Time series econometrics. Panel Data

CMF. Quantitive analytics,

Pyltsyna Ekaterina

Outline

1- Problem

2- Fixed effects models

3- Random effects models

4- Testing FE and RE models

5- Spatial Dynamic Panel Model

6- Bayesian Estimation of the Models

7- Conclusion

1. Problem

Observations correlated depending on their locations, are called spatial data.

Spatial data obtained in successive periods is called spatio-temporal data.

If they are independent over time, is called spatial panel data.

Due to the spatial or spatio-temporal correlation of data, it is necessary to determine their correlation structure and apply it in data analysis.

1. Problem

This requires determining the spatial or spatio-temporal covariance function, which is usually unknown and must be estimated.

A key issue in panel data modeling is variability among the experimental units.

Because of the heterogeneity between spatial locations each location may have different effects on data.

These effects can be either fixed or random.

2. Fixed Effects Models

- 2.1 Basic fixed-effects model
- 2.2 Estimation and inference
- 2.3 Model specification and diagnostics

2.1 Basic fixed effects model

- **Basic Elements**
- Subject i is observed on T_i occasions;
  - $i = 1, \ldots, n,$
  - $T_i \leq \square\, T$, the maximal number of time periods.
- The response of interest is y_{it}.
- The K explanatory variables are $\mathbf{x}_{it} = \{x_{it1}, x_{it2}, \ldots, x_{itK}\}'$, a vector of dimension $K \times 1$.
- The population parameters are $\square = (\square_1, \ldots, \square_K)'$, a vector of dimension $K \times 1$.

Observables Representation of the Linear Model

- E $y_{it} = \square + \square_1\, x_{it1} + \square_2\, x_{it2} + \ldots + \square_K\, x_{itK}$.
- $\{x_{it,1}, \ldots, x_{it,K}\}$ are nonstochastic variables.
- Var $y_{it} = \sigma^2$.
- $\{ y_{it} \}$ are independent random variables.
- $\{ y_{it} \}$ are normally distributed.
- The *observable* variables are $\{x_{it,1}, \ldots, x_{it,K}, y_{it}\}$.
- Think of $\{x_{it,1}, \ldots, x_{it,K}\}$ as defining a strata.
  - We take a random draw, y_{it}, from each strata.
  - Thus, we treat the x's as nonstochastic
  - We are interested in the distribution of y, conditional on the x's.

Error Representation of the Linear Model

- $y_{it} = \alpha + \beta_1 x_{it,1} + \beta_2 x_{it,2} + \ldots + \beta_K x_{it,K} + \varepsilon_{it}$
  where E $\varepsilon_{it} = 0$.
- $\{x_{it,1}, \ldots, x_{it,K}\}$ are nonstochastic variables..
- Var $\varepsilon_{it} = \sigma^2$.
- $\{ \varepsilon_{it} \}$ are independent random variables.

- This representation is based on the Gaussian theory of errors – it is centered on the *unobservable* variable ε_{it} .
- Here, ε_{it} are i.i.d., mean zero random variables.

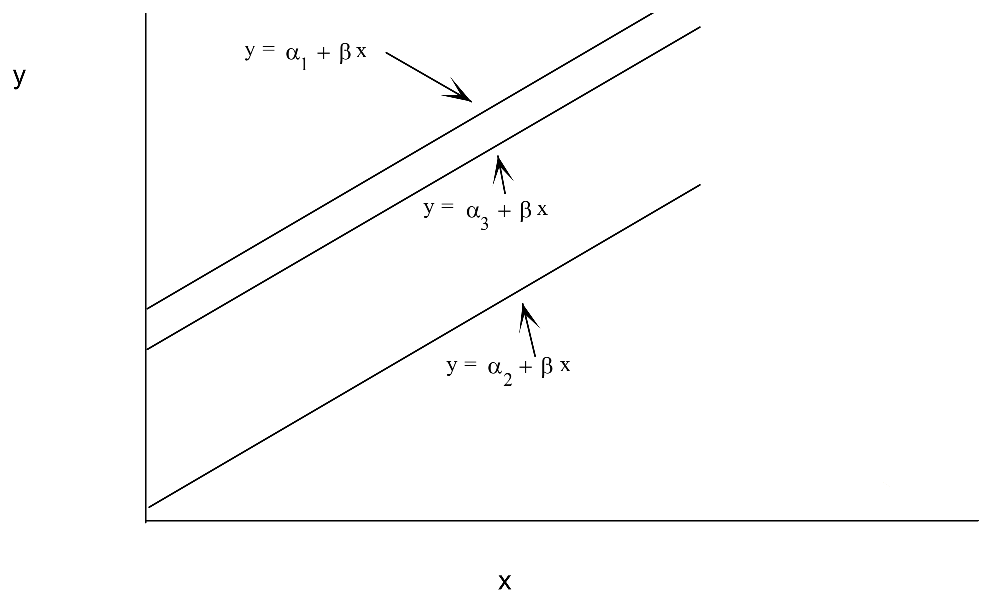
Heterogeneous model

- We now introduce a subscript on the intercept term, to account for heterogeneity.
- $\mathrm{E}\, y_{it} = \alpha_i + \beta_1\, x_{it,1} + \beta_2\, x_{it,2} + \ldots + \beta_K\, x_{it,K}$.
- For short-hand, we write this as

$$\mathrm{E}\, y_{it} = \alpha_i + \mathbf{x}_{it}'\, \boldsymbol{\beta}$$

Analysis of covariance model

- The intercept parameter, □, varies by subject.
- The population parameters □ do not but control for the common effect of the covariates **x**.
- Because the errors are mean zero, the expected response is $\mathrm{E}\, y_{it} = \square_i + \mathbf{x}_{it}'\, \square$.



Parameters of interest

- The common effects of the explanatory variables are dictated by the sign and magnitude of the betas (□´s)
  - These are the parameters of interest
- The intercept parameters vary by subject and account for different behavior of subjects.
  - The intercept parameters control for the heterogeneity of subjects.
  - Because they are of secondary interest, the intercepts are called *nuisance* parameters.

Time-specific analysis of covariance

- The basic model also is a traditional analysis of covariance model.
- The basic fixed-effects model focuses on the mean response and assumes:
  - no serial correlation (correlation over time)
  - no cross-sectional (contemporaneous) correlation (correlation between subjects)
- Hence, no special relationship between subjects and time is assumed.
- By interchanging *i* and *t*, we may consider the model

$$y_{it} = \square_t + \mathbf{x}_{it}' \, \square + \square_{it} .$$

- The parameters $\square_t$ are time-specific variables that do not depend on subjects.

Subject and time heterogeneity

- Typically, the number of subjects, n, substantially exceeds the maximal number of time periods, T.
- Typically, the heterogeneity among subjects explains a greater proportion of variability than the heterogeneity among time periods.
- Thus, we begin with the “basic” model $y_{it} = \square_i + \mathbf{x}_{it}' \square + \square_{it}$.
  - This model allows explicit parameterization of the subject-specific heterogeneity.
  - By using binary variables for the time dimension, we can easily incorporate time-specific parameters.

2.2 Estimation and inference

- Least squares estimates
- By the Gauss-Markov theorem, the best linear unbiased estimates are the ordinary least square (*ols*) estimates.
- These are given by:

$$\mathbf{b} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T_i} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T_i} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i) \right)$$

- and $a_i = \bar{y}_i - \bar{\mathbf{x}}_i'\mathbf{b}$
- Here, $\bar{y}_i$ and $\bar{\mathbf{x}}_i$ are averages of $\{y_{it}\}$ and $\{\mathbf{x}_{it}\}$ over time.
- Time-constant x's prevent one from getting unique estimates of **b** !!!

Estimation details

- Although there are $n+K$ unknown parameters, the calculation of the *ols* estimates requires inversion of only a $K \times K$ matrix.
- The *ols* estimate of β can also be expressed as a weighted average of estimates of subject-specific parameters.
  - Suppose that all parameters are subject-specific so that the model is $y_{it} = \alpha_i + \mathbf{x}_{it}' \beta_i + \varepsilon_{it}$
  - The ols estimate of β_i turns out to be

$$\mathbf{b}_i = \left(\sum_{t=1}^{T_i} \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i \right) \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i \right)' \right)^{-1} \left(\sum_{t=1}^{T_i} \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i \right) \left(y_{it} - \bar{y}_i \right) \right)$$

  - Define the weighting matrix $\mathbf{W}_i = \sum_{t=1}^{T_i} \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i \right) \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i \right)'$
  - With this weight, we can express the *ols* estimate of β as

$$\mathbf{b} = \left(\sum_{i=1}^{n} \mathbf{W}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{W}_i \mathbf{b}_i$$

  - a weighted average of subject-specific parameter estimates.

Properties of estimates

- Both a_i and $\mathbf{b}$ have the usual properties of *ols* regression estimators
  - They are unbiased estimators.
  - By the Gauss-Markov theorem, they are minimum variance among the class of unbiased estimates.
- To see this, consider an expression of the ols estimate of **b**,

$$\mathbf{b} = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \mathbf{W}_{it,1} y_{it} \qquad \mathbf{W}_{it,1} = \left(\sum_{i=1}^{n} \mathbf{W}_i \right)^{-1} \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i \right)$$

  - That is, **b** is a linear combination of responses.
  - If the responses are normally distributed, then so is **b**.
- The variance of **b** turns out to be $\operatorname{Var} \mathbf{b} = \sigma^2 \left(\sum_{i=1}^{n} \mathbf{W}_i \right)^{-1}$

ANOVA and standard errors

- This follows the usual regression set-up.
- We define the residuals as $e_{it} = y_{it}$ - $(a_i + \mathbf{x}_{it}' \mathbf{b})$.
- The error sum of squares is $Error\ SS = \Sigma_{it}\, e_{it}^{\,2}$.
- The mean square error is $$s^2 = \frac{Error\ SS}{N-(n+K)} = Error\ MS$$
- the residual standard deviation is s.
- The standard errors of the slope estimates are from the square root of the diagonal of the estimated variance matrix $$\widehat{\mathrm{Var}\,\mathbf{b}} = s^2 \left(\sum_{i=1}^{n} \mathbf{W}_i \right)^{-1}$$

Consistency of estimates

- As the number of subjects (n) gets large, then **b** approaches **b**.
  - Specifically, weak consistency means approaching (convergence) in probability.
  - This is a direct result of the unbiasedness and an assumption that $\sum_i \mathbf{W}_i$ grows without bound.
- As n gets large, the intercept estimates a_i do not approach a_i.
  - They are inconsistent.
  - Intuitively, this is because we assume that the number of repeated measurements of a_i is T_i, a bounded number.

Other large sample approximations

- Typically, the number of subjects is large relative to the number of time periods observed.
- Thus, in deriving large sample approximations of the sampling distributions of estimators, assume that $n \to \infty$ although T remains fixed.
- With this assumption, we have a central limit theorem for the slope estimator.
  - That is, **b** is approximately normally distributed even though responses are not.
  - The approximation improves as n becomes large.
- Unlike the usual regression set-up, this is not true for the intercepts. If the responses are not normally distributed, then α_i are not even approximately normal.

2.3 Model specification and diagnostics

- Pooling Test
- Added variable plots
- Influence diagnostics
- Cross-sectional correlations
- Heteroscedasticity

Pooling test

- Test whether the intercepts take on a common value, say a.
- Using notation, we wish to test the null hypothesis

$$H_0\text{: } a_1 = a_2 = \ldots = a_n = a.$$

- This can be done using the following partial F- (Chow) test:
  - Run the "full model" $y_{it} = \square_i + \mathbf{x}_{it}' \, \square + \square_{it}$ to get Error SS and s^2 .
  - Run the "reduced model" $y_{it} = \square + \mathbf{x}_{it}' \, \square + \square_{it}$ to get (Error SS)$_{reduced}$.
  - Compute the partial F-statistic,

$$F - ratio = \frac{(Error\ SS)_{reduced} - Error\ SS}{(n-1)s^2}$$

  - Reject H_0 if F exceeds a quantile from an F-distribution with numerator degrees of freedom $df_1 = n$-1 and denominator degrees of freedom $df_2 = N$-(n+K).

Added variable plot

- An *added variable plot* (also called a partial regression plot) is a standard graphical device used in regression analysis
- *Purpose:* To view the relationship between a response and an explanatory variable, after controlling for the linear effects of other explanatory variables.
- Added variable plots allow us to visualize the relationship between y and each x, without forcing our eye to adjust for the differences induced by the other x's.
- The basic added variable plot is a special case.

Procedure for making an added variable plot

- Select an explanatory variable, say x_j.
- Run a regression of y on the other explanatory variables (omitting x_j)
  - calculate the residuals from this regression. Call these residuals e_1.
- Run a regression of x_j on the other explanatory variables (omitting x_j)
  - calculate the residuals from this regression. Call these residuals e_2.
- The plot of e_1 versus e_2 is an *added variable plot*.

Correlations and added variable plots

- Let corr(e_1, e_2) be the correlation between the two sets of residuals.
  - It is related to the *t*-statistic of x_j, $t(b_j)$, from the full regression equation (including x_j) through:

$$\operatorname{corr}(e_1, e_2) = \frac{t(b_j)}{\sqrt{t(b_j)^2 + N - (n + K)}}$$

  - Here, K is the number of regression coefficients in the full regression equation and N is the number of observations.
- Thus, the *t*-statistic can be used to determine the correlation coefficient of the added variable plot without running the three step procedure.
- However, unlike correlation coefficients, the added variable plot allows us to visualize potential nonlinear relationships between y and x_j.

Influence diagnostics

- Influence diagnostics allow the analyst to understand the impact of individual observations and/or subjects on the estimated model
- Traditional diagnostic statistics are observation-level
  - of less interest in panel data analysis
  - the effect of unusual observations is absorbed by subject-specific parameters.
- Of greater interest is the impact that an entire subject has on the population parameters.
- We use the statistic $$B_i(\mathbf{b}) = \left(\mathbf{b} - \mathbf{b}_{(i)}\right)' \left(\sum_{i=1}^{n} \mathbf{W}_i \right) \left(\mathbf{b} - \mathbf{b}_{(i)}\right) / K$$
- Here, $\mathbf{b}_{(i)}$ is the *ols* estimate **b** calculated with the *i*th subject omitted.

Calibration of influence diagnostic

- The panel data influence diagnostic is similar to Cook's distance for regression.
  - Cook's distance is calculated at the observational level yet $B_i(\mathbf{b})$ is at the subject level
- The statistic $B_i(\mathbf{b})$ has an approximate c^2 (chi-square) with K degrees of freedom
  - Observations with a "large" value of $B_i(\mathbf{b})$ may be influential on the parameter estimates.
  - Use quantiles of the c^2 to quantify the adjective "large."
- Influential observations warrant further investigation
  - they may need correction, additional variable specification to accommodate differences or deletion from the data set.

Cross-sectional correlations

- The basic model assumes independence between subjects.
  - Looking at a cross-section of subjects, we assume zero cross-sectional correlation, that is, r_{ij} = Corr (y_{it}, y_{jt}) = 0 for $i \neq j$.
- Suppose that the "true" model is $y_{it} = l_t + \mathbf{x}_{it}'\mathbf{b} + e_{it}$, where l_t is a random temporal effect that is common to all subjects.
  - This yields Var $y_{it} = s_l^2 + s^2$
  - The covariance between observations at the same time but from different subjects is Cov $(y_{it}, y_{jt}) = s_l^2$, $i \neq j$.
  - Thus, the cross-sectional correlation is

$$\mathrm{Corr}\left(y_{it}, y_{jt}\right) = \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma^2}$$

Testing for cross-sectional correlations

- To test H_0: $r_{ij} = 0$ for all $i \neq j$, assume that $T_i = T$.
  - Calculate model residuals $\{e_{it}\}$.
  - For each subject i, calculate the ranks of each residual.
    - That is, define $\{r_{i,1}, \ldots, r_{i,T}\}$ to be the ranks of $\{e_{i,1}, \ldots, e_{i,T}\}$.
    - Ranks will vary from 1 to T, so the average rank is $(T+1)/2$.
  - For the ith and jth subject, calculate the rank correlation coefficient (Spearman's correlation)

$$sr_{ij} = \frac{\sum_{t=1}^{T} \left(r_{i,t} - (T+1)/2\right)\left(r_{j,t} - (T+1)/2\right)}{\sum_{t=1}^{T} \left(r_{i,t} - (T+1)/2\right)^2}$$

  - Calculate the average Spearman's correlation and the average squared Spearman's correlation

$$R_{AVE} = \frac{1}{n(n-1)/2} \sum_{\{i<j\}} sr_{ij} \qquad R_{AVE}^2 = \frac{1}{n(n-1)/2} \sum_{\{i<j\}} \left(sr_{ij}\right)^2$$

    - Here, $S_{\{i<j\}}$ means sum over i=1, ..., j-1 and j=2, ..., n.

Calibration of cross-sectional correlation test

- We compare R^2_{ave} to a distribution that is a weighted sum of chi-square random variables (Frees, 1995).
- Specifically, define

$$Q = \mathrm{a}(T)\ (\mathrm{c}_1^{\,2} - (T-1)) + \mathrm{b}(T)\ (\mathrm{c}_2^{\,2} - T(T-3)/2)\ .$$

  - Here, $\mathrm{c}_1^{\,2}$ and $\mathrm{c}_2^{\,2}$ are independent chi-square random variables with T-1 and $T(T$-3)/2 degrees of freedom, respectively.
  - The constants are

$$\mathrm{a}(T) = 4(T+2) / (5(T-1)^2(T+1))$$

and

$$\mathrm{b}(T) = 2(5T+6) / (5T(T-1)(T+1))\ .$$

Calculation short-cuts

- Rule of thumb for cut-offs for the Q distributon .
- To calculate R^2_{ave}
  - Define $$Z_{i,t,u} = \frac{1}{T^3 - T} 12\left(r_{i,t} - (T+1)/2\right)\left(r_{i,u} - (T+1)/2\right)$$
  - For each t, u, calculate $S_i\, Z_{i,t,u}$ and $S_i\, Z_{i,t,u}{}^2$.
  - We have $$R^2_{AVE} = \frac{1}{n(n-1)} \sum_{\{t,u\}} \left(\left(\sum_i Z_{i,t,u}\right)^2 - \sum_i Z^2_{i,t,u} \right)$$
  - Here, $S_{\{t,u\}}$ means sum over t=1, ..., T and u=1, ..., T.
  - Although more complex in appearance, this is a much faster computation form for R^2_{ave}.
  - Main drawback - the asymptotic distribution is only available for balanced data.

Heteroscedasticity

- Carroll and Ruppert (1988) provide a broad treatment
- Here is a test due to Breusch and Pagan (1980).
  - H_a: Var $e_{it} = s^2 + \mathbf{g}¢\ \mathbf{w}_{it}$, where $\mathbf{w}_{it}$ is a known vector of weighting variables and $\mathbf{g}$ is a p-dimensional vector of parameters.
  - H_0: Var $e_{it} = s^2$. This procedure is:
    - Fit a regression model and calculate the model residuals, $\{r_{it}\}$.
    - Calculate squared standardized residuals, $r_{it}^{*2} = r_{it}^2 / (Error\ SS / N)$
    - Fit a regression model of r_{it}^{*2} on $\mathbf{w}_{it}$.
    - The test statistic is $LM = (Regress\ SS_w)/2$, where $Regress\ SS_w$ is the regression sum of squares from the model fit in step 3.
    - Reject the null hypothesis if LM exceeds a percentile from a chi-square distribution with p degrees of freedom. The percentile is one minus the significance level of the test.

- Random Effects Model

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it} \ (t = 1, 2, ..., T_i)$$

$$\Downarrow$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + u_i\mathbf{i}_{T_i} + \mathbf{e}_i \ (i = 1, 2, ..., N)$$

  - u_i is random, independent of e_{it} and $\mathbf{x}_{it}$.
  - Define $e_{it} = u_i + e_{it}$ the error components.

Random Effects Model

- Assumptions
  - Strict Exogeneity

$$E(e_{it} \mid \mathbf{X}) = 0,\ E(u_i \mid \mathbf{X}) = 0 \Rightarrow E(\varepsilon_{it} \mid \mathbf{X}) = 0$$

    - **X** includes a constant term, otherwise $E(u_i|\mathbf{X})=u$.
  - Homoschedasticity

$$Var(e_{it} \mid \mathbf{X}) = \sigma_e^2,\ Var(u_i \mid \mathbf{X}) = \sigma_u^2,\ Cov(u_i, e_{it}) = 0$$

$$\Rightarrow Var(\varepsilon_{it} \mid \mathbf{X}) = \sigma_\varepsilon^2 = \sigma_e^2 + \sigma_u^2$$

  - Constant Auto-covariance (within panels)

$$Var(\boldsymbol{\varepsilon}_i \mid \mathbf{X}) = \sigma_e^2 \mathbf{I}_{T_i} + \sigma_u^2 \mathbf{i}_{T_i} \mathbf{i}_{T_i}'$$

Random Effects Model

- Assumptions
  - Cross Section Independence

$$Var(\boldsymbol{\varepsilon}_i \mid \mathbf{X}) = \Omega_i = \sigma_e^2 \mathbf{I}_{T_i} + \sigma_u^2 \mathbf{i}_{T_i} \mathbf{i}_{T_i}^{'}$$

$$Var(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \boldsymbol{\Omega} = \begin{bmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \Omega_N \end{bmatrix}$$

Random Effects Model

- Extensions
  - Weak Exogeneity

$$E(\varepsilon_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT_i}) = E(\varepsilon_{it} \mid \mathbf{X}_i) = 0$$

$$E(\varepsilon_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it}) = 0$$

$$E(\varepsilon_{it} \mid \mathbf{x}_{it}) = 0$$

  - Heteroscedasticity

$$Var(\varepsilon_{it} \mid \mathbf{X}_i) = \sigma_{u_i}^2 + Var(e_{it} \mid \mathbf{X}_i) = \sigma_{u_i}^2 + \begin{cases} \sigma_{e_{it}}^2 \\ \sigma_{e_i}^2 \end{cases}$$

Random Effects Model

- Extensions
  - Serial Correlation

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + u_i + e_{it},\ e_{it} = \begin{cases} \rho e_{it-1} + v_{it} \\ \rho_i e_{it-1} + v_{it} \end{cases}$$

  - Spatial Correlation

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it},\ \varepsilon_{it} = \lambda \sum_j w_{ij}\varepsilon_{jt} + e_{it},\ e_{it} = u_i + v_{it}$$

Model Estimation: GLS

- Model Representation

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i = u_i\mathbf{i}_{T_i} + \mathbf{e}_i$$

$$E(\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i) = \mathbf{0}$$

$$Var(\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i) = \Omega_i = \sigma_e^2\mathbf{I}_{T_i} + \sigma_u^2\mathbf{i}_{T_i}\mathbf{i}_{T_i}'$$

$$= \sigma_e^2\left[Q_i + \frac{\sigma_e^2 + T_i\sigma_u^2}{\sigma_e^2}\left(\mathbf{I}_{T_i} - Q_i\right)\right]$$

$$where\ Q_i = \mathbf{I}_{T_i} - \frac{1}{T_i}\mathbf{i}_{T_i}\mathbf{i}_{T_i}'$$

Model Estimation: GLS

- GLS

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{y} = \left[\sum_{i=1}^{N}\mathbf{X}_i\Omega_i^{-1}\mathbf{X}_i\right]^{-1}\sum_{i=1}^{N}\mathbf{X}_i\Omega_i^{-1}\mathbf{y}_i$$

$$Var(\hat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \left[\sum_{i=1}^{N}\mathbf{X}_i\Omega_i^{-1}\mathbf{X}_i\right]^{-1}$$

$$where\ \Omega_i^{-1} = \frac{1}{\sigma_e^2}\left[Q_i + \frac{\sigma_e^2}{\sigma_e^2 + T_i\sigma_u^2}\left(\mathbf{I}_{T_i} - Q_i\right)\right]$$

$$and\ \Omega_i^{-1/2} = \frac{1}{\sigma_e}\left[Q_i + \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T_i\sigma_u^2}}\left(\mathbf{I}_{T_i} - Q_i\right)\right]$$

Model Estimation: RE-OLS

- Partial Group Mean Deviations

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + (u_i + e_{it})$$

$$\overline{y}_i = \overline{\mathbf{x}}_i^{'}\boldsymbol{\beta} + (u_i + \overline{e}_i)$$

$$\Downarrow \quad \theta_i = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T_i\sigma_u^2}}$$

$$y_{it} - \theta_i\overline{y}_i = (\mathbf{x}_{it}^{'} - \theta_i\overline{\mathbf{x}}_i^{'})\boldsymbol{\beta} + [(1-\theta_i)u_i + (e_{it} - \theta_i\overline{e}_i)]$$

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}^{'}\boldsymbol{\beta} + \tilde{\varepsilon}_{it}$$

Model Estimation: RE-OLS

- Model Assumptions

$$E(\tilde{\varepsilon}_{it} \mid \tilde{\mathbf{x}}_i^{'}) = 0$$

$$Var(\tilde{\varepsilon}_{it} \mid \tilde{\mathbf{x}}_i^{'}) = (1-\theta_i)^2 \sigma_u^2 + (1 - 2\theta_i / T_i + \theta_i^2 / T_i)\sigma_e^2 = \sigma_e^2$$

$$Cov(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{is} \mid \tilde{\mathbf{x}}_i^{'}) = (1-\theta_i)^2 \sigma_u^2 + (-2\theta_i / T_i + \theta_i^2 / T_i)\sigma_e^2 = 0, t \neq s$$

$$Note: \theta_i = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T_i \sigma_u^2}}$$

- OLS

$$\hat{\boldsymbol{\beta}}_{OLS} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i^{'}\tilde{\mathbf{X}}_i\right]^{-1} \sum_{i=1}^{N} \tilde{\mathbf{X}}_i \tilde{\mathbf{y}}_i$$

$$\hat{Var}(\hat{\boldsymbol{\beta}}_{OLS}) = \hat{\sigma}_e^2 (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} = \hat{\sigma}_e^2 \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i^{'}\tilde{\mathbf{X}}_i\right]^{-1}$$

$$\hat{\sigma}_e^2 = \hat{\tilde{\boldsymbol{\varepsilon}}}'\hat{\tilde{\boldsymbol{\varepsilon}}} / (NT - K), \; \hat{\tilde{\boldsymbol{\varepsilon}}} = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}$$

Model Estimation: RE-OLS

- Need a consistent estimator of q:

$$\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2 + T_i \hat{\sigma}_u^2}}$$

  - Estimate the fixed effects model to obtain $\hat{\sigma}_e^2$
  - Estimate the between model to obtain $T\hat{\sigma}_u^2 + \hat{\sigma}_v^2$
  - Or, estimate the pooled model to obtain $\hat{\sigma}_e^2 + \hat{\sigma}_u^2$
  - Based on the estimated *large sample variances*, it is safe to obtain $0 < \hat{\theta} < 1$

Model Estimation: RE-OLS

- Panel-Robust Variance-Covariance Matrix
  - Consistent statistical inference for general heteroscedasticity, time series and cross section correlation.

$$\hat{V}ar(\hat{\boldsymbol{\beta}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']$$

$$= \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i\right]^{-1} \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \hat{\tilde{\boldsymbol{\varepsilon}}}_i \hat{\tilde{\boldsymbol{\varepsilon}}}_i' \tilde{\mathbf{X}}_i\right] \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i\right]^{-1}$$

$$= \left[\sum_{i=1}^{N} \sum_{t=1}^{T_i} \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}_{it}'\right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{T_i} \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}_{is}' \hat{\tilde{\varepsilon}}_{it} \hat{\tilde{\varepsilon}}_{is}\right] \left[\sum_{i=1}^{N} \sum_{t=1}^{T_i} \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}_{it}'\right]^{-1}$$

$$\hat{\tilde{\boldsymbol{\varepsilon}}}_i = \tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}}, \quad \hat{\tilde{\varepsilon}}_{it} = \tilde{y}_{it} - \tilde{\mathbf{x}}_{it}' \hat{\boldsymbol{\beta}}$$

Model Estimation: ML

- Log-Likelihood Function

$$y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + (u_i + e_{it}) = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it} \quad (t = 1, 2, \ldots, T_i)$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad (i = 1, 2, \ldots, N)$$

$$\boldsymbol{\varepsilon}_i \sim iidn(\mathbf{0}, \Omega_i), \quad \Omega_i = \sigma_e^2\mathbf{I}_{T_i} + \sigma_u^2\mathbf{i}_{T_i}\mathbf{i}_{T_i}^{'}$$

$$\Downarrow$$

$$ll_i(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}_i, \mathbf{X}_i) = -\frac{T_i}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\Omega_i\right| - \frac{1}{2}\boldsymbol{\varepsilon}_i\Omega_i^{-1}\boldsymbol{\varepsilon}_i$$

Model Estimation: ML

- ML Estimator

$$(\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2, \hat{\sigma}_u^2)_{ML} = \arg\max \sum_{i=1}^{N} ll_i(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}_i, \mathbf{X}_i)$$

where

$$ll_i(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}_i, \mathbf{X}_i) = -\frac{T_i}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\Omega_i\right| - \frac{1}{2}\boldsymbol{\varepsilon}_i\Omega_i^{-1}\boldsymbol{\varepsilon}_i$$

$$= -\frac{T_i}{2}\ln\left(2\pi\sigma_e^2\right) - \frac{1}{2}\ln\left(\frac{\sigma_e^2 + T\sigma_u^2}{\sigma_e^2}\right)$$

$$-\frac{1}{2\sigma_e^2}\left\{\left[\sum_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}^{'}\boldsymbol{\beta})^2\right] - \frac{\sigma_u^2}{\sigma_e^2 + T_i\sigma_u^2}\left[\sum_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}^{'}\boldsymbol{\beta})\right]^2\right\}$$

Hypothesis Testing
To Pool or Not To Pool, Continued

- Test for $Var(u_i) = 0$, that is

$$Cov(\varepsilon_{it},\varepsilon_{is}) = Cov(u_i + e_{it}, u_i + e_{is}) = Cov(e_{it}, e_{is})$$

  - If T_i=T for all i, the Lagrange-multiplier test statistic (Breusch-Pagan, 1980) is:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\hat{\mathbf{e}}'(I_N \otimes J_T)\hat{\mathbf{e}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}} - 1\right]^2 = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2} - 1\right]^2 \sim \chi^2(1)$$

$$where\ \hat{e}_{it} = y_{it} - \left[\mathbf{x}_{it}' \quad 1\right]\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{u}\end{bmatrix}_{Pooled},\ J_T = \mathbf{i}_T\mathbf{i}_T'$$

Hypothesis Testing
To Pool or Not To Pool, Continued

- For unbalanced panels, the modified Breusch-Pagan LM test for random effects (Baltagi-Li, 1990) is:

$$LM = \frac{\left(\sum_{i=1}^{N} T_i\right)^2}{2\left(\sum_{i=1}^{N} T_i(T_i - 1)\right)} \left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i} \hat{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T_i} \hat{e}_{it}^2} - 1 \right]^2 \sim \chi^2(1)$$

- Alternative one-side test:

$$\sqrt{LM} \sim N(0,1)\ under\ H_0$$

$$P-Value : \Pr_n(z > \sqrt{LM})$$

Hypothesis Testing
Fixed Effects vs. Random Effects

$$H_0 : Cov(u_i, \mathbf{x}_{it}^{'}) = 0 \; (random\ effects)$$

$$H_1 : Cov(u_i, \mathbf{x}_{it}^{'}) \neq 0 \; (fixed\ effects)$$

Estimator	Random Effects $E(u_i\|X_i) = 0$	Fixed Effects $E(u_i\|X_i)$ =/= 0
GLS or RE-OLS (Random Effects)	Consistent and Efficient	Inconsistent
LSDV or FE-OLS (Fixed Effects)	Consistent Inefficient	Consistent Possibly Efficient

Hypothesis Testing
Fixed Effects vs. Random Effects

- Alternative (Asym. Eq.) Hausman Test
  - Estimate any of the random effects models
    - $(y_{it} - \theta\bar{y}_i) = (\mathbf{x}_{it}^{'} - \theta\bar{\mathbf{x}}_i^{'})\boldsymbol{\beta} + (\mathbf{x}_{it}^{'} - \bar{\mathbf{x}}_i^{'})\boldsymbol{\gamma} + e_{it}$

      $(or, \text{random effects model}: y_{it} = \mathbf{x}_{it}^{'}\boldsymbol{\beta} + (\mathbf{x}_{it}^{'} - \bar{\mathbf{x}}_i^{'})\boldsymbol{\gamma} + e_{it})$
    - $(y_{it} - \theta\bar{y}_i) = (\mathbf{x}_{it}^{'} - \theta\bar{\mathbf{x}}_i^{'})\boldsymbol{\beta} + \bar{\mathbf{x}}_i^{'}\boldsymbol{\gamma} + e_{it}$
    - $(y_{it} - \theta\bar{y}_i) = (\mathbf{x}_{it}^{'} - \theta\bar{\mathbf{x}}_i^{'})\boldsymbol{\beta} + \mathbf{x}_{it}^{'}\boldsymbol{\gamma} + e_{it}$
  - F Test that $\gamma = 0$

$$H_0 : \boldsymbol{\gamma} = 0 \quad \Leftrightarrow \quad H_0 : Cov(u_i, \mathbf{x}_{it}) = 0$$

Hypothesis Testing
Fixed Effects vs. Random Effects

- Ahn-Low Test (1996)
  - Based on the estimated errors (GLS residuals) of the random effects model, estimate the following regression:

$$\hat{\tilde{\varepsilon}}_{it} = \alpha + (X_{it} - \hat{\theta}\bar{X}_i)\boldsymbol{\beta} + \bar{X}_i\boldsymbol{\gamma} + e_{it}$$

$$\Rightarrow NTR^2 \sim \chi^2(\#\boldsymbol{\gamma})$$

Panel Regression Model (Matrix Form)

If we set $\boldsymbol{y}_t = (y_{1t}, \dots, y_{Nt})'$, $\boldsymbol{X}_t = (\boldsymbol{x}'_{1t}, \dots, \boldsymbol{x}'_{Nt})'$,
$\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{Nt})'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})$

Then the matrix form of PRM is given by

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \qquad t = 1, \cdots, T$$

Dynamic Panel Regression Model (DPRM)

$$\boldsymbol{y}_t = \lambda\boldsymbol{y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \qquad t = 1, \cdots, T$$

where $\boldsymbol{y}_{t-1}$ is the lagged variable observed at time *t-1* and λ is the lagged autoregressive coefficient.

Spatial Dynamic Panel Regression Model (SDPRM)

$$\boldsymbol{y}_t = \rho \boldsymbol{W} \boldsymbol{y}_t + \lambda \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

where ρ is spatial autoregressive coefficient and $\boldsymbol{W}$ is a spatial weight matrix:

$$\boldsymbol{W} = \begin{pmatrix} \text{۰} & w_{\text{۱۲}} & \text{۰} \cdots & w_{\text{۱}n} \\ w_{\text{۲۱}} & \text{۰} & \cdots & w_{\text{۲}n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n\text{۱}} & w_{n\text{۲}} & \cdots & \text{۰} \end{pmatrix}$$

$$w_{ij} = d_{ij}^{\ -\alpha}, \alpha > 0$$

$$\mathrm{d_{ij}} = d(\mathrm{s_i} - \mathrm{s_j}) = [|\mathrm{x_i} - \mathrm{x_j}|^{\mathrm{p}} + |\mathrm{y_i} - \mathrm{y_j}|^{\mathrm{p}}]^{\frac{1}{\mathrm{p}}}, \quad p \geq 1$$

Bayesian Estimation of DPRM:

Prior distributions:

Conjugate priors: $\sigma^2 \sim IG(a, b)$, $\boldsymbol{\beta} \sim N(\beta_0, \Sigma_0)$,

and $\lambda \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1})$, where λ_{min} and λ_{max} are minimum and maximum Eigen values of the weight matrix (San et al, 1999).

The posterior distribution is given by

$$f(\boldsymbol{\beta}, \lambda, \boldsymbol{\mu}, \sigma^2 | \boldsymbol{y}) \propto f(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\mu}, \sigma^2) f(\boldsymbol{\beta}) f(\lambda) f(\boldsymbol{\mu}) f(\sigma^2)$$

But this distribution has not close form.

To use Gibbs sampling the full conditionals are needed:

Full conditional of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}|(\boldsymbol{y}, \lambda, \boldsymbol{\mu}, \sigma^2) \sim N(\boldsymbol{B}^{-1}b, \boldsymbol{B}^{-1})$$

where

$$\boldsymbol{B} = (\sigma^{-2} \sum_{t=1}^{T} \boldsymbol{X}_t' \boldsymbol{X}_t + \Sigma_0^{-1})$$

$$b = 2[\, \sigma^{-2} \sum_{t=1}^{T} \boldsymbol{X}_t' (\boldsymbol{y}_t - \lambda \boldsymbol{y}_{t-1} - \boldsymbol{\mu}_t) + \Sigma_0^{-1} \boldsymbol{\beta}_0 \,]$$

Full conditional of $\boldsymbol{\sigma}^2$:

$\sigma^2|(\boldsymbol{y}, \boldsymbol{\beta}, \lambda, \boldsymbol{\mu}) \sim IG(a^*, b^*),$

where

$a^* = a + \frac{NT}{2},$

$b^* = \frac{\mathbf{1}}{\mathbf{2}} \sum_{\boldsymbol{t}=\mathbf{1}}^{\boldsymbol{T}} (\boldsymbol{y}_t - \lambda \boldsymbol{y}_{t-1} - \boldsymbol{X}_t \boldsymbol{\beta} - \boldsymbol{\mu}_t)'(\boldsymbol{y}_{\boldsymbol{t}} - \lambda \boldsymbol{y}_{t-1} - \boldsymbol{X}_t \boldsymbol{\beta} - \boldsymbol{\mu}_t) + b$

Full conditional of λ

$$\lambda|(\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\mu},\sigma^2) \sim N(\lambda_n,\gamma)$$

where

$$\lambda_n = \left(\sum_{t=1}^{T} \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}\right)^{-1}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \boldsymbol{\mu}_t)'\boldsymbol{y}_{t-1}$$

$$\gamma = \sigma^2\left(\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}\right)^{-1}$$