across units (individuals, countries) but have little variation over time for each unit, than, unlike random effects, fixed effects estimates will be imprecise and have large standard errors.

- Fixed effects methods help to control for omitted variable bias by having units serve as their own controls – thus, in comparison to random effects, fixed effects models are less vulnerable to omitted variable bias

- The random effects model has a major drawback: it assumes that the random error associated with each cross-sectional unit is uncorrelated with the regressors, something that is not likely to be the case.

- **This will result in bias in the coefficient estimates from the random effects model.**

(This may explain why the slope estimates from the fixed and random effects models are often so different)

Tests

- In addition to strictly theoretical concerns, the selection between pooled OLS, fixed effects and random effects should be based on tests. Similarly, we can test for various model assumptions discussed in previous classes.

pooled OLS vs. fixed effects

- Fixed effects vs. pooled OLS are tested by the (incremental) F test. This hypothesis test is based on loss of goodness-of-fit. If the null hypothesis is rejected, you may conclude that the fixed group effect model is better than the pooled OLS model.

Tests

Breusch and Pagan (Lagrange multiplier) test for random effects

- Breusch and Pagan (1980) have devised a Lagrange multiplier test for the random effects model based on the OLS residuals for appropriateness of pooled OLS vis-à-vis random effects estimator.

- **Lagrange multiplier test tests whether or not the random effects have a variance.**

- Under the null hypothesis, the limiting distribution of Lagrange multiplier test is chi-squared with one degree of freedom.

Tests

Hausman test

- The Hausman specification test compares the fixed versus random effects under the null hypothesis that the individual effects are uncorrelated with the other regressors in the model (Hausman 1978).

- If correlated (H_0 is rejected), a random effect model produces biased estimators, violating one of the Gauss-Markov assumptions; so a fixed effect model is preferred.

Tests

Hausman test

- Hausman test is based on comparison of fixed and random effects models. The null hypothesis is that if the two estimation methods are both good, they should yield coefficients that are "similar". So, a significant Hausman test for correlation between the error and the regressors indicates that there is a major difference between fixed effects and random effects estimators

- The alternative hypothesis is that the fixed effects estimation is good and the random effects estimation is not. Thus, significant test indicates appropriateness of fixed effects

Tests

Mundlak test

- Mundlak's approach to TSCS/CSTS data is based on observation that if α_i is correlated with X_{it}in period t, then it will also be correlated with X_{it} in period s. Mundlak's assumption was that the individual-specific effect is equally correlated with all time-period. A benefit of the approach is that it provides a convenient approach to the test for correlation between the error and the regressors
- A linear panel-data model is given by

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

- The key to the Mundlak approach is to determine if α_i and x_{it} are correlated. we think of the mean of our outcome conditional on our covariates.

$$\alpha_i = \bar{x}_i \theta + \nu_i$$

- In the expression above, $\bar{x}_i$ is the panel-level mean of x_{it}, and ν_i is a time-invariant unobservable that is uncorrelated to the regressors. If $\theta = 0$, we know α_i and the covariates are uncorrelated.

$$y_{it} = \beta x_{it} + \bar{x}_i \theta + \nu_i + \varepsilon_{it}$$

Tests

Mundlak test

- Application of Mundlak model is simple:

  - Compute the panel-level average of your time-varying covariates.
  - Use a random-effects estimator to regress your covariates and the panel-level means against your outcome.
  - Test if the panel-level means are jointly zero

If you do not want to make the strong assumptions made by Mundlak you can use Chamberlain’s approach

Tests:

heteroscedactity and cross-sectional dependence

- We can test for cross-sectional independence of a fixed effect regression model or GLS. There are two options:

  1. Breusch-Pagan statistic, based on Greene (2000, p. 601), assumes that the deviation from independent errors in the context of pooled cross-section time-series data is likely to be due to contemporaneous correlations across cross-sectional units.

  2. Pesaran (2004) test for cross-section dependence in panel time-series data. The routine allows for multiple variable series to be tested at the same time

- Considering violation of the assumption of homoscedasticity, we can calculate a modified Wald statistic for groupwise heteroskedasticity in the residuals of a fixed effect regression model.

Swamy–Arora estimator

- Let’s discuss additional estimators:

- Considering random effects you can use the small-sample Swamy–Arora estimator instead of the default estimator

- Stata has two implementations of the Swamy–Arora method for estimating the variance components. They produce the same results in balanced panels, however, the two methods differ in unbalanced panels

Hausman–Taylor estimator for error-components models

- If you are confident you should model data using random effects, but your model includes time invariant variables or, more importantly, there is correlation between some of the covariates with the unobserved individual-level random effect, you can use Hausman and Tayloar (1981) or Amemiya and MaCurdy (1986) instrumental variables estimator.

- Let's assume we have a model:

$$y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + Z_{1i}\delta_1 + Z_{2i}\delta_2 + \mu_i + \varepsilon_{it}$$

  - X_{1it} is a vector of observations on exogenous, time-varying variables
  - X_{2it} is a vector of observations on endogenous, time-varying variables (possibly) correlated with μ_i
  - Z_{1it} is a vector of observations on exogenous, time-invariant variables
  - Z_{2it} is a vector of observations on endogenous, time-invariant variables (possibly) correlated μ_i

Basic steps of Hausman–Taylor/Amemiya–MaCurdy estimators

- We start by estimating β_1 and β_2 using within estimator and obtaining the residual υ_{it} (note, we do not have estimates of δ_1 and δ_2),
- Compute intra-group temporal means of residual, $\bar{\upsilon}_i$
- Estimates δ_{1IV} and δ_{2IV} are obtained by regressing the within residuals on Z_{1i} and Z_{2i}, using X_{1it} and Z_{1i} as instruments (i.e. we do 2SLS).
- Estimates of β_1 and β_2 and estimates δ_{1IV} and δ_{2IV}, can be used to obtain sets of within and overall residuals.
- The estimated variance components can then be used to perform the standard random-effects GLS transform on each of the variables.
- We can than use the transformed variables to perform 2SLS with instruments $X_{1it} - \bar{X}_{1i}$, $X_{2it} - \bar{X}_{2i}$, $\bar{X}_{1i}$ and Z_{1i}.

- Amemiya and MaCurdy (1986) place stricter requirements on the instruments that vary within panels to obtain a more efficient estimator. Specifically, they assume that X_{1it} is orthogonal to unit effects in every period.

Random coefficients model

- RCM generalizes random effects from the intercept to all parameters of interest. In random-coefficients models, we wish to treat the parameter vector as a realization (in each panel) of a stochastic process.

- The random coefficient model is a comprise between assuming absolute homogeneity and absolute heterogeneity. In other words, RCM is a compromise between estimating fully pooled equation and fully unpooled, that is separate OLS for each unit, data.

- The RCM uses idea of borrowing strength. This Bayesian notion shrinks each of the individual unit OLS estimate back to the overlap (pooled) estimates.

- **The RCM is very important approach in panel data but not that interesting in TSCS data.** Namely, RCS does not appear to solve the problem of modeling heterogeneity for TSCS data, but it may provide more accurate assessment of the standard errors of the parameters of interest.

Random coefficients model

- Random-coefficients model is based on Swamy (1970) estimator. The resulting GLS estimator is a matrix-weighted average of the panel-specific OLS estimators.

- In a random-coefficients model, the parameter heterogeneity is treated as stochastic variation. The derivation of the estimator assumes that the cross-sectional specific coefficient vector β_i is the outcome of a random process with mean vector β and covariance matrix Σ.

$$y_i = X_i\beta_i + \varepsilon_i$$
$$\beta_i = \beta + \nu_i \qquad E(\nu_i) = 0 \qquad E(\nu_i\nu_i') = \Sigma$$

- Our goal is to find $\hat{\beta}$ and $\hat{\Sigma}$.

Random coefficients model

- As we assume that the cross-sectional specific coefficient vector β_i is the outcome of a random process with mean vector β and covariance matrix Σ, the previous equation can be expressed as:

$$X_i(\beta + \nu_i) + \varepsilon_i = X_i\beta + (\boldsymbol{X_i \nu_i + \varepsilon_i}) = X_i\beta + \boldsymbol{\omega_i}$$
$$E(\omega_i\omega_i') = E\{(X_i\nu_i + \varepsilon_i)(X_i\nu_i + \varepsilon_i)'\} = \sigma_i^2 I + X_i\Sigma X_i' = \Pi_i$$

- The GLS estimator of $\hat{\beta}$

$$\hat{\beta} = \left(\sum_i X_i'\Pi_i^{-1} X_i\right)^{-1} \sum_i X_i'\Pi_i^{-1} y_i = \sum_{i=1}^{m} W_i b_i$$

b_i = OLS panel-specific estimator, $b_i = (X_i^T X_i)^{-1} X_i^T y_i$ - **(notice projection matrix)** - where:

$$W_i = \left\{\sum_{i=1}^{m} (\Sigma + V_i)^{-1}\right\}^{-1} (\Sigma + V_i)^{-1}$$

and $V_i = \sigma_i^2 (X_i^T X_i)^{-1}$ is variance covariance β estimate of each panel

Random coefficients model

- To calculate the above estimator $\hat{\beta}$ for the unknown Σ and V_i parameters, we use the two-step approach suggested by Swamy

- The two-step procedure begins with the usual OLS estimates of β_i. With those estimates, we may proceed by obtaining estimates of $\hat{V}_i$ i and $\hat{\Sigma}$ (and thus $\widehat{W}_i$) and then obtain an estimate of β.

- In the second step, the feasible best linear predictor of β_i is given by:

$$\hat{\beta}_i = \left(\hat{\Sigma}^{-1} + \hat{V}_i^{\;-1}\right)^{-1} \left(\hat{\Sigma}^{-1}\hat{\beta} + \hat{V}_i^{\;-1} b_i\right)$$