ECON 721: Analysis of Panel Data

Petra E. Todd

Fall, 2016

Contents

Chapter 1

Fixed Effect and Random Effect Models

1.1 Fixed effect model (Mundlak, 1961)

A major motivation for using panel data is that it allows you to control for unobserved heterogeneity.

The idea is that firms or individuals (or countries) have some unique characteristics that are known to the firm or individual but not to the econometrician and that need to be taken into account in estimation to avoid endogeneity bias.

For example, consider the problem of estimating a Cobb-Douglas production function of a firm:

$$\begin{aligned} Y_{it} &= AL_{it}^{\beta_1} K_{it}^{\beta_2} F_i^{\gamma} \\ & \quad \text{in logs} \\ y_{it} &= a + \beta_1 l_{it} + \beta_2 k_{it} + \gamma f_i \end{aligned}$$

Mundlak was estimating an agricultural production function, where each unit was a plot of land observed at a point in time and the unobserved effect represented soil quality.

More generally, the unobserved effect could represent a factor like entrepreneurial capacity.

Here, f_i varies across firms but is assumed to not change over time. We can write the firm-specific effect as:

$$\begin{aligned} \alpha_i &= a + \gamma f_i \\ y_{it} &= \alpha_i + \beta' x_{it} + u_{it}\,, \quad i = 1.., N; t = 1..T \end{aligned}$$

where

$$x_{it} = (x_{it1,}..., x_{itk})' \text{ if we have } k \text{ inputs.}$$

If you have at least two time periods, then you can difference out the unobserved effect:

$$y_{it} - y_{it'} = \beta(x_{it} - x_{it'}) + u_{it} - u_{it'} \quad \text{``within estimator''}$$

If we are willing to assume $E(u_i|x_i) = 0$ (so-called *strict exogeneity*, that the regressors are uncorrelated with past present and future shocks), then $\hat{\beta}_{OLS}$ is an unbiased estimator of β.

Remarks:

(1) The part of the variance accounted for by $x\beta$ might be very small relative to the part accounted for by the unobserved fixed effect.

(2) It is assumed that the goal is to measure the "causal effect" of x on y, not forecasting, so that we do not require an estimate of the fixed effect.

We can write the model as:

$$y_i = x_i\beta + l\alpha_i + u_i$$

where l is a $T \times 1$ vector of ones.

$$x_i = \begin{pmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{pmatrix}, \ y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad u_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}.$$

Stacking across all units (individuals or firms), we get

$$Y = Z_1\beta + Z_2\alpha + U$$

where

$$\begin{aligned} Y &= \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}_{NT\times 1}, \quad Z_1 = \begin{pmatrix} X_1 \\ \vdots \\ X_T \end{pmatrix}_{NT\times k}, Z_l = \begin{pmatrix} l & 0 & \cdots & 0 \\ \vdots & l & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & l \end{pmatrix}_{NT\times N} \\ \alpha &= (\alpha_1, ..., \alpha_N)'. \end{aligned}$$

Now, let

$$\tilde{Z} = (Z_1, Z_2) \text{ and } \gamma = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$$

The model can be written as:

$$Y = \tilde{Z}\gamma + U$$

Assumptions:

$$
\begin{aligned}
(i)\ \ E(U|\tilde{Z}) &= 0 \\
(ii)\ \ Var(U|\tilde{Z}) &= \sigma^2 I_{NT}
\end{aligned}
$$

That is, observations are uncorrelated both over time and across individuals.

Under these assumptions, the least squares estimator is the best linear unbiased estimator (BLUE)

$$\hat{\gamma} = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'Y$$

Note that

$$
\begin{aligned}
\hat{Y} &= \tilde{Z}\hat{\gamma} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'Y = HY, \quad H = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}' \quad \text{"hat matrix"} \\
e &= Y - \hat{Y} = (I - H)Y = MY, \qquad M = I - H, \quad M \text{ and } H \text{ are idempotent} \\
H\tilde{Z} &= \tilde{Z}, \quad M\tilde{Z} = 0
\end{aligned}
$$

The matrices H and M are idempotent.

Key orthogonality condition of the projection:

$$l'\tilde{Z} = Y'M\tilde{Z} = Y'(M\tilde{Z}) = 0$$

Least squares projection algebra:

$$
\begin{aligned}
Y &= Z_1\hat{\gamma}_1 + Z_2\hat{\gamma}_2 + e \\
\hat{\gamma} &= \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix}
\end{aligned}
$$

Define:

$$
\begin{aligned}
M_1 &= I - Z_1(Z_1'Z_1)^{-1}Z_1' \\
M_2 &= I - Z_2(Z_2'Z_2)^{-1}Z_2'
\end{aligned}
$$

Now,

$$
\begin{aligned}
M_2 Y &= M_2 Z_1 \hat{\gamma}_1 + \underbrace{M_2 Z_2}_{0} \hat{\gamma}_1 + \underbrace{M_2 e}_{0} \\
&= M_2 Z_1 \hat{\gamma}_1 + e \\
&\Longrightarrow \\
Z_1' M_2 Y &= Z_1' M_2 Z_1 \hat{\gamma}_1 \\
\hat{\gamma}_1 &= (Z_1' M_2 Z_1)^{-1} Z_1' M_2 Y \\
&= (Z_1' M_2' M_2 Z_1)^{-1} Z_1' M_2' M_2 Y \\
&= (Z_1^{*\prime} Z_1^*)^{-1} Z_1^{*\prime} Y^* \\
&= (Z_1^{*\prime} Z_1^*)^{-1} Z_1^{*\prime} Y
\end{aligned}
$$

What does M_2 look like?

$$
\begin{aligned}
(Z_2' Z_2)^{-1} &= \left[\begin{pmatrix} l' & & \\ & \ddots & \\ & & l' \end{pmatrix}_{N \times NT} \begin{pmatrix} l & & \\ & \ddots & \\ & & l \end{pmatrix}_{N \times NT} \right]^{-1} = \frac{1}{T} I_N \\
&\Longrightarrow \\
\underset{NT \times N}{Z_2} \underset{N \times N}{(Z_2' Z_2)^{-1}} \underset{N \times NT}{Z_2'} &= \begin{pmatrix} \frac{1}{T} l l' & & \\ & \ddots & \\ & & \frac{1}{T} l l' \end{pmatrix} \\
&= diag(Q, ..., Q) \text{ where } Q = \frac{1}{T} l l' \\
&= I_N \otimes Q
\end{aligned}
$$

Thus,

$$
\begin{aligned}
M_2 &= I_{NT} - I_N \otimes Q \\
&= I_N \otimes (I_T - Q)
\end{aligned}
$$

For $Z_1^* = M_2 Z_1$ we have

$$
\begin{aligned}
z_{it}^* &= z_{it} - l\ \bar{z}_i \ , \quad \bar{z}_i = \frac{1}{T}\sum_{t=1}^{T} z_{it} \\
y_{it}^* &= y_{it} - l\ \bar{y}_i \ , \ \ \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} \\
I_T - Q &= I_T - \frac{1}{T} l l' = \begin{pmatrix} 1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & 1 \end{pmatrix} - \frac{1}{T}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \qquad \text{“demeaning matrix”} \\
(I_T - Q)Y &= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{pmatrix} - \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_{it} \\ \frac{1}{T}\sum_{t=1}^{T} y_{it} \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T} y_{it} \end{pmatrix}
\end{aligned}
$$

1.2 Alternative Interpretation: Analysis of Covariance

Consider these alternative models:

(a) $y_{it} = \alpha_i + \beta_i' x_{it} + u_{it}$ (both intercept and slope vary across units)

(b) $y_{it} = \alpha_i + \beta' x_{it} + u_{it}$ (only intercept varies, “within estimator” - most common)

(c) $y_{it} = \alpha + \beta_i x_{it} + u_{it}$ (only slope varies - relatively uncommonly used)

(d) $y_{it} = \alpha + \beta x_{it} + u_{it}$ (neither intercept nor slope varies)

1.2.1 Model (a): Within Group estimators

$$
\begin{aligned}
\hat{\beta}_i &= W_{xx_i}^{-1} W_{xy_i} \\
\alpha_i &= \bar{y}_i - \hat{\beta}_i \bar{x}_i \\
(i &= 1, ..., N) \\
\bar{x}_i &= \frac{1}{T}\sum_{t=1}^{T} x_{it}, \quad \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}
\end{aligned}
$$

where

$$
\begin{aligned}
W_{xx_i} &= \sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \\
W_{xy_i} &= \sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{y}_i)'
\end{aligned}
$$

Define also

$$
W_{yy_i} = \sum_{t=1}^{T}(y_{it} - \bar{y}_i)^2
$$

The sum of squared residuals for individual i:

$$
SSR_i = W_{yy_i} - W_{xy_i}' W_{xx_i}^{-1} W_{xy_i},
$$

so the sum of squared residuals for the model is:

$$
SSR_1 = \sum_{i=1}^{N} SSR_i
$$

1.2.2 Model (b): Individual mean corrected model

The within estimator for β :

$$
\begin{aligned}
\hat{\beta}_w &= W_{xx}^{-1} W_{xy} \\
\alpha_i &= \bar{y}_i - \hat{\beta}_w' \bar{x}_i \quad i = 1, ..., N
\end{aligned}
$$

where

$$\begin{aligned} W_{xx} &= \sum_{i=1}^{N} W_{xx_i} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)' \\ W_{xy} &= \sum_{i=1}^{N} W_{xy_i} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)' \end{aligned}$$

Let

$$W_{yy} = \sum_{i=1}^{N} W_{yy_i} = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)^2.$$

Then, the sum of squared residuals for this model are:

$$SSR_2 = W_{yy} - W_{xy}' W_{xx}^{-1} W_{xy}$$

This estimator is identical to the Mundlak estimator.

1.2.3 Model (d): Pooled Model

$$\begin{aligned} \hat{\beta} &= T_{xx}^{-1} T_{xy} \\ \hat{\alpha} &= \bar{y} - \hat{\beta}'\bar{x} \end{aligned}$$

where

$$\begin{aligned} T_{xx} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})(x_{it}-\bar{x})' \\ T_{xy} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})(y_{it}-\bar{y}) \\ T_{yy} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y})^2 \\ & \text{and} \\ \bar{x} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}, \quad \bar{y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it} \end{aligned}$$

The sum of squared residuals:

$$SSR_3 = T_{yy} - T_{xy}' T_{xx}^{-1} T_{xy}$$

Each model can be tested against the other using the standard F-test. For example, by the Chow test:

$$\frac{(SSR_r - SSR_u) \ / \ (n - k_r) - (n - k_u)}{SSR_r \ / \ (n - k_u)} \sim F(k_u - k_r, n - k_u).$$

Note that the slope coefficient $\hat{\beta}_w$ is consistent if either $T \to \infty$ or $N \to \infty$, but $\hat{\alpha}_i$ is consistent only if $T \to \infty$.

1.3 Random Effect Model (Maddala, 1971)

There is an individual effect, but the effect is random and assumed to be uncorrelated with the included regressors. This individual effect may reflect omitted variables. The model:

$$y_{it} = \beta' x_{it} + v_i + w_{it} \quad (i = 1, .., N; \quad t = 1, .., T)$$

Define

$$u_{it} = v_i + w_{it}.$$

The error term u_{it} has two components, or in the more general case, it can be specified by:

$$u_{it} = v_i + \lambda_t + w_{it}$$

where both v_i and λ_t are random.

We assume that

$$\begin{aligned} E(v_i|x_i) &= 0, E(w_{it}|x_i) = 0 \\ E(v_i w_{it}|x_i) &= 0 \end{aligned}$$

Also,

$$\begin{aligned} E(v_i v_j|x_i) &= \sigma_v^2 \text{ if } i = j \\ &= 0 \text{ else} \\ E(w_{it} w_{is}|x_i) &= \sigma_w^2 \text{ if } i = j \text{ and } t = s \\ &= 0 \text{ else} \end{aligned}$$

Important assumption: All the error components are uncorrelated with the x's.

$$Var(y_i|x_i) = \sigma_v^2 + \sigma_w^2,$$

where σ_v^2 and σ_w^2 are the variance components.

$$\begin{aligned} u_{it} &= v_i + w_{it} \\ y_i &= x_i\beta + u_i \text{ equation for one individual} \\ E(u_i u_i') &= \sigma_w^2 I_T + \sigma_v^2 ll' = \Sigma = \sigma_w^2 (I_T + \frac{\sigma_v^2}{\sigma_w^2} ll') \end{aligned}$$

Therefore,

$$\Sigma^{-1} = \frac{1}{\sigma_w^2}(I_T - \frac{\sigma_v^2}{(T\sigma_v^2 + \sigma_w^2)}ll')$$

The error component v makes observations on the same individual correlated, i.e. $E(u_{is}u_{it}) \neq 0$.

If

$$\begin{aligned} Q &= I_T - \frac{1}{T}ll' \text{ as before, then} \\ y_i^* &= Qy_i = Qx_i\beta + \underbrace{Ql}_{=0} v_i + Qw_i \\ &= Qx_i\beta + Qw_i \\ &= x_i^*\beta + w_i^* \end{aligned}$$

Stacking and running LS regression of y^* on x^* gives as before

$$\begin{aligned} \hat{\beta}_w &= \hat{\beta}_{LS} = (\sum_{i=1}^N x_i'Qx_i)^{-1}(\sum_{i=1}^N x_i'Qy_i) \\ \hat{\alpha}_i &= \bar{y}_i - \hat{\beta}_w'\bar{x}_i. \end{aligned}$$

Now, however, the LS estimator is not BLUE. An efficient estimator can be obtained by GLS.

Recall,

$$\begin{aligned} y &= \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{pmatrix}, X = \begin{pmatrix} x_{11}' \\ \vdots \\ x_{1T}' \\ \vdots \\ x_{N1}' \\ \vdots \\ X_{NT}' \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, v = \begin{pmatrix} v_1 l \\ \vdots \\ \vdots \\ v_N l \end{pmatrix}, \\ w &= \begin{pmatrix} w_{11} \\ \vdots \\ w_{1T} \\ \vdots \\ w_{N1} \\ \vdots \\ w_{NT} \end{pmatrix}, \ u = v + w \end{aligned}$$

Because $E(v|x) = 0$ and $E(w|x) = 0$, we have that $E(u|x) = 0$.

$$var(u|x) = \begin{pmatrix} \sigma_v^2 ll' & & \\ & \ddots & \\ & & \sigma_v^2 ll' \end{pmatrix} + \sigma_w^2 I_{NT} = \Omega$$

where Ω is positive definite

We can find a nonsingular matrix D such that $\Omega^{-1} = D'D$ and $D\Omega D' = I$. Note that $(D'D)^{-1} = D^{-1}D'^{-1} = \Omega$.

Transformed model:

$$\begin{aligned} Dy &= DX\beta + Du \\ y^* &= X^*\beta + u^* \end{aligned}$$

Now, Var(u^*)= $DVar(u|x)D' = D\Omega D' = I$.

By the Gauss-Markov theorem, least squares applied to the transformed data is the MVLU (min variance linear unbiased) estimator.

$$\begin{aligned} \hat{\beta} &= (X'D'DX)^{-1}(X'D'Dy) \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \end{aligned}$$

For the random effect model,

$$\Omega = diag(\Sigma, ..., \Sigma) = \begin{pmatrix} \Sigma & & & \\ & \Sigma & & \\ & & \ddots & \\ & & & \Sigma \end{pmatrix} \longleftarrow \textit{block diagonal}$$

How do we interpret this estimator?

$$\begin{aligned} y_{it} &= \beta x_{it} + v_i + w_{it} \quad (*) \\ \bar{y}_i &= \beta \bar{x}_i + v_i + \bar{w}_i \quad (**) \end{aligned}$$

Subtracting, get

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (w_{it} - \bar{w}_i\ ) \quad (***)$$

From (**),

$$\begin{aligned}\hat{b}_b &= (\sum_{i=1}^{n} \bar{x}_i \bar{x}_i')^{-1} (\sum_{i=1}^{n} \bar{x}_i \bar{y}_i) \quad \text{“between estimator”} \\ &= B_{xx}^{-1} B_{xy}\end{aligned}$$

From (***),

$$\begin{aligned}\hat{b}_w &= (\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)')^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)) \quad \text{“within estimator”} \\ &= W_{xx}^{-1} W_{xy}\end{aligned}$$

Note that

$$\begin{aligned}cov(v_i + \bar{w}_i, w_{it} - \bar{w}_i) &= cov(\bar{w}_i, w_{it}) - cov(\bar{w}_i, \bar{w}_i) \\ &= \frac{1}{T}\sigma_w^2 - \frac{1}{T}\sigma_w^2 = 0\end{aligned}$$

Hence, the within and between estimators are uncorrelated, which suggests that they might be combined to get a more efficient estimator.

Let $\xi_{it} = w_{it} - \bar{w}_i$ and

$$X = \begin{pmatrix} x_{11} - \bar{x}_1 \\ x_{12} - \bar{x}_1 \\ \vdots \\ x_{1T} - \bar{x}_1 \\ \vdots \\ x_{N1} - \bar{x}_N \\ \vdots \\ x_{NT} - \bar{x}_N \end{pmatrix}$$

$$\begin{aligned}var(\xi_i|x_i) &= \sigma_w^2 (I_T - \frac{1}{T} ll') \\ var(\xi|x) &= \sigma_w^2 (I_N \otimes (I_T - \frac{1}{T} ll'))\end{aligned}$$

$$\begin{aligned}
var(\hat{b}_w|x) &= (X'X)^{-1}X'Var(\xi|x)X(X'X)^{-1} \\
&= \sigma_w^2(X'X)^{-1}X'[I_N \otimes I_T - \frac{1}{T}I_N \otimes ll']X(X'X)^{-1} \\
&= \sigma_w^2(X'X)^{-1}[X'X - \frac{1}{T}X'(I_N \otimes ll')X](X'X)^{-1} \\
&= \sigma_w^2 W_{xx}^{-1} \\
&\quad \text{using } (I_N \otimes ll')X = 0
\end{aligned}$$

Similarly, it can be shown that

$$Var(\hat{b}_b|x) = (T\sigma_v^2 + \sigma_w^2)B_{xx}^{-1}$$

Because $\hat{b}_b$ and $\hat{b}_w$ are two unbiased estimators for the same parameter, we can combine them to get a better estimator. Each estimator is weighted by the inverse of its covariance matrix:

$$\begin{aligned}
\hat{\beta}_{GLS} &= \left(\frac{1}{\sigma_w^2}W_{xx} + \frac{1}{(T\sigma_v^2 + \sigma_w^2)}B_{xx}\right)^{-1}\left(\frac{1}{\sigma_w^2}W_{xx}b_{ww} + \frac{1}{(T\sigma_v^2 + \sigma_w^2)}B_{xx}b_b\right) \\
&= (W_{xx} + \theta B_{xx})^{-1}(W_{xx}\hat{b}_w + B_{xx}\hat{b}_b) \\
&\quad \text{where}
\end{aligned}$$

$$\theta = \frac{\sigma_w^2}{T\sigma_v^2 + \sigma_w^2}$$

Feasible GLS requires consistent estimators of σ_v^2 and σ_w^2.

1.4 Correlated Random Effect (CRE) Model

The key difference between a random effects model and a correlated random effects (CRE) model is that CRE allows the individual random effect to be correlated with x_{it}. If random effect is c_i, it allows

$$E(c_i|X_i) = 0$$

Model:

$$y_{it} = \beta' x_{it} + c_i + \omega_{it}, i = 1, .., N, t = 1, .., T$$

Could specify c_i linear in x_{it} :

$$c_i = \sum_{t=1}^{T} \delta_t' x_{it} + v_i$$

More restrictive approach:

$$c_i = \delta_t' \bar{x}_i + v_i$$

which implies model:

$$\begin{aligned} y_{it} &= \beta' x_{it} + \delta' \bar{x}_i + v_i + \omega_{it} \\ &= \gamma' z_{it} + u_{it} \end{aligned}$$

where $\gamma = (\beta' \ \delta')$, $z_{it} = (x_{it}' \ \bar{x}_i')$, $u_{it} = v_i + \omega_{it}$.

In stacked form,

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \beta + \begin{pmatrix} l\bar{x}_1' \\ \vdots \\ l\bar{x}_N' \end{pmatrix} \delta + \begin{pmatrix} l & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & l \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} + \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \end{pmatrix}$$

Define

$$u = \begin{pmatrix} l & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & l \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} + \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \end{pmatrix}$$

$$\begin{aligned} E(u_i|x_i) &= 0 \\ E(u_i u_j) &= E((\omega_i + l v_i)(\omega_j + l v_j)') \\ &= \sigma_\omega^2 I_T + \sigma_v^2 l l' = \Sigma \text{ if } i = j \\ &= 0 \text{ else} \end{aligned}$$

and

$$\Sigma^{-1} = \frac{1}{\sigma_\omega^2}(I_T - \frac{\sigma_\omega^2}{T\sigma_v^2 + \sigma_\omega^2})$$

This is similar to the random effects model, only that the GLS estimator simplifies....

Rewrite model as

$$\begin{aligned} y_{it} &= \beta'(x_{it} - \bar{x}_i) + (\delta + \beta)'\bar{x}_i + v_i + \omega_{it} \\ &= \gamma' z_{it} + v_i + \omega_{it} \end{aligned}$$

where

$$\gamma = \begin{pmatrix} \beta \\ \delta + \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \tau \end{pmatrix} \qquad z_{it} = \begin{pmatrix} x_{it} - \bar{x}_i \\ \bar{x}_i \end{pmatrix}$$

The GLS estimator of γ is given by

$$\hat{\gamma} = \begin{pmatrix} \hat{\beta} \\ \hat{\tau} \end{pmatrix} = (W_{ZZ} + \theta B_{ZZ})^{-1}(W_{ZY} + \theta B_{ZY})$$

where

$$\theta = \frac{\sigma_\omega^2}{T\sigma_V^2 + \sigma_\omega^2}$$

but

$$\bar{z}_i = \begin{pmatrix} 0 \\ \bar{x}_i \end{pmatrix}$$

Therefore,
(i)

$$\begin{aligned} \mathrm{B}_{ZZ} &= T \textstyle\sum_{i=1}^{N} \bar{z}_i \bar{z}_i{}' = T \sum_{i=1}^{N} \begin{pmatrix} 0 \\ \bar{x}_i \end{pmatrix} (0' \ \bar{x}_i') \\ &= \mathrm{T} \textstyle\sum_{i=1}^{N} \begin{pmatrix} 0 & 0 \\ 0 & \bar{x}_i \bar{x}_i{}' \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & B_{XX} \end{pmatrix} \end{aligned}$$

(ii)

$$(z_{it} - \bar{z}_i) = \begin{pmatrix} x_{it} - \bar{x}_i \\ 0 \end{pmatrix}$$

$$\begin{aligned} W_{ZZ} &= \sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \bar{z}_i)(z_{it} - \bar{z}_i)' \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} \begin{pmatrix} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} W_{XX} & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

Similarly,

$$W_{ZY} = \begin{pmatrix} W_{XY} \\ 0 \end{pmatrix} \qquad B_{ZY} = \begin{pmatrix} B_{XY} \\ 0 \end{pmatrix}$$

$$\begin{aligned}
\hat{\gamma}_{GLS} &= (W_{ZZ} + \theta B_{ZZ})^{-1}(W_{ZY} + \theta B_{ZY}) \\
&= \left(\begin{pmatrix} W_{XY} & 0 \\ 0 & 0 \end{pmatrix} + \theta \begin{pmatrix} 0 & 0 \\ 0 & B_{XX} \end{pmatrix} \right)^{-1} \left(\begin{pmatrix} W_{XY} \\ 0 \end{pmatrix} + \theta \begin{pmatrix} 0 \\ B_{XY} \end{pmatrix} \right) \\
&= \begin{pmatrix} W_{XX}^{-1} W_{XY} \\ B_{XX}^{-1} B_{XY} \end{pmatrix} = \begin{pmatrix} b_w \\ b_b \end{pmatrix}
\end{aligned}$$

$$\hat{\beta} = b_w$$
$$\hat{\tau} = \hat{\beta} + \hat{\delta}$$

so

$$\hat{\delta} = b_b - b_w$$

$\hat{\delta}$ is a measure of how much the "between" and the "within" estimators differ.

$$\hat{c}_i = \bar{x}_i(b_b - b_w)$$

How do we test between the RE model and the CRE model?

H$_0$:$\delta = 0$ (pure random effect model) H$_1$:$\delta \neq 0$

Under H_0, the RE model holds and hence

$$\begin{aligned}
SSR_0 &= \sum_{i=1}^{n} (y_i - x_i \hat{\beta}_{GLS})' \Sigma^{-1} (y_i - x_i \hat{\beta}_{GLS}) \,\tilde{}\, \chi^2(NT - k) \\
SSR_1 &= \sum_{i=1}^{n} (y_i - x_i \hat{\beta}_{GLS} - l \bar{x}_i' \hat{\delta}_{GLS})' \Sigma^{-1} (y_i - x_i \hat{\beta}_{GLS} - l \bar{x}_i' \hat{\delta}_{GLS}) \,\tilde{}\, \chi^2(NT - 2k)
\end{aligned}$$

These are like GMM χ^2 statistics with optimal weighting matrices.

We could base test directly on SSR_0 or else do a Chow test:

$$\frac{(SSR_0 - SSR_1)/k}{SSR_1/(NT - 2k)} \,\tilde{}\, F(k, NT - 2k)$$

Alternative Test: Hausman Test

Under the null, both estimators are consistent. Under the alternative, β_0 is inconsistent. Under the null,

$$\sqrt{N}(\hat{\beta}_0 - \beta) \,\tilde{}\, N(0, \Sigma_0)$$
$$\sqrt{N}(\hat{\beta}_1 - \beta) \,\tilde{}\, N(0, \Sigma_1)$$

(Σ_0 is Cramer-Rao bound)

$$N(\hat{\beta}_1 - \hat{\beta}_0)'[var(\hat{\beta}_1) - var(\hat{\beta}_0)]^{-1}(\hat{\beta}_1 - \hat{\beta}_0) \,\tilde{}\, \chi^2(k)$$

where $\hat{\beta}_1 - \hat{\beta}_0$ correspond to $\hat{\beta}_w - \hat{\beta}_{GLS}$.

Under the alternative, $\hat{\beta}_0 = \hat{\beta}_{GLS}$ is biased, so test statistic is $\tilde{}\,\chi^2(k, \lambda)$ where

$$\lambda = \bar{q}' var(\hat{q})\bar{q}$$
$$\hat{q} = \hat{\beta}_w - \hat{\beta}_{GLS}$$
$$\bar{q} = plim\hat{q}$$

Note: within estimator is consistent under CRE, random effects GLS estimator is not.

Chapter 2

Dynamic Panel Data Models

x_{it} : $K \times 1$ vector of explanatory variables

α_i : individual effect

λ_t : time effect

Dynamic model:

$$y_{it} = \gamma y_{it-1} + \beta' x_{it} + \alpha_i + \lambda_t + u_{it}, \; i = 1, .., N \text{ and } t = 1, ..T$$

$$E(u_{it}u_{js}) = \sigma_u^2 \text{ if } i = j, t = s$$
$$else = 0 \text{ otherwise}$$

2.1 Fixed effect model:

Assume $|\gamma| < 1, \lambda_t = 0$

$\alpha_i = \mu + c_i$

y_{i0} observable, fixed

The least squares estimator is not consistent

$$y_{it} - y_{it-1} = \gamma(y_{it-1} - y_{it-2}) + (x_{it} - x_{it-1})'\beta + (u_{it} - u_{it-1})$$

$$cov(y_{it-1} - y_{it-2}, u_{it} - u_{it-1}) \neq 0$$

Remark: The same is the case if the individual fixed effect is eliminated by deviation from the individual's mean.

Solution: Instrumental Variables

$$\begin{aligned} y_{i2} - y_{i1} &= \gamma(y_{i1} - y_{i0}) + (x_{i2} - x_{i1})'\beta + (u_{i2} - u_{i1}) \\ &\vdots \\ y_{iT} - y_{iT-1} &= \gamma(y_{iT-1} - y_{iT-2}) + (x_{iT} - x_{iT-1})'\beta + (u_{iT} - u_{iT-1}) \end{aligned}$$

$y_{it-2} - y_{it-3}$ can serve as as instrument for $y_{it-1} - y_{it-2}$.

$$\begin{aligned} cov(y_{it-1} - y_{it-2}, y_{it-2} - y_{it-3}) &\neq 0 \\ cov(y_{it-2} - y_{it-3}, u_{it} - u_{it-1}) &= 0 \end{aligned}$$

Note that other lagged $y's$ can also serve as instruments.

Include all possible lags be in vector

$$Z = \begin{pmatrix} y_{it-2} - y_{it-3} \\ \vdots \\ y_{i2} - y_{i1} \end{pmatrix}$$

Moment conditions:

$$E\left\{[(y_{it} - y_{it-1}) - (y_{it-1} - y_{it-2}) - (x_{it} - x_{it-1})\beta - (u_{it} - u_{it-1})]\ Z\right\} = 0$$

- Require at least three periods of data.

- Can set this up as a GMM problem with optimal weighting matrix, to choose the optimal combination of instruments. For data periods further in the future, there are more available lags to use as instruments.

- See Hansen (1982), Arellano & Bond (1991). Arellano and Bond's estimation routines are available in Stata.

- Under strict exogeneity of the covariates, can use past, present and future values as instruments.

2.2 Predetermined verses strictly exogenous regressors

Model

$$y_{it} = \alpha y_{it-1} + \beta x^*_{it} + \eta_i + v_{it}$$

Two cases:

(1) x^*_{it} predetermined, implies

$$\begin{aligned} E(x^*_{it} v_{is}) &\neq 0 \text{ for } s < t \\ &= 0 \text{ otherwise} \end{aligned}$$

(2) x^*_{it} strictly exogenous implies

$$E(x^*_{it} v_{is}) = 0 \text{ for all } s, t$$

For the predetermined case,

$$\begin{aligned} y_{it} - y_{it-1} &= \alpha(y_{it-1} - y_{it-2}) + \beta(x^*_{it} - x^*_{it-1}) + (v_{it} - v_{it-1}) \\ &\quad (x^*_{it} - x^*_{it-1}) \text{ and } (v_{it} - v_{it-1}) \text{ potentially correlated} \\ &\quad x^*_{it-1} \text{ is a valid instrument} \end{aligned}$$

- The predetermined assumption might be reasonable, for example, if x^*_{it} chosen after seeing y_{it-1}.

- This problem can also be solved using instrumental variables.

Example #1: Estimating an educational production function

Observations on test scores of child i at time t along with information on family and school inputs (x_{it}).

Value-added model:

$$T_{it} = \alpha T_{it-1} + \beta x_{it} + \eta_i + v_{it}$$

In differenced form (to eliminate the fixed effect),

$$T_{it} - T_{it-1} = \alpha(T_{it-1} - T_{it-2}) + \beta(x_{it} - x_{it-1}) + (v_{it} - v_{it-1})$$

Reason to believe that x_{it} might be responsive to v_{it-1}. For example, if a child has a low reading score, the parents might purchase more books (one of the inputs). This leads to a correlation between x_{it} and v_{it-1}.

Anything that is in the fixed effect is also potentially a valid instrument (due to assumption that fixed effect no longer enters equation after differencing).

Also, can use information on inputs of older siblings - chosen at a time before the test score for the younger sibling could be observed.

Example #2: Cigarette Addiction (Becker, Murphy)

Model

$$C_{it} = \theta_1 C_{it-1} + \beta\theta_2 C_{it+1} + \gamma p_{it} + \eta_i + \delta_t + v_{it}$$

C_{it} : consumption at time t

C_{it-1} : past consumption C_{it+2} : future consumption p_{it} : price

η_i : captures variations in marginal utility of wealth (person specific fixed effect)
δ_t : time effect

(a) a rational addict will decrease current consumption in response to an anticipated decrease in future consumption

(b) addiction implies that past consumption increases current consumption

(c) Assume prices are exogenous with respect to η_i, use lagged and future prices as instruments

Chapter 3

Discrete Panel Data

Consider model

$$
\begin{aligned}
y_i &= 1 \text{ if } x_i\beta - \varepsilon_i \geq 0 \\
&= 0 \text{ else}
\end{aligned}
$$

With panel data, we can allow for individual unobserved heterogeneity.

Individual probability

$$F(x_{it}\beta + \alpha_i)$$

The average probability conditional on x

$$= \int F(\beta' x_{it} + \alpha_i) dH(\alpha|x)$$

3.1 MLE Estimator

$\Pr(y_{it} = 1|x_{it}) = F(\beta' x_{it} + \alpha_i)$

As $T \to \infty$ could get a consistent estimator for α_i using data on person i.

If T is small, α_i cannot be consistently estimated, and this problem translates into inconsistency of β.

Even if $N \to \infty$, MLE for β is inconsistent.

3.1.1 Example: Logit Model, Show that the MLE estimator is inconsistent

$$\begin{aligned}
\log L &= \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}\ln(e^{x_{it}'\beta+\alpha_i}) - y_{it}\ln(1+e^{x_{it}'\beta+\alpha_i}) - \\
&\quad \sum_{i=1}^{n}\sum_{t=1}^{T}(1-y_{it})\ln(1+e^{x_{it}'\beta+\alpha_i}) \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}(x_{it}'\beta+\alpha_i) - \sum_{i=1}^{n}\sum_{t=1}^{T}\log(1+e^{x_{it}'\beta+\alpha_i})
\end{aligned}$$

Assume $T = 2, x_{i1} = 0$ and $x_{i2} = 1$.

$$\begin{aligned}
\frac{\partial \log L}{\partial \beta} &= \sum_{i=1}^{n}\sum_{t=1}^{2}(y_{it}x_{it} - \frac{e^{\beta' x_{it}+\alpha_i}}{1+e^{\beta' x_{it}+\alpha_i}}x_{it}) = 0 \\
&= \sum_{i=1}^{n}\left(y_{i2} - \frac{e^{\beta+\alpha_i}}{1+e^{\beta+\alpha_i}}\right) = 0 \quad \text{(plugging in for } x_{i1}, x_{i2}) \quad (*) \\
\frac{\partial \log L}{\partial \alpha_i} &= \sum_{t=1}^{2}(y_{it} - \frac{e^{\beta' x_{it}+\alpha_i}}{1+e^{\beta' x_{it}+\alpha_i}}) = 0 \\
&\Longrightarrow \\
\sum_{t=1}^{2} y_{it} &= \sum_{t=1}^{2}\frac{e^{\beta' x_{it}+\alpha_i}}{1+e^{\beta' x_{it}+\alpha_i}}
\end{aligned}$$

Solve for $\hat{\alpha}_i$:

$$\begin{aligned}
\text{If } \sum_t y_{it} &= 2, \text{ then } 2 = \frac{e^{\alpha}}{1+e^{\alpha}} + \frac{e^{\alpha+\beta}}{1+e^{\alpha+\beta}} \\
&= \frac{1}{1+e^{-\alpha}} + \frac{e^{\beta}}{e^{-\alpha}+e^{\beta}} \\
&\Longrightarrow \alpha_i = \infty \\
\text{If } \sum_t y_{it} &= 0, \text{ then } \exp(\beta' x_{it} + \alpha_i) = 0, \ \alpha_i = -\infty \\
\text{If } y_{i1} + y_{i2} &= 1, \text{ then } 1 = \frac{e^{\alpha}}{1+e^{\alpha_i}} + \frac{e^{\alpha+\beta}}{1+e^{\alpha_i+\beta}} \\
&\Longrightarrow \alpha_i = \frac{-\beta}{2}
\end{aligned}$$

Suppose that n_2 individuals have $\sum_t y_{it} = 2$

Suppose that n_1 individuals have $\sum_t y_{it} = 1$

and suppose that $N - n_1 - n_2$ individuals have $\sum_t y_{it} = 0$.

Substitute into (*) to get

$$\sum_{i=1}^{n} \frac{e^{\beta+\alpha_i}}{1+e^{\beta+\alpha_i}} = n_2 + n_1 \frac{\exp(\beta/2)}{1+\exp(\beta/2)} = \sum_{i=1}^{n} y_{i2}$$

Solve for $\hat{\beta}$:

$$\hat{\beta} = 2 \left\{ \ln\left(\frac{1}{N}\left(\sum_{i=1}^{n} y_{i2} - n_2\right)\right) - \ln\left(\frac{1}{N}\left(n_1 + n_2 - \sum_{i=1}^{n} y_{i2}\right)\right) \right\}$$

Analyze term by term:

$$\begin{aligned} \text{plim}\frac{1}{N}\sum_{i=1}^{n} y_{i2} &= \Pr(y_{i1}=0, y_{i2}=1) + \Pr(y_{i1}=1, y_{i2}=1) \\ \text{plim}\frac{n_2}{N} &= Pr(y_{i1}=1, y_{i2}=1) \end{aligned}$$

so,

$$\begin{aligned} & \frac{1}{N}(\sum_{i=1}^{n} y_{i2} \ - n_2) \underset{p}{\rightarrow} \Pr(y_{i1}=0, y_{i2}=1) \\ = \ & \frac{1}{N}\sum_{i=1}^{n} \frac{\exp(\beta+\alpha_i)}{[1+\exp(\alpha_i)][1+\exp(\beta+\alpha_i)]} \end{aligned}$$

Note

$$\begin{aligned} \Pr(y_{i1}=0) &= 1 - \frac{e^{\alpha_i}}{1+e^{\alpha_i}} \\ &= \frac{1}{1+e^{\alpha_i}} \\ \Pr(y_{i2} &= 1) = \frac{e^{\alpha_i+\beta}}{1+e^{\alpha_i+\beta}} \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{1}{N}(n_1 + n_2 + \sum_{i=1}^{n} y_{i2}) &= \frac{1}{N}\sum_{i=1}^{n} \Pr(y_{i1}=1, y_{i2}=0) \\ &= \frac{1}{N}\sum_{i=1}^{n} \frac{e^{\alpha_i}}{1+e^{\alpha_i}} \frac{1}{1+e^{\alpha_i+\beta}} \end{aligned}$$

Using these results, we get

$$\text{plim } \hat{\beta} = 2\left\{ ln(\frac{1}{N}\sum_{i=1}^{n} \frac{e^{\alpha_i+\beta}}{(1+e^{\alpha_i+\beta})(1+e^{\alpha_i})}) - ln(\frac{1}{N}\sum_{i=1}^{n} \frac{e^{\alpha_i}}{(1+e^{\alpha_i})(1+e^{\alpha_i+\beta})}) \right\}$$

If α_i were constant, would get

$$= \quad 2\{\beta + \alpha - \alpha\} = 2\beta, \text{ inconsistent.}$$

For the linear probability model, can take differences to eliminate α_i.

Can't do this for the logit or probit model.

3.1.2 Solution for the fixed effects logit model: (Neyman and Scott, 1948)

Find functions $\psi(y_1, ...y_N|\beta)$ such that they are independent of α_i (incidental parameters) at the true parameter value β.

For example, in the linear model, first differences eliminates α_i.

In nonlinear models, ψ can be harder to find

Fixed effect logit model:

$$\begin{aligned}
y_i &= (y_{i1}, ..., y_{iT}) \\
\Pr(y_i|x_i) &= \Pi_{t=1}^T \left(\frac{\exp(\beta' x_{it} + \alpha_i)}{1 + \exp(\beta' x_{it} + \alpha_i)} \right)^{y_{it}} \left(1 - \frac{\exp(\beta' x_{it} + \alpha_i)}{1 + \exp(\beta' x_{it} + \alpha_i)} \right)^{1-y_{it}} \\
&= \frac{\exp(\beta' \sum_{t=1}^T x_{it} y_{it} + \alpha_i \sum_{t=1}^T y_{it})}{\Pi_{t=1}^T 1 + \exp(\beta' x_{it} + \alpha_i)}
\end{aligned}$$

Define

$$
\begin{aligned}
y_i^s &= \sum_{t=1}^{T} y_{it} \\
\Pr(y_i^s &= m_i) = \sum_{d \in \tilde{B}_i} \frac{\exp(\beta' \sum_{t=1}^{T} d_{it} x_{it} + \alpha_i m_i)}{\Pi_{t=1}^{T}(1 + \exp\{\beta' x_{it} + \alpha_i)} \\
d &\in \tilde{B}_i \text{ represents all the different ways of summing to } m_i \\
\tilde{B}_i &= \{(d_{i1}, ..., d_{iT}) : d_{it} = 0 \text{ or } 1 \text{ and } \sum_{t=1}^{T} d_{it} = m_1\} \\
\Pr(y_i) &= \Pr(y_i | y_i^s = m_i) \Pr(y_i^s = m_i) \\
&\Longrightarrow \\
\Pr(y_i | y_i^s &= m_i) = \frac{\Pr(y_i)}{\Pr(y_i^s = m_i)} \\
\Pr(y_i | y_i^s &= m_i) = \frac{\exp(\beta' \sum_{t=1}^{T} x_{it} y_{it})}{\sum_{d \in \tilde{B}_i} \exp(\beta' \sum_{t=1}^{T} x_{it} y_{it})} \\
&\Longrightarrow \text{ can estimate the conditional likelihood, which does not} \\
&\quad \text{depend on } \alpha_i
\end{aligned}
$$

Here, the estimator for β will be consistent.

Conditioning on $y_i^s = m_i$ removes dependence on α_i. y_i^s is a sufficient statistic for α_i.

Can estimate β consistently for $N \to \infty$, T fixed by solving the conditional likelihood problem (note that you do throw away data)

3.2 Random Effects Models

$$\Pr(y_{it} = 1|x_{it}) = F(\beta' x_{it} + \alpha_i)$$

Assume that the incidental parameters α_i are independent of x_i and are a randomly sampled from a univariate distribution H indexed by a finite number of parameters δ.

Log likelihood function

$$\log L = \sum_{i=1}^{n} \log \int \Pi_{t=1}^{T} F(\beta' x_{it} + \alpha)^{y_{it}} [1 - F(\beta' x_{it} + \alpha)]^{1-y_{it}} dH(\alpha|\delta).$$

If a_i is correlated with x_{it}, then will be subject to omitted variable bias.

3.3 Correlated random effect model

Would like to relax assumption that incidental parameters are independent from observables.

One approach due to Chamberlain (1980, 1984) assumes that

$$\alpha_i = \sum_{t=1}^{T} a_t' x_{it} + \eta_i = a' x_i + \eta_i$$

Given these assumptions, the log likelihood is given by

$$\log L = \sum_{i=1}^{n} \log \int \Pi_{t=1}^{T} F(\beta' x_{it} + a' x_i + \eta)^{y_{it}} [1 - F(\beta' x_{it} + a' x_i + \eta)]^{1-y_{it}} dH^*(\eta),$$

where H^* is a univariate distribution function for η.

For example, could assume that F is standard normal and choose H^* to be a normal random variable with mean zero and variance σ_η^2.

$$y_{it} \;=\; 1 \text{ if } {}'x_{it} + a'x_i + \eta_i + u_{it} > 0$$

where

$$u_i + 1\eta_i \;\tilde{}\; N(0, I_T + \sigma_\eta^2 11')$$

This is a multivariate probit model.

Requires evaluation of T dimensional integrals.

Usually need to use simulation techniques to solve these problems.

Chapter 4

Semiparametric Panel Data Models

4.1 Panel Maximum Score Estimator

model

$$d_{it} = 1(x_{it}\beta_0 + z_i\gamma_0 + c_i - v_{it} > 0), \quad i = 1..N, t = 1..T$$

examples

- individual choice with unobserved ability effect

$$\begin{aligned} E(d_{i0}|x_{i0}, x_{i1}, z_i, c_i) &= F_{v_0}(x'_{i0}\beta_0 + Z_i\gamma_0 + c_i|x_{i0}, x_{i1}, , z_i, c_i) \\ E(d_{i1}|x_{i0}, x_{i1}, z_i, c_i) &= F_{v1}(x'_{i1}\beta_0 + Z_i\gamma_0 + c_i|x_{i0}, x_{i1}, , z_i, c_i) \end{aligned}$$

Suppose the marginal distribution of v_0 and v_1 are the same. Then

$$\begin{aligned} x'_{i0}\beta_0 &> x'_{i1}\beta_1 &\Leftrightarrow\ & E(d_{i0}|x_{i0}, x_{i1}, z_i, c_i) > E(d_{i1}|x_{i0}, x_{i1}, z_i, c_i) \\ x'_{i0}\beta_0 &= x'_{i1}\beta_1 &\Leftrightarrow\ & E(d_{i0}|x_{i0}, x_{i1}, z_i, c_i) = E(d_{i1}|x_{i0}, x_{i1}, z_i, c_i) \\ x'_{i0}\beta_0 &< x'_{i1}\beta_1 &\Leftrightarrow\ & E(d_{i0}|x_{i0}, x_{i1}, z_i, c_i) < E(d_{i1}|x_{i0}, x_{i1}, z_i, c_i) \end{aligned}$$

Can use the same approach as in Manski's maximum score estimator to get a consistent estimate of β_0.

Recall that the maximum score assumed:

$$
\begin{aligned}
med(\varepsilon|x) &= 0.5, \quad F_\varepsilon(x'\beta_0) = 0.5 \;\text{ iff } x'\beta_0 = 0 \\
S_N(b) &= \frac{1}{n}\sum_{i=1}^{n}(2d_i - 1)sgn(x_i b)
\end{aligned}
$$

In this case, the objective function is

$$
\begin{aligned}
S_N(b) &= \frac{1}{n}\sum_{i=1}^{n} sgn(x'_{i0}\beta_0 - x'_{i1}\beta_0)(d_{0i} - d_{1i}) \\
&= \frac{1}{n}\sum_{i=1}^{n} sgn(w'\beta_0)(d_{0i} - d_{1i})
\end{aligned}
$$

where

$$
\begin{aligned}
w'\beta_0 &> 0 \text{ if } E(d_{0i}|\Omega) > E(d_{1i}|\Omega) \qquad (*) \\
w'\beta_0 &< 0 \text{ if } E(d_{0i}|\Omega) < E(d_{1i}|\Omega)
\end{aligned}
$$

Show the limiting function of the objective function is maximized at the true value, i.e.

$$
\begin{aligned}
H(\beta_0) - H(b) &> 0 \quad b \neq \beta_0. \\
& E\{sgn(w'\beta_0)(d_{0i} - d_{1i})\ \} - E\{sgn(w'b)(d_{0i} - d_{1i})\ \} \\
&= E\{sgn(w_i\beta_0) - sgn(w_i b))(d_0 - d_1)\} \\
&= E\{sgn(w_i\beta_0) - sgn(w_i b))E(d_0 - d_1|w)\} \\
&= \int_{\{w \text{ s.t. } w'\beta_0 > 0 > w'b\}} sgn(w_i\beta_0)(1 - (-1))E(d_0 - d_1|w)dF_w \\
& + \int_{\{w \text{ s.t. } w'b > 0 > w'\beta_0\}} (-1)sgn(w_i\beta_0)(-1 - 1))E(d_0 - d_1|w)dF_w \\
& +0 \\
&= 2\int_{\{w \text{ s.t. } sgn(w'b) \neq sgn(w'\beta_0)\}} sgn(w'\beta_0)E(d_0 - d_1|w)dF_w \geq 0 \\
& \text{(using (*))}
\end{aligned}
$$

Remarks:

(1) CDF can differ across t, but marginals must be the same (stationary)

(2) Same idea generalized for $T > 2$

$$\sum_{i=1}^{n} \sum_{\tau=1}^{T} \sum_{\substack{s \neq \tau \\ s < \tau}} sgn\{(x_{is} - x_{i\tau})'b)(d_s - d_\tau)$$

(3) Could do smoothed version of the estimator

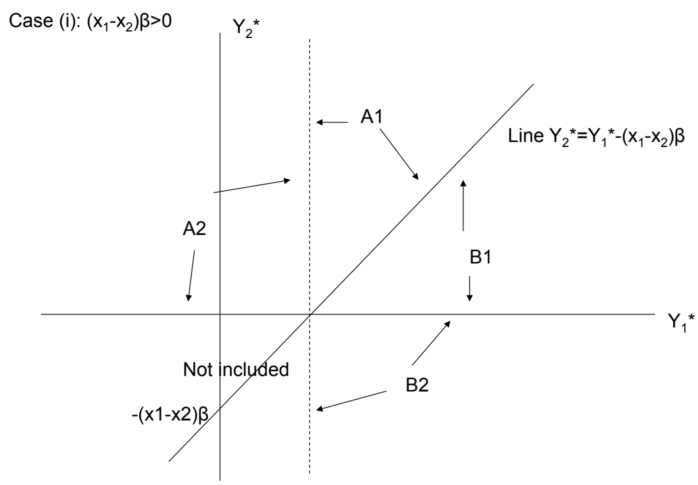
4.2 Honore's estimator

"Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects"

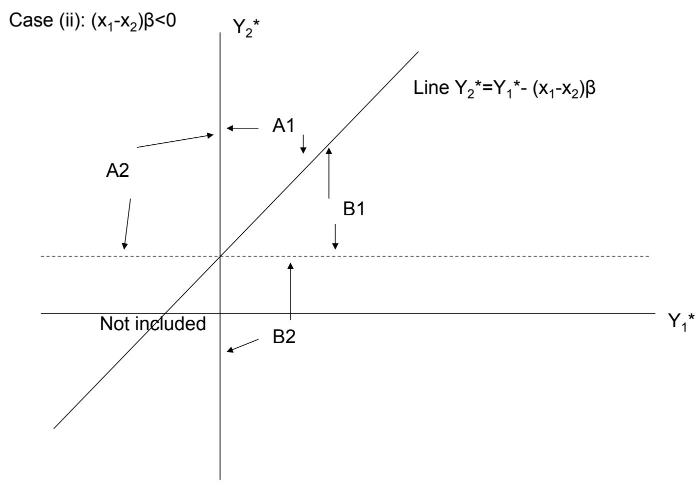
$$\begin{aligned}
y_t^* &= \alpha + x_t\beta + \varepsilon_t \quad t = 1, 2 \\
& \quad \varepsilon_1, \varepsilon_2 \text{ iid} \\
y_1 &= y_1^* \text{ if } \; y_1^* \geq 0 \\
&= 0 \text{ else} \\
y_2 &= y_2^* \text{ if } \; y_2^* \geq 0 \\
&= 0 \text{ else}
\end{aligned}$$

Can write as

$$\begin{aligned}
y_2^* &= y_1^* - (x_1 - x_2)\beta + \varepsilon_2 - \varepsilon_1 \\
&= y_1^* - \Delta x\beta + \varepsilon_2 - \varepsilon_1
\end{aligned}$$



$$
\begin{aligned}
A_1 &= \{(y_1^*, y_2^*) : y_1^* > \Delta x\beta, \quad y_2^* > y_1^* - \Delta x\beta\} \\
A_2 &= \{(y_1^*, y_2^*) : y_1^* \leq \Delta x\beta, \quad y_2^* > 0\} \\
B_1 &= \{(y_1^*, y_2^*) : y_1^* > \Delta x\beta, \; 0 < \; y_2^* < y_1^* - \Delta x\beta\} \\
B_2 &= \{(y_1^*, y_2^*) : y_1^* > \Delta x\beta, \; y_2^* \leq 0\}
\end{aligned}
$$



$$
\begin{aligned}
A_1 &= \{(Y_1^*, Y_2^*): \quad Y_1^* > 0,\ Y_2^* > Y_1^* - \Delta x\beta\} \\
A_2 &= \{(Y_1^*, Y_2^*): \quad Y_1^* \leq 0,\ Y_2^* > -\Delta x\beta\} \\
B_1 &= \{(Y_1^*, Y_2^*): \quad Y_1^* > 0,\ -\Delta x\beta < Y_2^* < Y_1^* - \Delta x\beta\} \\
B_2 &= \{(Y_1^*, Y_2^*): \quad Y_1^* > 0,\ Y_2^* \leq -\Delta x\beta\}
\end{aligned}
$$

(1) Truncated data, only observe $Y_1^* > 0$ and $Y_2^* > 0$

$$
\Pr((Y_1, Y_2) \in A_1 \ |X_1, X_2) = \Pr((Y_1, Y_2) \in B_1 | X_1, X_2)
$$

which implies

$$
E[[1\{(Y_1, Y_2) \in A_1 \ \} - 1\{(Y_1, Y_2) \in B_1\}]\Delta x] = 0
$$

Note that y_1 and y_2 symmetrically distributed around the line follows as a consequence of ε_1 and ε_2 being iid. We do not have to assume symmetry (as in Powell's LAD estimator).

(2) Censored data

$$\Pr((Y_1, Y_2) \in A_1 \cup A_2 \ |X_1, X_2) = \Pr((Y_1, Y_2) \in B_1 \cup B_2 \ |X_1, X_2)$$

which implies

$$E[[1\{(Y_1, Y_2) \in A_1 \cup A_2\} - 1\{(Y_1, Y_2) \in B_1 \cup B_2\}]\Delta x] = 0$$