1. BASIC PANEL DATA MODELS

[1] Introduction to panel-data models

(1) Data structure:

Individuals, i = 1, 2, ... , N;

Time, t = 1, 2, ... , T, for each i.

(2) Types of Data:

- large N and small T (most labor data).
- small N and large T (macroeconomic data on G7).
- Both large (farm production).

(3) Balanced v.s. Unbalanced Data:

- Balanced: for any i, there are T observations.
- Unbalanced: T may different over i.

Comment:

- Unbalanced data can be used for regression model, but have some limitations on analysis of non-linear model such as probit or logit.
- The lecture will focus on balanced data.

(4) Available Panel Data:

- PSID (Panel Study of Income Dynamics)
  - Starts in 1968 with 4802 families
  - Currently, over than 10,000 families are included.
  - Over 5,000 variables.
  - Available through the internet.

- NLS (National Longitudinal Surveys of Labor Market Experience)
  - Includes five distinct segments of the labor force:

    Older men (age between 45 and 49 in 1966)

    Young men (between 14 and 24 in 1966)

    Mature women (age between 30 and 44 in 1966)

    Young women (age between 14 and 21 in 1966)

    Youths (age between 14 and 27 in 1979)

- CPS (Current Population Survey)
  - Monthly national household survey conducted by Census Bureau.
  - Focuses on unemployment rate and other labor force statistics.

(5) Benefits and Limitations of Panel Data Analysis

- Benefits:
  - Can control unobservable individual heteroskedasticity.
  - Rich information about C-S variations and dynamics.
  - Can avoid problems in T-S data, e.g., multicollinearity, aggregation bias and nonstationarity.
  - Can identify individual and time effects which cannot be identified by pure C-S or T-S data. (Union members are paid better because they are more productive or because their negotiation power is strong?)

- Limitations:
  - Large parts of panel data are unbalanced.
  - Measurement errors.
  - Most existing estimation techniques are for panel data with short-time horizon.

[2] Is Controlling Unobservables Important?

[Example from Stock and Watson, Ch. 8]

- Issue:
  - Do alcohol taxes help decrease traffic deaths?
- Data: auto_1.txt
  - 48 U.S. states (excluding Alaska and Hawaii): N = 48.
  - 1982 -1988: T =7.

Series	**Descriptions**
state	State ID (FIPS) Code
year	Year
spircons	Spirits Consumption
unrate	Unemployment Rate
perinc	Per Capita Personal Income
emppop	Employment/Population Ratio
beertax	Tax on Case of Beer
mlda	Minimum Legal Drinking Age
vmiles	Ave. Mile per Driver
jaild	Mandatory Jail Sentence
comserd	Mandatory Community Service
allmort	# of Vehicle Fatalities (#VF)
mrall	Vehicle Fatality Rate (VFR)

- OLS results:

dependent variable: VFR

variable	coeff.	std. err.	t-st
beertax	**0.1112**	**0.0624**	**1.7832**
mlda	-0.0297	0.0317	-0.9367
jailed	0.1959	0.0723	2.7085
comserd	0.1460	0.0813	1.7951
unrate	-0.0227	0.0143	-1.5852
lpinc	-1.9018	0.2265	-8.3957
yr83	-0.0900	0.0959	-0.9389
yr84	-0.0648	0.0996	-0.6504
yr85	-0.0783	0.1006	-0.7782
yr86	0.0632	0.1022	0.6185
yr87	0.1032	0.1067	0.9671
yr88	0.1404	0.1107	1.2679
cons	20.7805	2.3157	8.9738

R-Square = 0.3482

- What is going on here?

[Digression]

- Consider a simple multiple regression model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \ .$$

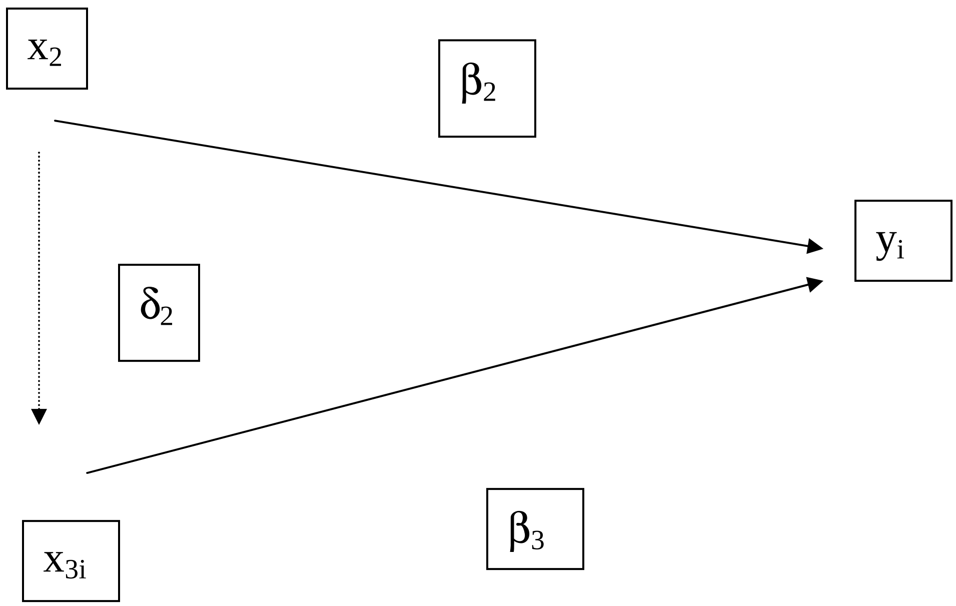
- What do β_2 and β_3 measure?

β_2 measures the direct (pure) effect of x_{2i} on y_i with x_{3i} held constant. Similarly, β_3 measures the direct effect of X_{3i} on Y_i with X_{2i} held constant.

- If you estimate $y_i = \alpha_1 + \alpha_2 x_{2i}$ + error instead?
  - Let $x_{3i} = \delta_1 + \delta_2 x_{2i} + v_i$. Substitute it into the correct model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3(\delta_1+\delta_2 x_{2i}+v_i) + \varepsilon_i$$
$$= (\beta_1 + \beta_3\delta_2) + (\beta_2 + \beta_3\delta_2)x_{2i} + (\varepsilon_i + \beta_3 v_i).$$

  - Thus, $\alpha_2 = \beta_2 + \delta_2\beta_3$.
  - Direct v.s. Indirect Effects



Total effect of $x_{2i} = \beta_2$ (direct) + $\delta_2\beta_3$ (indirect).

- When you do a regression omitting an important regressor, your estimated coefficients capture the total effects of your regressors!

[End of Digression]

- Return to our example:
  - Each state would have different preference for alcohol (say, Pal).
  - Pal and Beertax could be positively related ($\delta_2 > 0$).
  - Pal would have a positive effect on VFR ($\beta_3 > 0$).
  - The coefficient on Beertax captures the total effect:

$$\beta_2(-) + \delta_2(+)\times\beta_3(+) = (+).$$

- How could we control Pal using panel data?

[2] Fixed effects vs. Random effects

(1) Basic Model:

$$y_{it} = x_{it}\beta + z_i\gamma + u_{it} = h_{it}\delta + u_{it};\ u_{it} = \alpha + \varepsilon_{it}\ , \qquad \textbf{(1)}$$

where $i = 1, \ldots, N$ (cross-section unit),

$t = 1, \ldots, T$ (time),

$h_{it} = [x_{it}, z_i]$, $\delta = [\beta', \gamma']'$.

- Assumptions:
  - x_{it}: $1\times k$ vector of time-varying regressors.
  - z_i: $1\times g$ vector of time invariant regressors (overall intercept term will be included here).
  - ε_{it} i.i.d. $N(0,\sigma_\varepsilon^2)$.
  - α_i varies over i but constant over time (individual effects).

- Matrix Notation:
  - Define:

$$y_i = \begin{pmatrix} y_{i1} \\ : \\ y_{iT} \end{pmatrix}; X_i = \begin{pmatrix} x_{i1} \\ : \\ x_{iT} \end{pmatrix}; u_i = \begin{pmatrix} u_{i1} \\ : \\ u_{iT} \end{pmatrix}; \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ : \\ \varepsilon_{iT} \end{pmatrix}.$$

  - With this notation, we can write (1) by:

$$y_i = X_i\beta + e_T z_i + u_i = H_i\delta + u_i; u_i = \alpha_i e_T + \varepsilon_i\ , \qquad \textbf{(2)}$$

where $H_i = [X_i, e_T z_i]$ and e_T is $T\times 1$ vector of ones.

- Notation:
  - For T×p matrix M_i or T×1 vector m_i,

    $$\rightarrow M = \begin{pmatrix} M_1 \\ \vdots \\ M_N \end{pmatrix}; m = \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix}.$$

    → H denotes the data matrices of NT rows.
  - With this notation, model (2) can be rewritten for all observations as y = Hδ + u.

- Other matrix notation:

  $$e_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \text{ (T×1 vector of ones);}$$

  $$P_T = e_T(e_T'e_T)^{-1}e_T'$$

  $$= (1/T)e_Te_T' = \begin{pmatrix} \frac{1}{T} & \frac{1}{T} & \cdots & \frac{1}{T} \\ \frac{1}{T} & \frac{1}{T} & \cdots & \frac{1}{T} \\ \vdots & \vdots & & \vdots \\ \frac{1}{T} & \frac{1}{T} & & \frac{1}{T} \end{pmatrix} \text{ (T×T mean operator);}$$

$$Q_T = I_T - P_T = \begin{pmatrix} \frac{T-1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} \\ -\frac{1}{T} & \frac{T-1}{T} & \cdots & -\frac{1}{T} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{T} & -\frac{1}{T} & & \frac{T-1}{T} \end{pmatrix};$$

($T \times T$ deviation-from-mean operator);

$P_T Q_T = 0_{T\times T}$; $P_T e_T = e_T$; $Q_T e_T = 0_{T\times 1}$.

Example:

Let $y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}$. Then, $P_T y_i = \begin{pmatrix} \overline{y}_i \\ \overline{y}_i \\ \vdots \\ \overline{y}_i \end{pmatrix}$; $Q_T y_i = \begin{pmatrix} y_{i1} - \overline{y}_i \\ y_{i2} - \overline{y}_i \\ \vdots \\ y_{iT} - \overline{y}_i \end{pmatrix}$;

$P_T(e_T z_i) = e_T z_i$; $Q_T(e_T z_i) = 0_{T\times g}$, where $\overline{y}_i = \frac{1}{T}\Sigma_t y_{it}$.

(2) Fixed Effects Model:

1. Assumptions:

   a) The α_i are treated as parameters (1980, JEC, Kiefer)
      (i.e., different intercepts for individuals)

   b) The α_i are random variables which are correlated with all the regressors.

- Mundlak (1978, ECON): a) and b) are equivalent.

2. Within Estimation (Least Square Dummy Variables (LSDV))

- Treat α_i as parameters (individual intercepts).

$$y_i = X_i\beta + (e_T z_i)\gamma + e_T\alpha_i + \varepsilon_i = X_i\beta + e_T(z_i\gamma+\alpha_i) + \varepsilon_i \,. \qquad \textbf{(3)}$$

Observe:

$$Q_T y_i = Q_T X_i\beta + Q_T(e_T z_i\gamma+e_T\alpha_i) + Q_T\varepsilon_i = Q_T X_i\beta + Q_T\varepsilon_i \,. \qquad \textbf{(4)}$$

- Within estimator of β:

$$\hat{\beta}_W = \text{OLS on (4)} = (\Sigma_i X_i' Q_V X_i)^{-1}(\Sigma_i X_i' Q_V y_i)$$

$$= \text{OLS on (3) with dummy variables for individuals.}$$

Digression:

- Kronecker product:

  Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{p\times q}$. The two matrices do not have to have the same dimensions. Then,

$$A\otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \ldots & a_{1n}B \\ a_{21}B & a_{22}B & \ldots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{pmatrix}_{mp\times nq} .$$

- Facts: $(A\otimes B)(C\otimes D) = AC\otimes BD$

  $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$

- Notation:
  - $V = I_N\otimes e_T$ $(NT\times N)$ (Matrix of individual dummy variables)

$$V = \begin{pmatrix} e_T & 0_{T\times 1} & \ldots & e_T \\ 0_{T\times 1} & e_T & \ldots & 0_{T\times 1} \\ \vdots & \vdots & & \vdots \\ 0_{T\times 1} & 0_{T\times 1} & \ldots & e_T \end{pmatrix}.$$

  - $P_V = V(V'V)^{-1}V' = I_N\otimes P_T$; $Q_V = I_{NT} - P_V = I_N\otimes Q_T$.

- Observe:

$$y = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \cdots \\ \vdots \\ \cdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \rightarrow Q_V y = \begin{pmatrix} Q_T y_1 \\ Q_T y_2 \\ \vdots \\ Q_T y_N \end{pmatrix} = \begin{pmatrix} y_{11} - \overline{y}_1 \\ y_{12} - \overline{y}_2 \\ \vdots \\ y_{1T} - \overline{y}_1 \\ \cdots \\ \vdots \\ \cdots \\ y_{N1} - \overline{y}_N \\ y_{N2} - \overline{y}_N \\ \vdots \\ y_{NT} - \overline{y}_N \end{pmatrix}; \; P_V y = \begin{pmatrix} P_T y_1 \\ P_T y_2 \\ \vdots \\ P_T y_N \end{pmatrix} = \begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \\ \vdots \\ \overline{y}_1 \\ \cdots \\ \vdots \\ \cdots \\ \overline{y}_N \\ \overline{y}_N \\ \vdots \\ \overline{y}_N \end{pmatrix}$$

- $P_V Q_V = 0_{TN\times TN}$; $Q_V V = 0_{NT\times NT}$.

End of Digression

- Within estimator of β:

  $\hat{\beta}_W$ = {OLS on $Q_V y = Q_V X\beta + Q_V\varepsilon$} = $(X'Q_V X)^{-1}X'Q_V y$.

  $\rightarrow$ OLS on $y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)\beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$.

  - Properties of the within estimator:
  - unbiased.
  - consistent as either $T \rightarrow \infty$ or $N \rightarrow \infty$.
  - $$Cov(\hat{\beta}_W) = s^2\left(\Sigma_{i=1}^N \Sigma_{t=1}^T (x_{it} - \overline{x}_i)'(x_{it} - \overline{x}_i)\right)^{-1} = s^2\left(\Sigma_{i=1}^T X_i' Q_T X_i\right)^{-1} = s^2\left(X'Q_V X\right)^{-1},$$

  where s^2 = SSE from within estimation /{N(T-1)-k}.

  $\rightarrow$ s^2 is a consistent estimator of σ_ε^2.

- Notes on within estimation:
  - Individual effects are differenced away.
  - Can't estimate γ (coefficients of time-invariant regressors).
  - $\xi_i = \bar{y}_i - \bar{x}_i\hat{\beta}_W \to_p z_i\gamma + \alpha_i$, as $T \to \infty$.
  - $$R^2 = 1 - \frac{\Sigma_{i,t}(y_{it} - x_{it}\hat{\beta}_W - \hat{\xi}_i)^2}{\Sigma_{it}(y_{it} - \bar{y})^2} = 1 - \frac{\Sigma_{i,t}(y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)\hat{\beta}_W)^2}{\Sigma_{it}(y_{it} - \bar{y})^2}$$ $$= 1 - \frac{SSE \ \ from\ within\ regression}{SST}.$$

[VFR example from Stock and Watson, Ch. 8]

- Data: auto_1.txt
  - 48 U.S. states (excluding Alaska and Hawaii): N = 48.
  - 1982 -1988: T =7.
- We can think of Pal as α_i. Pal would be time-invariant.
- Within estimation results:

dependent variable: VFR

variable	coeff.	std. err.	t-st
beertax	**-0.4768**	**0.1657**	**-2.8773**
mlda	-0.0019	0.0178	-0.1053
jailed	0.0147	0.1201	0.1222
comserd	0.0345	0.1377	0.2503
unrate	-0.0629	0.0111	-5.6629
lpinc	1.7964	0.3625	4.9560
yr83	-0.0972	0.0322	-3.0232
yr84	-0.2812	0.0371	-7.5740
yr85	-0.3745	0.0389	-9.6220

yr86	-0.3376	0.0422	-8.0090
yr87	-0.4347	0.0481	-9.0369
yr88	-0.5213	0.0537	-9.7103

R-Square = 0.9390

- Now, the estimated coeff. on Beertax has the expected sign and is significant!

3. Other estimators:
   - OLS of y_{it} on x_{it} and z_i, that is, OLS on $y = X\beta + VZ\gamma + u$,
     $\rightarrow$ biased and inconsistent.
   - "Between" estimator of δ (β and γ) = OLS of $\bar{y}_i$ on $\bar{x}_i$ and z_i

     $= \text{OLS on } P_T y_i = P_T X_i\beta + e_T z_i\gamma + P_T u_i = P_T H_i\delta + P_T u$

     $= \text{OLS on } P_V y = P_V X\beta + VZ\gamma + P_V u = P_V H\delta + P_V u$

     $= (H'P_V H)^{-1}H'P_V y.$

     $\rightarrow$ Biased and inconsistent.

4. F test for H_o: $\alpha_1 = \alpha_2 = ... = \alpha_N$ ((N-1) restrictions).

STEP 1: Do OLS on $y_{it} = x_{it}\beta + z_i\gamma + \varepsilon_{it}$ and get SSE (RSSE: restricted SSE) (Make sure Z_i includes the overall intercept term.)

STEP 2: Do within estimation, i.e., OLS on

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i),$$

and get SSE (USSE: unrestricted SSE).

STEP3:

$$F = \frac{(RSSE - USSE)/(N-1)}{USSE/(NT-N-k)} \sim F(N-1, NT-N-k).$$

- Comment: This test is effective only if N is small.

5. Testing for poolability of panel data (with small N and large T):

- For each individual,

  $y_{it} = x_{it}\beta_i + \alpha_i^* + \varepsilon_{it}$; $\alpha_i^* = \alpha_i + z_i\gamma_i$.

  [Not only α^{i*} but also β_i may differ across i. In this cases, we cannot pool all the data. β_i and α_i^* should be estimated by OLS on each individual.

- Testing H_o: $\beta_1 = ... = \beta_N$ and $\alpha_1^* = ... = \alpha_N^*$ (assuming ε_{it} iid)

  STEP 1: Do OLS on each individual i and get SSE (SSE_i)

  Let USSE = $\Sigma_i SSE_i$.

  STEP 2: Do OLS on $y_{it} = x_{it}\beta + \alpha + \varepsilon_{it}$ using all obervations and get SSE (RSSE).

  STEP 3: F = $\frac{(RSSE - USSE)/((N-1)(k+1))}{USSE/(NT - N(k+1))}$.

6. Testing H_o: $\beta_1 = ... = \beta_N$ (assuming ε_{it} iid)

   STEP 1: Do OLS on each individual i and get SSE (SSE_i)

   Let USSE = $\Sigma_i SSE_i$

   STEP 2: Do OLS on $(y_{it} - \overline{y}_i) = (x_{it} - \overline{x}_i)\beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$ using all observations and get SSE (RSSE).

   STEP 3: F = $\frac{(RSSE - USSE)/((N-1)k)}{USSE/(NT - N(k+1))}$.

- For tests under general assumptions on ε_{it}, see Baltagi, Ch. 4.

(3) Random Effects Model [Balestra-Nerlove (ECON, 1966)]

1. Assumptions:

- α_i i.i.d. $N(0,\sigma_\alpha^2)$.
- α_i uncorrelated with regressors.

2. Possible estimation methods:

- OLS of y_{it} on x_{it} and z_i: consistent as $N \to \infty$.
- Between: unbiased and consistent as $N \to \infty$.
- Within: unbiased and consistent as $N \to \infty$ or $T \to \infty$.
- All of these are inefficient.

3. GLS estimator (efficient)

- The model:

$$y_i = H_i\delta + u_i \; ; \; u_i = e_T\alpha_i + \varepsilon_i \, ,$$

where $Cov(\varepsilon_i) = \sigma_\varepsilon^2 I_T$, $Cov(e_T\alpha_i) = T\sigma_\alpha^2 P_T$, $cov(\alpha_i,\varepsilon_{it}) = 0$.

$$y = H\delta + u = X\beta + VZ\gamma + u; \; u = V\alpha + \varepsilon$$

where,

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix}; VZ = \begin{pmatrix} e_T z_1 \\ e_T z_2 \\ \vdots \\ e_T z_N \end{pmatrix}; \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix}; V\alpha = \begin{pmatrix} e_T\alpha_1 \\ e_T\alpha_2 \\ \vdots \\ e_T\alpha_N \end{pmatrix};$$

$Cov(\varepsilon) = \sigma_\varepsilon^2 I_{NT}$ and $Cov(V\alpha) = \sigma_\alpha^2 P_V$.

- Assume ε and α are uncorrelated:

$$\text{Cov}(u_i) = \text{Cov}(e_T\alpha_i+\varepsilon_i) = \text{Cov}(e_T\alpha_i) + \text{Cov}(\varepsilon_i)$$
$$= e_T v(\alpha_i)e_T' + \text{Cov}(\varepsilon_i) = \sigma_\alpha^2 e_T e_T' + \sigma_\varepsilon^2 I_T$$
$$= \sigma_\alpha^2 e_T e_T' + \sigma_\varepsilon^2 I_T = T\sigma_\alpha^2 e_T(e_T'e_T)^{-1}e_T' + \sigma_\varepsilon^2 I_T$$
$$= T\sigma_\alpha^2 P_T + \sigma_\varepsilon^2 I_T = \sigma_\varepsilon^2[(T\sigma_\alpha^2/\sigma_\varepsilon^2)P_T+I_T]$$
$$= \sigma_\varepsilon^2[\{(T\sigma_\alpha^2+\sigma_\varepsilon^2)/\sigma_\varepsilon^2\}P_T + Q_T]$$
$$\equiv \sigma_\varepsilon^2(\theta^{-2}P_T + Q_T) \equiv \sigma_\varepsilon^2\Sigma .$$
$$\text{Cov}(u) = \sigma_\varepsilon^2\Omega,\ \Omega = \theta^{-2}P_V + Q_V.$$

- $\Sigma \neq I_T$ unless $\theta = 1$ (that is, $\sigma_\alpha^2 = 0$).
- $\Sigma^{-1} = \theta^2 P_T + Q_T \rightarrow \Sigma^{-1/2} = \theta P_T + Q_T$.

  $\Omega^{-1} = \theta^2 P_V + Q_V \rightarrow \Omega^{-1/2} = \theta P_V + Q_V$.

- Whitening the error in the model:

$$\Sigma^{-1/2}y_i = \Sigma^{-1/2}H_i\delta + \Sigma^{-1/2}u_i \qquad \textbf{(5)}$$
$$\rightarrow \text{Cov}(\Sigma^{-1/2}u_i) = \sigma_\varepsilon^2 I_T.$$
$$\Omega^{-1/2}y = \Omega^{-1/2}H\delta + \Omega^{-1/2}u$$
$$\rightarrow \text{Cov}(\Omega^{-1/2}u) = \sigma_\varepsilon^2 I_{NT}.$$

- GLS estimator of δ:

$$\hat{\delta}_{GLS} = \text{OLS on (5)} = (\Sigma_i H_i'\Sigma^{-1}H_i)^{-1}\Sigma_i H_i'\Sigma^{-1}y_i$$

$$= (H'\Omega^{-1}H)^{-1}H'\Omega^{-1}y,$$

where $\Omega = I_N \otimes \Sigma$.

4. A practical way to obtain GLS

$$y_{it}^* = y_{it} - (1-\theta)\overline{y}_i ; x_{it}^* = x_{it} - (1-\theta)\overline{x}_i ; z_i^* = z_i - (1-\theta)z_i = \theta z_i .$$

(quasi-differenced data)

$y_i^* = \Sigma^{-1/2}y_i = y_i-(1-\theta)P_Ty_i$; $X_i^* = \Sigma^{-1/2}X_i = X_i-(1-\theta)P_TX_i$;

$e_Tz_i^* = \Sigma^{-1/2}e_Tz_i = \theta e_Tz_i$.

$y^* = \Omega^{-1/2}y = y-(1-\theta)P_Vy$; $X^* = \Omega^{-1/2}X = X-(1-\theta)P_VX$;

$\Omega^{-1/2}VZ = \theta VZ$.

- GLS estimator of $\delta = (\beta',\gamma')'$

= OLS on $y_{it}^* = x_{it}^*\beta + z_i^*\gamma$ + error

= OLS on $y_i^* = X_i^*\beta + e_Tz_i^*\gamma$ + error

= OLS on $\Omega^{-1/2}y = \Omega^{-1/2}H\delta$ + error.

5. Estimation of θ

- Between on $P_T y_i = P_T X_i \beta + e_T z_i \gamma$ + error (for NT observation) {= OLS on $P_V y = P_V X\beta + VZ\gamma$ + error} and get the residual vector v.
- $s_B^2 = SSE_B/(N\text{-}k\text{-}g) = v'v/(N\text{-}k\text{-}g) = \Sigma_i \Sigma_t v_{it}^2/(N\text{-}k\text{-}g)$.

  [$\text{plim}\ s_B^2 = T\sigma_\alpha^2 + \sigma_\varepsilon^2$, as $N \to \infty$.]
- $\hat{\theta} = [s^2/s_B^2]^{1/2}$ [$\text{plim}_{N\to\infty}\ \hat{\theta} = \theta$].

6. R^2:

- $\hat{\xi}_{i,GLS} = \bar{y}_i - \bar{x}_{it}\hat{\beta}_{GLS} - z_i\hat{\gamma}_{GLS}$.
- $$R^2 = 1 - \frac{\Sigma_{i,t}(y_{it} - x_{it}\hat{\beta}_{GLS} - z_i\hat{\gamma}_{GLS} - \hat{\xi}_{i,GLS})'(y_{it} - x_{it}\hat{\beta}_{GLS} - z_i\hat{\gamma}_{GLS} - \hat{\xi}_{i,GLS})}{\Sigma_{it}(y_{it} - \bar{y})^2}$$
  $$= 1 - \frac{\Sigma_{i,t}(y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)\hat{\beta}_{GLS})'(y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)\hat{\beta}_{GLS})}{\Sigma_{it}(y_{it} - \bar{y})^2}.$$

7. Statistical properties of GLS:

- Consistent if N is large.
- More efficient than Within.
- GLS of $\beta \approx$ Within of β, if T is large.

  $\to \theta = [\sigma_\varepsilon^2/(T\sigma_\alpha^2 + \sigma_\varepsilon^2)]^{1/2} \to 0$ as $T \to \infty$.

  $y_i^* = y_i - (1-\theta)P_T y_i \to y_{it} - \bar{y}_i$,

  $X_i^* = X_i - (1-\theta)P_T X_i \to x_{it} - \bar{x}_i$,

  $z_i^* = \theta z_i \to 0$ (but can estimate γ even if T is large).

8. Covariance matrix of GLS:

   $\delta = (\beta',\gamma')'$ and $h_{it}^* = [x_{it}^*, z_i^*]$.

   $\rightarrow \text{Cov}(\hat{\delta}_{GLS}) = s^2\,[\Sigma_i\Sigma_t h_{it}^{*\prime} h_{it}^*]^{-1} = s^2[\Sigma_i H_i^{*\prime} H_i^*]^{-1}$

   $= s^2[H'\Omega^{-1}H]^{-1}$.

9. GLS is the optimally weighted average of within and between estimators. [See Judge et al (book, 1984, ch. 13) or Maddala (1971, ECON), Baltagi, p. 16.]

10. Testing the existence of individual effects:

- H_o: $\sigma_\alpha^2 = 0$.
- Let e_{it} be the residual from OLS on $y_{it} = x_{it}\beta + z_i\gamma + \varepsilon_{it}$. Breusch and Pagan (1980, RESTUD) derive LM test:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\Sigma_i(\Sigma_t e_{it})^2}{\Sigma_i\Sigma_t e_{it}^{\;2}} - 1\right]^2$$

11. Hausman test of GLS V.S. Within estimator:

- GLS is efficient if the α_i are random and uncorrelated with regressors. If these assumptions are violated, GLS is inconsistent.
- $\hat{\beta}_W$ is consistent whether the α_i are random or fixed effects.
- How to test consistency of GLS:
  - Let $\hat{\beta}_{GLS}$ be the GLS estimator of β.
  - $H_T = (\hat{\beta}_W - \hat{\beta}_{GLS})'[Cov(\hat{\beta}_W) - Cov(\hat{\beta}_{GLS})]^+(\hat{\beta}_W - \hat{\beta}_{GLS})$

    $\to \chi^2$, df = rank[Cov($\hat{\beta}_W$)-Cov($\hat{\beta}_{GLS}$)].

- Alternatives of H_T:

  [Ahn and Low (1996, Journal of Econometrics)].

  - s there any link between J-test and Hausman test?
  - H_T = J-statistic for testing

    $E(X_i'Q_Tu_i) = 0$ ($\to E[X_i'Q_T(e_T\alpha_i+\varepsilon_i)] = E(X_i'Q_T\varepsilon_i) = 0$),

    $E(z_i'e_T'u_i) = 0$ ($\to E(z_i'\overline{u}_i) = 0$),

    $E(X_i'P_Tu_i) = 0$ ($\to E(\overline{x}_i'\overline{u}_i) = 0$).

- AL test procedure:
  - OLS on $\Omega^{-1/2}y = \Omega^{-1/2}H\delta$ + error,

    get v^*(the residual vector from this).
  - Regress v^* on Q_VX, P_VX and VZ, and get R^2 and f_{it} = the vector of fitted value.

    $\rightarrow AL_{T1} = NT\times R^2 \rightarrow H_T$, if z_i includes ONE.

    $\rightarrow AL_{T2} = f'f/s^2 = H_T$, numerically.

12. MLE of the Random Effects Model

- See Baltagi, pp. 18-19.
- GLS of β and $\gamma \approx$ MLE of β and γ.
- MLE behaves well in finite samples.

13. Two-Way Error Component Models:

- The model:

$$y_{it} = x_{it}\beta + z_i\gamma + u_{it};\ u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}\ , \qquad (6)$$

where λ_t = time-specific effects (e.g., macro shock).

- Cases where both α_i and λ_t are fixed:
  - Using $\tilde{Q}_V = I_{NT} - (I_N\otimes P_T) - (P_N\otimes I_T) + (P_N\otimes P_T)$,

$$\tilde{Q}_V y = \tilde{Q}_V X\beta + \tilde{Q}_V \varepsilon .$$

  - $\hat{\beta}_{TW,W} = (X'\tilde{Q}_V X)^{-1} X'\tilde{Q}_V y$.

- Equivalent procedure I:
  - Define: $\overline{y}_t = \frac{1}{N}\Sigma_i y_{it}$; $\overline{x}_t = \frac{1}{N}\Sigma_i x_{it}$;

    $\overline{y} = \frac{1}{NT}\Sigma_i\Sigma_t y_{it}$; $\overline{x} = \frac{1}{NT}\Sigma_i\Sigma_t x_{it}$.
  - Within = OLS on:

    $$(y_{it} - \overline{y}_t - \overline{y}_i + \overline{y}) = (x_{it} - \overline{x}_t - \overline{x}_i + \overline{x})\beta + (\varepsilon_{it} - \overline{\varepsilon}_t - \overline{\varepsilon}_i + \overline{\varepsilon}).$$

- Equivalent procedure II.
  - $\hat{\beta}_{TW,W} = \hat{\beta}_W$ on the model with time-dummy variables.

- Cases where the λ_t are random.

  Do GLS. (See Baltagi.)

- Can we treat λ_t as random?
  - If T is large. May assume λ_t (macro shocks) iid $N(0,\sigma_\lambda^2)$.
  - If T is small, treat λ_t as fixed.

[3] Heteroskedasticity and Autocorrelation

(1) The case of time-series heteroskedasticity/autocorrelation, but not cross-sectional heteroskedasticity:

- $Cov(u_i) = \Sigma \neq \sigma_\varepsilon^2(\theta^{-2}P_T + Q_T)$, for all i.
- The ε_{it} could be autocorrelated and heteroskedastic over time.

- GLS for random effects models
  - Let $\hat{u}_i = y_i - H_i\hat{\delta}$, where $\hat{\delta}$ is the OLS estimator.
  - Compute $\hat{\Sigma} = \frac{1}{N}\Sigma_i \hat{u}_i \hat{u}_i'$.
  - $\hat{\delta}_{GLS} = (\Sigma_i H_i'\hat{\Sigma}^{-1}H_i)^{-1}\Sigma_i H_i'\hat{\Sigma}^{-1}y_i$

- GLS for fixed effects models (Kiefer, 1981)
  - $Q_T y_i = Q_T X_i \beta + Q_T \varepsilon_i$. $[Q_V y = Q_V X\beta + Q_V\varepsilon]$
  - $\Phi = Cov(Q_T\varepsilon_i)$. Since Q_T is singular, so is Φ.

    → Use Φ^+ (Moor-Penrose generalized Inverse)

    $[\Lambda = Cov(Q_V\varepsilon) = I_N \otimes \Phi.]$
  - $$\hat{\beta}_{W,GLS} = \left(\Sigma_i X_i' Q_T \Phi^+ Q_T X_i\right)^{-1}\Sigma_i X_i' Q_T \Phi^+ Q_T y_i$$
    $$= (X'Q_V\Lambda^+ Q_V X)^{-1}X'Q_V\Lambda^+Q_V y.$$
  - Estimate Φ by $\hat{\Phi} = \frac{1}{N}\Sigma_i(Q_T y_i - Q_T X_i\hat{\beta}_W)'(Q_T y_i - Q_T X_i\hat{\beta}_W)$.

(2) The case of time-series heteroskedasticity/autocorrelation, and cross-sectional heteroskedasticity:

- $Cov(u_i) = \Sigma_i \neq \Sigma_j = Cov(u_j)$, for all $i \neq j$.
- The ε_{it} could be autocorrelated over time, and heteroskedastic over time and individuals.

- For the case of fixed effects:
  - Use $\hat{\beta}_W$.
  - $Cov(\hat{\beta}_W) = \left(\Sigma_i X_i{}' Q_T X_i\right)^{-1} (\Sigma X_i{}' Q_T \hat{\Phi}_i Q_T X_i) \left(\Sigma_i X_i{}' Q_T X_i\right)^{-1}$,

    where $\hat{\Phi}_i = (Q_T y_i - Q_T X_i \hat{\beta}_W)'(Q_T y_i - Q_T X_i \hat{\beta}_W)$.

- For the case of random effects:
  - Use $\hat{\delta}_{OLS}$.
  - $Cov(\hat{\delta}_{OLS}) = (\Sigma_i H_i{}' H_i)^{-1} (\Sigma_i H_i{}' \hat{u}_i \hat{u}_i{}' H_i)^{-1} (\Sigma_i H_i{}' H_i)^{-1}$,

    where $\hat{u}_i = (y_i - H_i \hat{\delta}_{OLS})'(y_i - H_i \hat{\delta}_{OLS})$.

[Our alcohol example again]

Within Estimation Results (HETERO/AUTO ADJUSTED)

variable	coeff.	std. err.	t-st
beertax	-0.4768	0.2949	-1.6167
mlda	-0.0019	0.0209	-0.0894
jailed	0.0147	0.0158	0.9292
comserd	0.0345	0.1285	0.2684
unrate	-0.0629	0.0127	-4.9657
lpinc	1.7964	0.6243	2.8775

Kiefer's Within Estimation Results

variable	coeff.	std. err.	t-st
beertax	-0.1833	0.1837	-0.9982
mlda	-0.0058	0.0175	-0.3343
jailed	-0.0026	0.0882	-0.0295
comserd	0.0420	0.1110	0.3784
unrate	-0.0518	0.0112	-4.6255
lpinc	2.0991	0.3932	5.3387

Kiefer's Within Estimation Results (HETERO ADJUSTED)

variable	coeff.	std. err.	t-st
beertax	-0.1833	0.2254	-0.8135
mlda	-0.0058	0.0157	-0.3720
jailed	-0.0026	0.0138	-0.1885
comserd	0.0420	0.0924	0.4545
unrate	-0.0518	0.0107	-4.8367
lpinc	2.0991	0.5934	3.5373