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 heterogeneity.
  • 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.
Is Controlling Unobservables Important?

[Example from Stock and Watson, Ch. 10]

- Issue:
  - Do alcohol taxes help decrease traffic deaths?
- Data: auto_1.txt

<table>
<thead>
<tr>
<th>Series</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>State ID (FIPS) Code</td>
</tr>
<tr>
<td>year</td>
<td>Year</td>
</tr>
<tr>
<td>spircons</td>
<td>Spirits Consumption</td>
</tr>
<tr>
<td>unrate</td>
<td>Unemployment Rate</td>
</tr>
<tr>
<td>perinc</td>
<td>Per Capita Personal Income</td>
</tr>
<tr>
<td>emppop</td>
<td>Employment/Population Ratio</td>
</tr>
<tr>
<td>beertax</td>
<td>Tax on Case of Beer</td>
</tr>
<tr>
<td>mlda</td>
<td>Minimum Legal Drinking Age</td>
</tr>
<tr>
<td>vmiles</td>
<td>Ave. Mile per Driver</td>
</tr>
<tr>
<td>jaild</td>
<td>Mandatory Jail Sentence</td>
</tr>
<tr>
<td>comserd</td>
<td>Mandatory Community Service</td>
</tr>
<tr>
<td>allmort</td>
<td># of Vehicle Fatalities (#VF)</td>
</tr>
<tr>
<td>mrall</td>
<td>Vehicle Fatality Rate (VFR)</td>
</tr>
<tr>
<td></td>
<td>(traffic deaths per 10,000 people)</td>
</tr>
</tbody>
</table>

Basic Panel-4
• OLS results:

dependent variable: VFR

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

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 \( \beta_2 \) and \( \beta_3 \) measure?

\( \beta_2 \) measures the direct (pure) effect of \( x_{2i} \) on \( y_i \) with \( x_{3i} \) held constant.

Similarly, \( \beta_3 \) measures the direct effect of \( X_{3i} \) on \( Y_i \) with \( X_{2i} \) held constant.
• If you estimate $y_i = \alpha_1 + \alpha_2 x_{2i} + \text{error}$ instead?

• Let $x_{3i} = \delta_1 + \delta_2 x_{2i} + v_i$. Substitute it into the correct model:

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

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

• Thus, $\alpha_2 = \beta_2 + \delta_2 \beta_3$.

• Direct v.s. Indirect Effects

<table>
<thead>
<tr>
<th>$x_{2i}$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2$</td>
<td>$y_i$</td>
</tr>
<tr>
<td>$x_{3i}$</td>
<td>$\beta_3$</td>
</tr>
</tbody>
</table>

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 a different level of preference for alcohol (say, Pal).
  • Pal and Beertax could be positively related ($\delta_2 > 0$).
  • Pal would have a positive direct 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?
Fixed effects vs. Random effects

(1) Basic Model:

\[ y_{it} = x_{it}'\beta + z_{it}'\gamma + u_{it} = h_{it}'\delta + u_{it} ; \ u_{it} = \alpha_i + \varepsilon_{it}, \]  

where \( i = 1, \ldots, N \) (cross-section unit), \( t = 1, \ldots, T \) (time),

\[ h_{it} = \begin{pmatrix} x_{it} \\ z_i \end{pmatrix} , \ \delta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} . \]

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

- Matrix Notation:
  - Define:

\[
y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} ; \ X_i = \begin{pmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{pmatrix} ; \ u_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix} ; \ e_i = \begin{pmatrix} e_{i1} \\ \vdots \\ e_{iT} \end{pmatrix} .
\]
With this notation, we can write (1) by:

\[ y_i = X_i \beta + e_i \gamma' + u_i = H_i \delta + u_i; u_i = \alpha_i e_T + \epsilon_i, \]  

where \( H_i = [X_i, e_T z_i'] \) and \( e_T \) is \( T \times 1 \) vector of ones.

**Notation:**

- For \( T \times p \) matrix \( M_i \) or \( T \times 1 \) 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 \( N_T \) rows.

- With this notation, model (2) can be rewritten for all observations as \( y = H \delta + u \).

**Other matrix notation:**

\( e_T = (1 \ 1 \ldots \ 1)' \) \((T \times 1 \) vector of ones); \n
\( P_T = e_T (e_T e_T)' e_T T^{-1} e_T \)

\[
= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ T & T & \cdots & T \\ 1 & 1 & \cdots & 1 \\ T & T & \cdots & T \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \\ T & T & \cdots & T \\ \end{pmatrix} \quad (T \times T \text{ mean operator});
\]
\[ Q_T = I_T - P_T = \begin{pmatrix} \frac{T-1}{T} & -1 & \cdots & -1 \\ \frac{T}{T} & \frac{T}{T} & \cdots & \frac{T}{T} \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & \frac{T-1}{T} \end{pmatrix} \]

\((T \times T \text{ deviation-from-mean operator})\);

\[ P_T Q_T = 0_{T \times T}; \quad P_T e_T = e_T; \quad Q_T e_T = 0_{T \times 1}. \]

Example:

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

\[ P_T (e_T z_i) = e_T z_i; \quad Q_T (e_T z_i) = 0_{T \times g}, \text{ where } \bar{y}_i = \frac{1}{T} \sum_i y_{it}. \]
(2) Fixed Effects Model:

1. Assumptions:
   a) The $\alpha_i$ are treated as parameters (1980, JEC, Kiefer)
      (i.e., different intercepts for individuals)
   b) The $\alpha_i$ 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))

\[
y_i = X_i \beta + (e_T z_i') \gamma + e_T \alpha_i + \varepsilon_i = X_i \beta + e_T (z_i' \gamma + \alpha_i) + \varepsilon_i. \tag{3}
\]

Observe:

\[
Q_T y_i = Q_T X_i \beta + Q_T e_T (z_i' \gamma + \alpha_i) + Q_T \varepsilon_i = Q_T X_i \beta + Q_T \varepsilon_i. \tag{4}
\]

- Within estimator of $\beta$:

\[
\hat{\beta}_w = \text{OLS on (4)} = \left( \Sigma_i X_i' Q_T X_i \right)^{-1} \Sigma_i X_i' Q_T y_i \\
= \text{OLS on (3) with dummy variables for individuals.}
\]
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,

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

- 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)
    \[
    V = \begin{pmatrix}
      e_T & 0_{T \times 1} & \ldots & 0_{T \times 1} \\
      0_{T \times 1} & e_T & \ldots & 0_{T \times 1} \\
      \vdots & \vdots & \ddots & \vdots \\
      0_{T \times 1} & 0_{T \times 1} & \ldots & e_T
    \end{pmatrix}.
    \]
  - \( P_V = V(V'V)^{-1}V' = I_N \otimes P_T \); \( Q_V = I_{NT} - P_V = I_N \otimes Q_T \).
• Observe:

\[ y = \begin{pmatrix} y_11 \\ y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} \rightarrow \quad Q_T y = \begin{pmatrix} Q_T y_1 \\ Q_T y_2 \\ \vdots \\ Q_T y_N \end{pmatrix} = \begin{pmatrix} y_{11} - \bar{y}_1 \\ y_{12} - \bar{y}_2 \\ \vdots \\ y_{1T} - \bar{y}_1 \\ \vdots \\ y_{N1} - \bar{y}_N \\ y_{N2} - \bar{y}_N \\ \vdots \\ y_{NT} - \bar{y}_N \end{pmatrix}; \]

\[ P_T y = \begin{pmatrix} P_T y_1 \\ P_T y_2 \\ \vdots \\ P_T y_N \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix} \]

• \( P_V Q_V = 0_{TN\times TN}; \quad Q_V V = 0_{NT\times NT}. \)

End of Digression
• Within estimator of $\beta$:
\[
\hat{\beta}_w = \{\text{OLS on } Q'\v y = Q'X\beta + Q'\v \v = (X'Q'X)^{-1}X'Q'v\v y.}
\]
\[
\to \text{OLS on } y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (\v_{it} - \v_i).
\]

• Properties of the within estimator:
  • unbiased.
  • consistent as either $T \to \infty$ or $N \to \infty$.
\[
\text{Cov}(\hat{\beta}_W) = s^2 \left( \Sigma_{i=1}^N \Sigma_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i) \right)^{-1}
\]
\[
= s^2 \left( \Sigma_{i=1}^N X'_iQ_TX_i \right)^{-1} = s^2 \left( X'Q'_YX \right)^{-1},
\]
where $s^2 = \text{SSE from within estimation / \{N(T-1)-k\}}$.
\[
\to s^2 \text{ is a consistent estimator of } \sigma^2_\v.
\]

• Why is $s^2$ consistent?
  • To make sense of this, let $q_i = Q_Tu_i = Q_T\v_i$. Then we can observe:
\[
E \left( \Sigma_{i=1}^T q_i^2 \right) = E(q'_iq_i) = E(\v'_iq_i) = E(tr(Q_T\v_i\v_i'))
\]
\[
= \sigma^2_\v tr(Q_T) = (T-1)\sigma^2_\v.
\]
Then, by the central limit theorem,
\[
\frac{q'q}{N(T-1)} = \frac{1}{N(T-1)} \Sigma_{i=1}^N q'_iq_i
\]
\[
\to_p \lim \frac{1}{N(T-1)} E(\Sigma_i q'_iq_i) = \sigma^2_\v.
\]
Notes on within estimation:

- Individual effects are differenced away.
- Can't estimate $\gamma$ (coefficients of time-invariant regressors).
- $\hat{\xi}_i = \bar{y}_i - \bar{x}_i' \hat{\beta}_W \to p z_i' \gamma + \alpha_i$, as $T \to \infty$.

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

$$= 1 - \frac{\text{SSE from within regression}}{\text{SST}}.$$
• Data: auto_1.txt
• We can think of Pal as $\alpha_i$. 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 $\delta$ ($\beta$ and $\gamma$) = 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.
Random Effects Model [Balestra-Nerlove (ECON, 1966)]

1. Assumptions:
   - $\alpha_i$ i.i.d. $N(0, \sigma^2_\alpha)$.
   - $\alpha_i$ 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:
     \[ y_i = H_i\delta + u_i; u_i = e_T\alpha_i + \epsilon_i, \]
   where $\text{Cov}(\epsilon_i) = \sigma^2_\epsilon I_T$, $\text{Cov}(e_T\alpha_i) = T\sigma^2_\alpha P_T$, $\text{cov}(\alpha_i, \epsilon_{it}) = 0$.

   \[ y = H\delta + u = X\beta + VZ\gamma + u; u = V\alpha + \epsilon \]

   where,
   \[ Z = \begin{pmatrix} z_1' \\ z_2' \\ \vdots \\ z_N' \end{pmatrix}; VZ = \begin{pmatrix} e_Tz_1' \\ e_Tz_2' \\ \vdots \\ e_Tz_N' \end{pmatrix}; \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix}; V\alpha = \begin{pmatrix} e_T\alpha_1 \\ e_T\alpha_2 \\ \vdots \\ e_T\alpha_N \end{pmatrix}; \]

   $\text{Cov}(\epsilon) = \sigma^2_\epsilon I_{NT}$ and $\text{Cov}(V\alpha) = T\sigma^2_\alpha P_V$. 

Basic Panel-18
• Assume $\varepsilon$ and $\alpha$ are uncorrelated:

\[
\text{Cov}(u_i) = \text{Cov}(e_T\alpha_i + \varepsilon_i) = \text{Cov}(e_T\alpha_i) + \text{Cov}(\varepsilon_i)
\]

\[
= e_T v(\alpha_i)e_T' + \text{Cov}(\varepsilon_i) = \sigma^2_\alpha e_T e_T' + \sigma^2_\varepsilon I_T
\]

\[
= \sigma^2_\alpha e_T e_T' + \sigma^2_\varepsilon I_T = T\sigma^2_\alpha e_T(e_T' e_T)^{-1} e_T' + \sigma^2_\varepsilon I_T
\]

\[
= T\sigma^2_\alpha P_T + \sigma^2_\varepsilon I_T = \sigma^2_\varepsilon [(T\sigma^2_\alpha/\sigma^2_\varepsilon)P_T + I_T]
\]

\[
= \sigma^2_\varepsilon [\{(T\sigma^2_\alpha + \sigma^2_\varepsilon)/\sigma^2_\varepsilon\}P_T + Q_T]
\]

\[
= \sigma^2_\varepsilon (\theta^2 P_T + Q_T) \equiv \sigma^2_\varepsilon \Sigma.
\]

\[
\text{Cov}(u) = \sigma^2_\varepsilon \Omega, \; \Omega = \theta^2 P_V + Q_V.
\]

• $\Sigma \neq I_T$ unless $\theta = 1$ (that is, $\sigma^2_\alpha = 0$).

• $\Sigma^{-1} = \theta^2 P_T + Q_T \rightarrow \Sigma^{-1/2} = \theta P_T + Q_T$.

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

• Whitening the error in the model:

\[
\Sigma^{-1/2} y_i = \Sigma^{-1/2} H_i \delta + \Sigma^{-1/2} u_i
\]

\[
\rightarrow \text{Cov}(\Sigma^{-1/2} u_i) = \sigma^2_\varepsilon I_T.
\]

\[
\Omega^{-1/2} y = \Omega^{-1/2} H \delta + \Omega^{-1/2} u
\]

\[
\rightarrow \text{Cov}(\Omega^{-1/2} u) = \sigma^2_\varepsilon I_{NT}.
\]
• GLS estimator of $\delta$:

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

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

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

4. A practical way to obtain GLS

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

(quasi-differenced data)

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

$$e_T z_i^* = \Sigma^{-1/2} e_T z_i' = \theta e_T z_i'.$$

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

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

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

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

$$= \text{OLS on } y_i^* = X_i \beta + e_T z_i' \gamma + \text{error}$$

$$= \text{OLS on } \Omega^{-1/2} y = \Omega^{-1/2} H \delta + \text{error}.$$
5. Estimation of θ

- Between on \( P_T Y_i = P_T X_i \beta + e_T Z_i' \gamma + \text{error} \) (for NT observations) \{= OLS on \( P_V Y = P_V X \beta + VZ \gamma + \text{error} \}\} and get the residual vector \( \hat{\nu} \).
  
  \[ s_B^2 = \frac{\text{SSE}_B}{(N-k-g)} = \frac{\hat{\nu}^T \hat{\nu}}{(N-k-g)} = \frac{\Sigma_i \Sigma_i \hat{v}_{it}^2}{(N-k-g)}. \]
  
  \[ \text{plim} s_B^2 = T\sigma^2_\alpha + \sigma^2_\varepsilon, \text{ as } N \rightarrow \infty. \]

- To make sense of these, let \( \nu_i = P_T u_i = (\alpha_i + \bar{\varepsilon}_i) e_T \). Then we can observe:
  
  \[ E\left( \Sigma_{t=1}^T v_{it}^2 \right) = E(\nu_i' \nu_i) = E(e_T' e_T (\alpha_i + \bar{\varepsilon}_i)^2) = T\sigma^2_\alpha + \sigma^2_\varepsilon. \]

Then, by the central limit theorem,

\[
\frac{\nu_i' \nu_i}{N} = \frac{1}{N} \sum_{i=1}^{N} \nu_i' \nu_i \rightarrow_p \lim_{N} \frac{1}{N} E(\nu_i' \nu_i) = T\sigma^2_\alpha + \sigma^2_\varepsilon.
\]

- \( \hat{\theta} = \sqrt{s^2 / s_B^2} \) [plim\( N \rightarrow \infty \) \( \hat{\theta} = 0 \)].

- Alternatively, OLS on \( y = X \beta + VZ + \text{error} \) and get \( \hat{u} \).
  
  \[ s_{OLS}^2 = \frac{\text{SSE}_{OLS}}{(NT - k - g)}. \]
  
  \[ \text{plim} s_{OLS}^2 = T\sigma^2_\alpha + \sigma^2_\varepsilon. \]

- Why? Observe:
  
  \[ E\left( \Sigma_{t=1}^T u_{it}^2 \right) = E(u_i' u_i) = T(\sigma^2_\alpha + \sigma^2_\varepsilon). \]

Then, by the central limit theorem,

\[
\frac{u_i' u_i}{NT} = \frac{1}{NT} \sum_{i=1}^{N} u_i' u_i \rightarrow_p \lim_{NT} \frac{1}{NT} E(\Sigma u_i' u_i) = \sigma^2_\alpha + \sigma^2_\varepsilon.
\]

- \( \hat{\theta} = \sqrt{s^2 / [T(s_{OLS}^2 - s^2) + s^2]} \) [plim\( N \rightarrow \infty \) \( \hat{\theta} = 0 \)].

Basic Panel-21
6. R$^2$:

- $\hat{\xi}_{i,\text{GLS}} = \bar{y}_i - \bar{x}_i' \hat{\beta}_{\text{GLS}} - z_i' \hat{\gamma}_{\text{GLS}}$.

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

7. Statistical properties of GLS (or feasible GLS):

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

$$\theta = \left[\frac{\sigma^2_\epsilon}{(T\sigma^2_\alpha + \sigma^2_\epsilon)}\right]^{1/2} \rightarrow 0 \text{ as } T \rightarrow \infty.$$

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

$$X_i^* = X_i - (1-\theta)P_T X_i \rightarrow (x_{it} - \bar{x}_i)' .$$

$$z_i^* = \theta z_i \rightarrow 0 \text{ (but can estimate } \gamma \text{ even if } T \text{ is large).}$$

8. Covariance matrix of (feasible) GLS:

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

$$\rightarrow \text{Cov}(\hat{\delta}_{\text{GLS}}) = s^2 \left( \sum_{i} \sum_{t} h_{it}^* h_{it}^{**} \right)^{-1} = \sigma^2_\epsilon \left( \sum_{i} H_{i}'' H_{i}^* \right)^{-1} = \sigma^2_\epsilon [H'\Omega^{-1}H]^{-1}.$$
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:
- \( H_0: \sigma_{\alpha}^2 = 0. \)
- 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:
  \[
  LM = \frac{NT}{2(T-1)} \left[ \frac{\sum_i (\sum_t e_{it})^2}{\sum_i \sum_t e_{it}^2} - 1 \right]^2
  \]

11. Hausman test of GLS V.S. Within estimator:
- GLS is efficient if the \( \alpha_i \) are random and uncorrelated with regressors. If these assumptions are violated, GLS is inconsistent.
- \( \hat{\beta}_W \) is consistent whether the \( \alpha_i \) are random or fixed effects.
- How to test consistency of GLS:
  - Let \( \hat{\beta}_{GLS} \) be the GLS estimator of \( \beta \).
  - \( H_T = (\hat{\beta}_W - \hat{\beta}_{GLS})' [\text{Cov}(\hat{\beta}_W) - \text{Cov}(\hat{\beta}_{GLS})] (\hat{\beta}_W - \hat{\beta}_{GLS}) \)
    \[ \rightarrow \chi^2, \text{df} = \text{rank}[\text{Cov}(\hat{\beta}_W) - \text{Cov}(\hat{\beta}_{GLS})]. \]
• Alternatives of $H_T$:


• AL test procedure:
  
  • OLS on $\Omega^{-1/2}y = \Omega^{-1/2}H\delta + \text{error}$,
    
    get $v^*$ (the residual vector from this).
  
  • Regress $v^*$ on $Q_vX$, $P_vX$ and $VZ$, and get $R^2$ and $f_{it} = \text{the vector of fitted value}$.

  \[ \rightarrow AL_{T1} = NT \times R^2 \rightarrow H_T, \text{ if } z_i \text{ includes ONE.} \]
  \[ \rightarrow AL_{T2} = f'f/s^2 = H_T, \text{ numerically.} \]

12. MLE of the Random Effects Model

• See Baltagi, pp. 18-19.

• GLS of $\beta$ and $\gamma \approx \text{MLE of } \beta \text{ and } \gamma$.

• MLE behaves well in finite samples.
13. Two-Way Error Component Models:

- The model:
  \[ y_{it} = x_{it}'\beta + z_{i}'\gamma + u_{it}; \quad u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}, \]  
  where \( \lambda_t \) = time-specific effects (e.g., macro shock).

- Cases where both \( \alpha_i \) and \( \lambda_t \) 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: \( \bar{y}_t = \frac{1}{N} \sum_i y_{it}; \quad \bar{x}_i = \frac{1}{N} \sum_i x_{it}; \)
      \[ \bar{y} = \frac{1}{NT} \sum_i \sum_t y_{it}; \quad \bar{x} = \frac{1}{NT} \sum_i \sum_t x_{it}. \]
    - Within = OLS on:
      \[ (y_{it} - \bar{y}_t - \bar{y}_i + \bar{y}) = (x_{it} - \bar{x}_i - \bar{x}_i + \bar{x})'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_i + \bar{\varepsilon}). \]
• Equivalent procedure II.
  • \( \hat{\beta}_{tw,w} = \hat{\beta}_w \) on the model with time-dummy variables.

• Cases where the \( \lambda_t \) are random.
  Do GLS. (See Baltagi.)

• Can we treat \( \lambda_t \) as random?
  • If \( T \) is large. May assume \( \lambda_t \) (macro shocks) iid N(0,\( \sigma_\lambda^2 \)).
  • If \( T \) is small, treat \( \lambda_t \) as fixed.
Heteroskedasticity and Autocorrelation

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

- \( \text{Cov}(u_i) = \Sigma_u = \sigma^2(\theta^2 P_T + Q_T) \), for all \( i \).

- The \( \varepsilon_{it} \) could be autocorrelated and heteroskedastic over time.

(feasible) GLS for random effects models

- Let \( \hat{u}_i = y_i - H_i \hat{\delta} \), where \( \hat{\delta} \) is the OLS estimator.

- Compute \( \hat{\Sigma}_u = \frac{1}{N} \sum_i \hat{u}_i \hat{u}_i' \) or \( \hat{\Omega}_u = I_N \otimes \hat{\Sigma}_u \).

- \( \hat{\delta}_{RE\_GLS} = (\Sigma_i H_i \hat{\Sigma}_{u}^{-1} H_i)^{-1} \Sigma_i H_i \hat{\Sigma}_{u}^{-1} y_i = (H' \hat{\Omega}_{u}^{-1} H)^{-1} H' \hat{\Omega}_{u}^{-1} y \).

- \( \text{Cov}(\hat{\delta}_{RE\_GLS}) = (\Sigma_i H_i \hat{\Sigma}_{u}^{-1} H_i)^{-1} \). [Why?]
• (feasible) GLS for fixed effects models (Kiefer, 1980)
  • An efficient fixed-effect estimator if $\varepsilon_{it}$ are normal.
  • $Q_T y_i = Q_T X_i \beta + Q_T \varepsilon_i$.  $[Q_V y = Q_V X \beta + Q_V \varepsilon]$
  • $\Phi = \text{Cov}(Q_T \varepsilon_i)$.  Since $Q_T$ is singular, so is $\Phi$.

  $\rightarrow$ Use $\Phi^+$ (Moor-Penrose generalized Inverse)

  $[\Lambda = \text{Cov}(Q_V \varepsilon) = I_N \otimes \Phi . ]$

  $\hat{\beta}_{W,\text{GLS}} = \left( \Sigma_i X_i' Q_T \Phi^+ Q_T X_i \right)^{-1} \Sigma_i X_i' Q_T \Phi^+ Q_T y_i$

  $= (X'Q_V \Lambda^+ Q_V X)^{-1} X'Q_V \Lambda^+ Q_V y .$

  • $\text{Cov} (\hat{\beta}_{W,\text{GLS}}) = \left( \Sigma_i X_i' Q_T \Phi^+ Q_T X_i \right)^{-1}$

  • Estimate $\Phi$ 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)' .
The case of time-series heteroskedasticity/autocorrelation, and cross-sectional heteroskedasticity:

- \( \text{Cov}(u_i) = \Sigma_{u,i} \neq \Sigma_{u,j} = \text{Cov}(u_j) \), for all \( i \neq j \).
- The \( \varepsilon_{it} \) could be autocorrelated over time, and heteroskedastic over time and individuals.

For the case of fixed effects:

- Use \( \hat{\beta}_W \) with
  \[
  \text{Cov}(\hat{\beta}_w) = \left( \Sigma_i X_i' Q_T X_i \right)^{-1} \left( \Sigma_i X_i' Q_T \hat{\Phi}_i Q_T X_i \right) \left( \Sigma_i X_i' Q_T X_i \right)^{-1},
  \]
  where \( \hat{\Phi}_i = (Q_T y_i - Q_T X_i \hat{\beta}_W) (Q_T y_i - Q_T X_i \hat{\beta}_W)' \).
- Or use \( \hat{\beta}_{W, GLS} \) with
  \[
  \text{Cov}(\hat{\beta}_{W, GLS}) = \left( \Sigma_i X_i' Q_T \hat{\Phi}^+ Q_T X_i \right)^{-1} \left( \Sigma_i X_i' Q_T \hat{\Phi}^+ \hat{\Phi}_i \hat{\Phi}^+ Q_T X_i \right) \left( \Sigma_i X_i' Q_T X_i \right)^{-1},
  \]

For the case of random effects:

- Use \( \hat{\delta}_{RE, GLS} \).
- \( \text{Cov}(\hat{\delta}_{RE, GLS}) = (\Sigma_i H_i' \hat{\Sigma}_u^{-1} H_i)^{-1} (\Sigma_i H_i' \hat{\Sigma}_u^{-1} \hat{u}_i \hat{u}_i' \hat{\Sigma}_u^{-1} H_i)^{-1} (\Sigma_i H_i' \hat{\Sigma}_u^{-1} H_i)^{-1} \),
  where \( \hat{u}_i = (y_i - H_i \hat{\delta}_{OLS}) \).
[Our alcohol example again]

Within Estimation Results (HETERO/AUTO ADJUSTED)

<table>
<thead>
<tr>
<th>variable</th>
<th>coeff.</th>
<th>std. err.</th>
<th>t-st</th>
</tr>
</thead>
<tbody>
<tr>
<td>beertax</td>
<td>-0.4768</td>
<td>0.2949</td>
<td>-1.6167</td>
</tr>
<tr>
<td>mlda</td>
<td>-0.0019</td>
<td>0.0209</td>
<td>-0.0894</td>
</tr>
<tr>
<td>jailed</td>
<td>0.0147</td>
<td>0.0158</td>
<td>0.9292</td>
</tr>
<tr>
<td>comserd</td>
<td>0.0345</td>
<td>0.1285</td>
<td>0.2684</td>
</tr>
<tr>
<td>unrate</td>
<td>-0.0629</td>
<td>0.0127</td>
<td>-4.9657</td>
</tr>
<tr>
<td>lpinc</td>
<td>1.7964</td>
<td>0.6243</td>
<td>2.8775</td>
</tr>
</tbody>
</table>

Kiefer's Within Estimation Results

<table>
<thead>
<tr>
<th>variable</th>
<th>coeff.</th>
<th>std. err.</th>
<th>t-st</th>
</tr>
</thead>
<tbody>
<tr>
<td>beertax</td>
<td>-0.1833</td>
<td>0.1837</td>
<td>-0.9982</td>
</tr>
<tr>
<td>mlda</td>
<td>-0.0058</td>
<td>0.0175</td>
<td>-0.3343</td>
</tr>
<tr>
<td>jailed</td>
<td>-0.0026</td>
<td>0.0882</td>
<td>-0.0295</td>
</tr>
<tr>
<td>comserd</td>
<td>0.0420</td>
<td>0.1110</td>
<td>0.3784</td>
</tr>
<tr>
<td>unrate</td>
<td>-0.0518</td>
<td>0.0112</td>
<td>-4.6255</td>
</tr>
<tr>
<td>lpinc</td>
<td>2.0991</td>
<td>0.3932</td>
<td>5.3387</td>
</tr>
</tbody>
</table>

Kiefer's Within Estimation Results (HETERO ADJUSTED)

<table>
<thead>
<tr>
<th>variable</th>
<th>coeff.</th>
<th>std. err.</th>
<th>t-st</th>
</tr>
</thead>
<tbody>
<tr>
<td>beertax</td>
<td>-0.1833</td>
<td>0.2254</td>
<td>-0.8135</td>
</tr>
<tr>
<td>mlda</td>
<td>-0.0058</td>
<td>0.0157</td>
<td>-0.3720</td>
</tr>
<tr>
<td>jailed</td>
<td>-0.0026</td>
<td>0.0138</td>
<td>-0.1885</td>
</tr>
<tr>
<td>comserd</td>
<td>0.0420</td>
<td>0.0924</td>
<td>0.4545</td>
</tr>
<tr>
<td>unrate</td>
<td>-0.0518</td>
<td>0.0107</td>
<td>-4.8367</td>
</tr>
<tr>
<td>lpinc</td>
<td>2.0991</td>
<td>0.5934</td>
<td>3.5373</td>
</tr>
</tbody>
</table>