Lecture Notes:
Econometrics II

Based on lectures by **Marko Mlikota** in Spring semester, 2025

Draft updated on March 21, 2025

These lecture notes were taken in the course *Econometrics II* taught by **Marko Mlikota** at Graduate of International and Developoment Studies, Geneva as part of the International Economics program (Semester II, 2024).

Currently, these are just drafts of the lecture notes. There can be typos and mistakes anywhere. So, if you find anything that needs to be corrected or improved, please inform at `jingle.fu@graduateinstitute.ch`.

Contents

Lecture 1.

Review of Econometrics I

1.1 Basic assumtions

As awe know,

$$\hat{\beta} = (X'X)^{-1}X'y \xrightarrow{p} \beta$$

if

1. Model is correctly specified: $y_i = x_i'\beta + u_i$
2. X is full rank
3. $\mathbb{E}[x_i u_i] = 0$: x_i is exogenous.
4. Unbiased CIA: $\mathbb{E}[u_i|x_i] = 0$

Theorem 1.1.1 (Frisch-Waugh-Lovell (FWL) theorem)**.**
Recall: $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = P_X Y$, $Y = \hat{Y} + \hat{U} \to \hat{U} = (I - P_X)Y = M_X Y$.
Take $Y = X_1\beta_1 + X_2\beta_2 + U = X\beta' + U$, let $P_1 = X_1(X_1'X_1)^{-1}X_1'$, $M_1 = I - P_1$.
And write $M_1 Y = M_1 X_2 b + M_1 U$, then

$$\hat{\beta}_{2,OLS} = \hat{b}.$$

1.2 Endogeneity

Three reasons for endogeneity:

1. Measurement error: x_i is measured with error.
   Assume the true Regression is: $y_i = x_i^{*\prime}\beta + \varepsilon$, $\mathbb{E}[x_i^*\varepsilon_i] = 0$, we run: $y_i = x_i'\beta + u_i$, $x_i = x_i^* + v_i$, $u_i = \varepsilon_i - v_i'\beta$.

$$\begin{aligned}
\mathbb{E}[x_i u_i] &= \underbrace{\mathbb{E}[x_i\varepsilon_i]}_{0} - \mathbb{E}[x_i v_i']\beta \\
&= -\mathbb{E}[(x_i^* + v_i)v_i']\beta \\
&= -\underbrace{\mathbb{E}[x_i^* v_i']}_{0}\beta - \mathbb{E}[v_i v_i']\beta \\
&= -\mathbb{E}[v_i v_i']\beta
\end{aligned}$$

2. Simultaneity(Reverse causality): x_i is endogenous.
   $y_i = x_i'\beta + u_i = x_{i1}^*\beta_1 + x_{i2}\beta_2 + u_i$, $x_i = z_i'\gamma + y_i\delta + v_i$.

3. Omitted variables: x_i is correlated with u_i.
   The true regression is: $y_i = x_i'\beta + w_i'\delta + \varepsilon_i$, $\mathbb{E}[x_i\varepsilon_i] \neq 0$, $\mathbb{E}[w_i\varepsilon_i] = 0$.

We run: $y_i = x_i'\beta + u_i$, then

$$\begin{aligned}\mathbb{E}[x_i u_i] &= \mathbb{E}[x_i(w_i'\delta + \varepsilon_i)] \\ &= \mathbb{E}[x_i w_i']\delta + \underbrace{\mathbb{E}[x_i\varepsilon_i]}_{0}\end{aligned}$$

For our general regression model $y_i = x_i'\beta + u_i$, we have $\mathbb{E}[x_i u_i] \neq 0$, thus $\hat{\beta}_{OLS} \xrightarrow{p} \beta$ doesn't hold. We take $z_i \in \mathbb{R}^r$, which is a good IV if:

1. Relevance: $\mathbb{E}[z_i x_i] \neq 0$;
2. Exogeneity: $\mathbb{E}[z_i u_i] = 0$.

Then, we have the 2SLS method:

Definition 1.2.1 (2SLS Method).

1. Estimate: $x_i = z_i'\gamma + e_i \Rightarrow \hat{\gamma} = (Z'Z)^{-1}Z'X \Rightarrow \hat{X} = Z'\hat{\gamma} = P_Z X$;
2. Estimate: $y_i = \hat{x}_i'\beta + u_i^*$.

$$\begin{aligned}\hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y \\ &= ((P_Z X)'P_Z X)^{-1}(P_Z X)'Y \\ &= (X'P_Z X)^{-1}X'P_Z Y \\ &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y \\ &= \beta + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'u \\ &\xrightarrow{p} \beta + Q_{xz}^{-1}\mathbb{E}[x_i z_i']\mathbb{E}[z_i z_i']\mathbb{E}[z_i u_i] \\ &= \beta.\end{aligned}$$

$$\begin{aligned}\mathbb{V}[\hat{\beta}_{2SLS}|X,Z] &= \mathbb{V}\left[(X'P_Z X)^{-1}X'P_Z U|X,Z\right] \\ &= (X'P_Z X)^{-1}\,\mathbb{V}\left[X'P_Z U|X,Z\right](X'P_Z X)^{-1} \\ &= (X'P_Z X)^{-1}X'P_Z\mathbb{V}[U]P_Z X\,(X'P_Z X)^{-1} \\ &= (X'P_Z X)^{-1}\sigma^2\end{aligned}$$

As we know $\mathbb{V}[\hat{\beta}_{OLS}] = (X'X)^{-1}\sigma^2$,

$$\begin{aligned}\mathbb{V}\left[\hat{\beta}_{OLS}\right]^{-1} - \mathbb{V}\left[\hat{\beta}_{2SLS}\right]^{-1} &= (\sigma^2)^{-1}X'X - (\sigma^2)^{-1}X'P_Z X \\ &= (\sigma^2)^{-1}X'(I - P_Z)X \\ &= (\sigma^{-2})X'M_Z X \\ &= \sigma^{-2}(\underbrace{M_Z X}_{\hat{E}})'M_Z X \\ &= \sigma^{-2}SSR_{1SLS}.\end{aligned}$$

Theorem 1.2.1 (Anderson-Rubin Method).
$y_i = x_i'\beta_0 + u_i$, $\mathbb{E}[z_i u_i] = 0$, $y_i - x_i'\beta = \delta z_i + v_i$. $\Rightarrow \hat{\delta}(\beta) = (Z'Z)^{-1}Z'(Y - X\beta) \to \hat{\delta}(\beta_0) =$

$(Z'Z)^{-1}Z'U$. For many βs, test: $H_0 : \delta(\beta) = 0$, e.g. using t-test.

$$T_t = \frac{\hat{\delta}(\beta)}{se(\hat{\delta}(\beta))} \sim \mathbf{N}(0,1)$$

The 90% CI for β is the set of βs at which $\delta(\beta) = 0$ cannot be rejected at 90% confidence level.

Lecture 2.

Causal Inference

2.1 Potential Outcomes Framework

Definition 2.1.1 (Stable Unit Treatment Value Assumption (SUTVA)).

$$y_i = \begin{cases} y_{0i} & d_i = 0 \\ y_{1i} & d_i = 1 \end{cases}$$

Causal effect of d_i on y_i for individual i: $y_{1i} - y_{0i}$.

$$y_i = d_i y_{1i} + (1 - d_i) y_{0i}$$

SUTVA(2.1.1) ensures that the individual treatment effect is well defined.

For a population, we know that $\mathbb{E}[d_i], \mathbb{E}[y_i], \mathbb{E}[y_{0i}], \mathbb{E}[y_{1i}]$ exist, we can define the treatment conditional expectations:

$$\mathbb{E}[y_i|d_i = 1], \mathbb{E}[y_{0i}|d_i = 1], \mathbb{E}[y_{1i}|d_i = 1] = \mathbb{E}[y_i|d_i = 1]$$

that denote the averages of the outcome y_i.

Analogously, we can define the control conditional expectations:

$$\mathbb{E}[y_i|d_i = 0], \mathbb{E}[y_{0i}|d_i = 0] = \mathbb{E}[y_i|d_i = 0], \mathbb{E}[y_{1i}|d_i = 0]$$

for the non-treated subpopulation.

Then, we candefine the Average Treatment Effect (ATE), the Average Treatment Effect for the Treatment-Group (ATT) and the Average Treatment Effect for the Control-Group (ATC) as distinct objects:

$$\begin{aligned} \text{ATE} &= \mathbb{E}[y_{1i} - y_{0i}] \\ \text{ATT} &= \mathbb{E}[y_{1i} - y_{0i}|d_i = 1] \\ \text{ATC} &= \mathbb{E}[y_{1i} - y_{0i}|d_i = 0] \end{aligned}$$

$$\mathbb{E}[z] = \mathbb{E}[z|d = 1]\mathbb{P}[d = 1] + \mathbb{E}[z|d = 0]\mathbb{P}[d = 0] = \mathbb{E}[\mathbb{E}[z|d]].$$

For sample, $\{d_i, y_i\}_{i=1}^n = \{d_i, y_{d_i,i}\}_{i=1}^n$, because $y_i = y_{1i}d_i + y_{0i}(1 - d_i)$.

$N = \{i = 1, 2, \cdots, n\}$, $N_1 = \{i \in N : d_i = 1\} \leftarrow n_1 = |N_1|$, $N_0 = \{i : d_i = 0\} \leftarrow n_0 = |N_0|$.

$$\begin{aligned} \frac{1}{n_1}\sum_{i\in N_1} y_i &= \frac{1}{n_1}\sum_{i\in N_1} y_{1i} \xrightarrow{p} \mathbb{E}[y_{1i}|d_i = 1] = \mathbb{E}[y_i|d_i = 1] \\ \frac{1}{n_0}\sum_{i\in N_0} y_i &= \frac{1}{n_0}\sum_{i\in N_0} y_{0i} \xrightarrow{p} \mathbb{E}[y_{0i}|d_i = 0] = \mathbb{E}[y_i|d_i = 0] \end{aligned}$$

$$\frac{1}{n_1}\sum_{i\in N_1} y_i - \frac{1}{n_0}\sum_{i\in N_0} y_i \xrightarrow{p} \mathbb{E}[y_{1i}|d_i = 1] - \mathbb{E}[y_{0i}|d_i = 0] = \text{ATE} = \text{ATT} = \text{ATC}.$$

We define teh difference of treated and non-treated as: *Naive Difference*.

$$\begin{aligned}\text{ND} &= \mathbb{E}[y_{1i}|d_i = 1] - \mathbb{E}[y_{0i}|d_i = 0] \\ &= \mathbb{E}[y_{1i}|d_i = 1] - \mathbb{E}[y_{0i}|d_i = 1] + \mathbb{E}[y_{0i}|d_i = 1] - \mathbb{E}[y_{0i}|d_i = 0] \\ &= ATT + \mathbb{E}[y_{0i}|d_i = 1] - \mathbb{E}[y_{0i}|d_i = 0]\end{aligned}$$

For LRM, $y_i = \beta_0 + \beta_1 d_i + u_i$,

$$\begin{aligned}\text{ND} &= \mathbb{E}[y_i|d_i = 1] - \mathbb{E}[y_i|d_i = 0] \\ &= \mathbb{E}[\beta_0 + \beta_1 + u_i|d_i = 1] - \mathbb{E}[\beta_0 + u_i|d_i = 0] \\ &= \beta_1 + \mathbb{E}[u_i|d_i = 1] - \mathbb{E}[u_i|d_i = 0]\end{aligned}$$

$$\{Y_d\} \perp\!\!\!\perp D \mid X \Rightarrow \{Y_d\} \perp\!\!\!\perp D \mid \pi(X), \quad D \perp\!\!\!\perp X \mid \pi(X)$$

Lecture 3.

Panel Data Analysis

3.1 Incidental Parameters Problem

3.1.1 Consistency

Suppose we are estimating the following panel data regression:

$$y_{it} = \alpha + x'_{it}\beta + u_{it}, \quad \mathbb{E}[u_{it}x_{it}] = 0, \quad \mathbb{V}[u_{it}|x_{it}] = \sigma^2$$

Omitting the distinction between intercept and slope, we can write the model as:

$$\begin{aligned} y_{it} &= \tilde{x}'_{it}\tilde{\beta} + u_{it} \\ \tilde{x}_{it} &= \begin{bmatrix} 1 \\ x_{it} \end{bmatrix} \\ \tilde{\beta} &= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \end{aligned}$$

where $i = 1 : n$, $T = 1 : t$.

Or, we can write the model as:

$$\underset{T\times 1}{y_i} = \underset{T\times K}{\tilde{X}_i}\underset{K\times 1}{\tilde{\beta}} + \underset{T\times 1}{u_i}$$

Using OLS method to estimate $\tilde{\beta}$, we have:

$$\min_{\tilde{\beta}} \sum_i \sum_t u_{it}^2 = \min_{\beta} \sum_i u'_i u_i = \min_{\beta} (y_i - \tilde{X}_i\tilde{\beta})'(y_i - \tilde{X}_i\tilde{\beta})$$

The FOC of this equation is:

$$\begin{aligned} \sum_i -\tilde{X}'_i(y_i - \tilde{X}_i\tilde{\beta}) = 0 \\ \left(\sum_i \tilde{X}'_i\tilde{X}_i\right)\tilde{\beta} = \sum_i \tilde{X}'_i y_i \\ \hat{\tilde{\beta}} &= \left(\sum_i \tilde{X}'_i\tilde{X}_i\right)^{-1} \sum_i \tilde{X}'_i y_i \\ &= \left(\sum_i\sum_t \tilde{x}_{it}\tilde{x}'_{it}\right)^{-1}\left(\sum_i\sum_t \tilde{x}_{it}y_{it}\right) \\ &= \tilde{\beta} + \left(\frac{1}{n}\sum_i\sum_t \tilde{x}_{it}\tilde{x}'_{it}\right)^{-1}\left(\sum_i\sum_t \tilde{x}_{it}u_{it}\right) \\ &\xrightarrow{p} \tilde{\beta} + \mathbb{E}\left[\sum_t \tilde{x}_{it}\tilde{x}'_{it}\right]\mathbb{E}\left[\sum_t \tilde{x}_{it}u_{it}\right] \\ &= \tilde{\beta} \end{aligned}$$

Hence $\hat{\beta_{OLS}}$ is consistent provided that x_{it} and u_{it} are contemperaneously uncorrelated, as $\mathbb{E}[x_{it}u_{it}] = 0$.

3.1.2 Asymptotic Normality

From the analysis of consistency, we know that:

$$\hat{\tilde{\beta}} = \left(\sum_i \tilde{X}_i' \tilde{X}_i\right)^{-1} \sum_i \tilde{X}_i' y_i$$

Hence:

$$\begin{aligned}\sqrt{n}(\hat{\tilde{\beta}} - \tilde{\beta}) &= \left(\frac{1}{n}\sum_i \tilde{X}_i'\tilde{X}_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i \tilde{X}_i' u_i\right) \\ &\xrightarrow{p} \mathbb{E}[\tilde{X}_i'\tilde{X}_i]^{-1} \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\left(\tilde{X}_i'u_i\right)\left(\tilde{X}_i'u_i\right)'\right]\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\tilde{X}_i'\tilde{X}_i\right]^{-1}\mathbb{E}\left[\tilde{X}_i'u_iu_i'\tilde{X}_i\right]\mathbb{E}\left[\tilde{X}_i'\tilde{X}_i\right]\right)\end{aligned}$$

The above model is honmogeneous, which is unattractive, as the data generating process would differ across i, with some units having a higher level of the outcome variable y_{it} than others, regardless of covariates x_{it}(with a higher intercept α) or a stronger effect of some covariates $x_{it,k}$ on y_{it} than others.

At the other extreme, we assume the fulle heterogenous estimation:

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}, \quad \mathbb{E}[u_{it}x_{it}] = 0, \quad \mathbb{V}[u_{it}|x_{it}] = \sigma_i^2.$$

Under $T = 1$, we run $y_i = \beta_0 + x_i'\beta + v_i$, where $v_i = u_i + \underbrace{\alpha_i - \beta_0}_{\tilde{\alpha}_i}$ and $\mathbb{E}[v_i] = 0$.

Under $T > 1$, we run:

$$\begin{aligned}y_i &= x_i'\beta + \sum_{j=1}^{n} \alpha_j \mathbf{1}\{i = j\} + u_{it} \\ &= \tilde{x}_{it}'\tilde{\beta} + u_{it}\\ \tilde{x}_{it} &= \begin{bmatrix} x_{it} \\ \mathbf{1}\{i=1\} \\ \mathbf{1}\{i=2\} \\ \vdots \\ \mathbf{1}\{i=n\}\end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix}\beta \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n\end{bmatrix}\end{aligned}$$

In a similar way, we can write the regression as

$$y_i = \tilde{X}_i\tilde{\beta}_i + u_i$$

with $\tilde{\beta}_i$ is specific for each i. We have n separate time series regressions, one for each unit i.

Following the same analyzing process, we can get:

$$\hat{\tilde{\beta}}_{i,OLS} = \left(\sum_i \tilde{X}_i'\tilde{X}_i\right)\sum_i \tilde{X}_i' y_i = \left(\sum_t \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\sum_t \tilde{x}_{it}y_{it}\right),$$

which obviously shows that $\hat{\tilde{\beta}}$ is consistent $\Leftrightarrow T \to \infty$.

3.1.3 One-way error component model

With the fully homogeneous specification unattractive and the fully heterogeneous specifi- cation infeasible, researchers usually go for a compromise and let intercepts (and error term variances) be unit-specific.

Definition 3.1.1 (One-way error component model).

$$y_{it} = \alpha_i + x_{it}' + u_{it}, \quad \mathbb{E}[u_{it}x_{it}] = 0, \quad \mathbb{V}[u_{it}|x_{it}] = \sigma^2, \tag{3.1}$$

where α_i is an individual-specific effect, and u_{it} are idiosyncratic(i.i.d.) errors.

In any case, the equation above makes clear that α_i contains all factors that affect y_{it}, that are not included in x_{it} and that are fixed over time (the time-varying factors are in u_{it}).

Suppose hte model is correctly specified, and we have a cross-sectional detaset available, i.e. $T = 1$. Then, we would estimate:

$$y_{it} = \beta_0 + x_{it}' + v_i, \text{ for} t = 1,$$

where $v_i = \alpha_i + u_{it} - \beta_0$.

If the unobserved heterogeneity α_i is correlated with the covariate x_{it}, our standard OLS estimator is biased and inconsistent.

If we have a panel dataset, i.e. $T > 1$, we can write the above model into a regression of $k+n$ regressors:

$$y_{it} = x_{it}'\beta + \sum_{j=1}^{n} \mathbf{1}\{i = j\}\alpha_j + u_{it} = x_{it}^{*'}\beta^* + u_{it},$$

where $x_{it}^* = \Big(x_{it}', \mathbf{1}\{i = 1\}, \cdots, \mathbf{1}\{i = n\}\Big)'$, and $\beta^* = \Big(\beta', \alpha_1, \cdots, \alpha_n\Big)'$.
This leads to the pooled OLS estimator for β^*:

$$\hat{\beta}^* = \Big(\sum_i \sum_t x_{it}^* x_{it}^{*'}\Big) \sum_i \sum_t x_{it}^* y_{it}.$$

HOwever, the estimator suffers from the so-called **IPP problem**, as the number of parameters increase with $n \to \infty$, the limit of $\frac{1}{n}\sum_i x_{it}^* x_{it}^{*'}$ is not well-defined and as a result, we can't establish consistency of $\hat{\beta}_{OLS}$.

3.2 Random Effects

By defining $v_{it} = u_{it} + \alpha_i - \beta_0$, we can transform the random effect model to the following:

$$\begin{aligned} y_{it} &= \alpha_1 + x_{it}'\beta + u_{it} \\ &= \underbrace{\beta_0 + x_{it}'\beta}_{\tilde{x}_{it}'} + \underbrace{u_{it} + \alpha_i - \beta_0}_{\equiv v_{it}} \end{aligned}$$

Defining again $\tilde{x}_{it} = (1, x_{it}')'$, $\tilde{\beta} = (\beta_0, \beta')'$, we can rewrite the model as:

$$\begin{aligned} y_{it} &= \tilde{x}_{it}'\beta + v_{it} \Leftrightarrow y_i = \tilde{X}_i'\tilde{\beta} + v_i \\ \to \hat{\tilde{\beta}} &= \left(\sum_i \tilde{X}_i'\tilde{X}_i\right)^{-1} \sum_i \tilde{X}_i' y_i \end{aligned}$$

With this intercept $beta_0$, $\mathbb{E}[v_i] = 0$ is guaranteed to hold. Define $\tilde{\alpha}_i = \alpha_i - \beta_0$ as the mean-zero unit-specific heterogenety so that $v_i = u_i + \tilde{\alpha}_i$.

Note (POLS).

Homogenous spec: $y_{it} = \alpha + x'_{it}\beta + u_{it} = \tilde{x}'_{it}\tilde{\beta} + u_{it}$. $\hat{\tilde{\beta}}$ is consistent if $\mathbb{E}[v_{it}x_{it}] = 0, \forall t$.

Using pooled OLS to estimate $\hat{\tilde{\beta}}$,

$$\begin{aligned}
\hat{\tilde{\beta}}_{RE-OLS/POLS} &= \left(\frac{1}{n}\sum_i \tilde{X}'_i\tilde{X}_i\right)^{-1}\frac{1}{n}\sum_i \tilde{X}'_i y_i \\
&= \tilde{\beta} + \left(\frac{1}{n}\sum_i \tilde{X}'_i\tilde{X}_i\right)^{-1}\frac{1}{n}\sum_i \tilde{X}'_i v_i \\
&\xrightarrow{p} \tilde{\beta} + \mathbb{E}[\tilde{X}'_i\tilde{X}_i]^{-1}\mathbb{E}[\tilde{X}'_i v_i] \\
\text{where} \quad \mathbb{E}[\tilde{X}'_i v_i] &= \mathbb{E}\left[\sum_t \tilde{x}'_{it} v_{it}\right] \\
&= \sum_t \mathbb{E}\left[\tilde{x}'_{it} v_{it}\right] \\
&= \sum_t \mathbb{E}\left[\tilde{x}_{it}(u_{it} + \alpha_i - \beta_0)\right]
\end{aligned}$$

Here, the error term v_i is not equal to the original error term u_{it}.

Note.

Under the random effect, you have to use the heteroskedasticity-robust methods. Because even if we assume u_{it} to be homoskedastic, v_{it} is not, as it includes also the unit-specific heterogeneity α_i.

So, to obtain consistency, we need to assume that:

- $\mathbb{E}[u_{it}|\tilde{x}_{it}, \tilde{\alpha}_i] = 0, \forall t$.
- $\mathbb{E}[\tilde{\alpha}_i|\tilde{x}_{it}] = 0, \forall t$.

And, we are also obliged to use HAC-robust standard error because:

$$\Omega \equiv \mathbb{E}[v_i v'_i|\tilde{X}_i] = \mathbb{E}[(\alpha_i \mathbf{1}_i + u_i)(\tilde{\alpha}_i\mathbf{1}_i + u_i)'|\tilde{X}_i] = \mathbb{E}[\tilde{\alpha}_i^2\mathbf{1}_i\mathbf{1}'_i|\tilde{X}_i] + \mathbb{E}[u_i u'_i|\tilde{X}_i]$$

is not diagonal.

Given the error structure the natural estimator for β is GLS. The GLS eimator for β is:

$$\begin{aligned}
\hat{\tilde{\beta}}_{RE-GLS} &= \left(\sum_i \tilde{X}'_i\Omega^{-1}\tilde{X}_i\right)^{-1}\sum_i \tilde{X}'_i\Omega^{-1}y_i \\
\Omega^{-\frac{1}{2}}y_i &= \Omega^{-\frac{1}{2}}\tilde{X}'_i\tilde{\beta} + \Omega^{-\frac{1}{2}}v_i \\
\Omega = \mathbb{E}[v_i v'_i|\tilde{X}_i] &= \mathbb{E}\left[\begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT}\end{bmatrix}\begin{bmatrix} v_{i1} & v_{i2} & \cdots & v_{iT}\end{bmatrix}|\tilde{X}_i\right] \\
&= \mathbb{E}\begin{bmatrix} \mathbb{E}[v_{i1}^2|\tilde{X}_i] & \mathbb{E}[v_{i1}v_{i2}|\tilde{X}_i] & \cdots & \mathbb{E}[v_{i1}v_{iT}|\tilde{X}_i] \\ \mathbb{E}[v_{i2}v_{i1}|\tilde{X}_i] & \mathbb{E}[v_{i2}^2|\tilde{X}_i] & \cdots & \mathbb{E}[v_{i2}v_{iT}|\tilde{X}_i] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[v_{iT}v_{i1}|\tilde{X}_i] & \mathbb{E}[v_{iT}v_{i2}|\tilde{X}_i] & \cdots & \mathbb{E}[v_{iT}^2|\tilde{X}_i] \end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix}
\mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{i1}^2|\tilde{X}_i] & \mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{i1}u_{i2}|\tilde{X}_i] & \cdots & \mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{i1}u_{iT}|\tilde{X}_i] \\
\mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{i2}u_{i1}|\tilde{X}_i] & \mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{i2}^2|\tilde{X}_i] & \cdots & \mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{i2}u_{iT}|\tilde{X}_i] \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{iT}u_{i1}|\tilde{X}_i] & \mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{iT}u_{i2}|\tilde{X}_i] & \cdots & \mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right] + \mathbb{E}[u_{iT}^2|\tilde{X}_i]
\end{bmatrix} \\
&= \mathbb{E}\left[\mathbb{E}\left[\alpha_i^2|\tilde{X}_i\right]\mathbf{1}\right] + \mathbb{E}[u_i u_i'|\tilde{X}_i] \\
&= \mathbb{E}[\alpha_i^2]\mathbf{1}_i\mathbf{1}_i' + \begin{bmatrix}
\mathbb{E}[u_{i1}^2|\tilde{X}_i] & 0 & \cdots & 0 \\
0 & \mathbb{E}[u_{i2}^2|\tilde{X}_i] & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \mathbb{E}[u_{iT}^2|\tilde{X}_i]
\end{bmatrix} \\
&= \sigma_{\alpha_i}^2 \mathbf{1}_i' +_i^2 I \\
&= \sigma_\alpha^2 \mathbf{1}_i\mathbf{1}_i' + \sigma^2 I \\
\text{beacuse } & \mathbb{V}[\tilde{\alpha}_i|\tilde{X}_i] = \sigma_{\alpha_i}^2 = \sigma_\alpha^2 \\
& \mathbb{V}[u_{it}|\tilde{X}_i] = \sigma_i^2 = \sigma^2, \forall i.
\end{aligned}
$$

Under the assumption $\mathbb{E}[u_{it}x_{it}] = 0$, we now describe some statistical properties of $\hat{\tilde{\beta}}_{RE-GLS}$.

$$
\begin{aligned}
\hat{\tilde{\beta}}_{RE-GLS} - \tilde{\beta} &= \Big(\sum_i \tilde{X}_i'\Omega^{-1}\tilde{X}_i\Big)^{-1}\Big(\sum_i \tilde{X}_i'\Omega^{-1}v_i\Big) \\
&\to \mathbb{E}\Big[\sum_i \tilde{X}_i'\Omega^{-1}\tilde{X}_i\Big]\mathbb{E}\Big[\sum_i \tilde{X}_i'\Omega^{-1}v_i\Big] \\
\text{where } \mathbb{E}\Big[\sum_i \tilde{X}_i'\Omega^{-1}v_i\Big] &= \sum_i \mathbb{E}\Big[\tilde{X}_i'\Omega^{-1}v_i\Big] \\
&= \sum_i \tilde{X}_i'\Omega^{-1}\mathbb{E}[v_i|\tilde{X}_i] \\
&= \sum_i \tilde{X}_i'\Omega^{-1}\mathbb{E}[u_i + \tilde{\alpha}_i|\tilde{X}_i] \\
&= 0
\end{aligned}
$$

Thus, $\hat{\tilde{\beta}}_{RE-GLS}$ is conditionally unbiased for $\tilde{\beta}$. The asymptotic variance of $\hat{\tilde{\beta}}_{RE-GLS}$ is:

$$
\begin{aligned}
\sqrt{n}\left(\hat{\tilde{\beta}}_{RE-GLS} - \tilde{\beta}\right) &\xrightarrow{d} \mathcal{N}(0, V) \\
\text{where } V &= \mathbb{E}\left[\tilde{X}_i'\Omega^{-1}\tilde{X}_i\right]^{-1}\mathbb{E}\left[\tilde{X}_i'\Omega^{-1}v_i v_i'\Omega^{-1}\tilde{X}_i\right]\mathbb{E}\left[\tilde{X}_i'\Omega^{-1}\tilde{X}_i\right] \\
&= \mathbb{E}\left[\tilde{X}_i'\Omega^{-1}\tilde{X}_i\underbrace{\mathbb{E}[v_i v_i'|\tilde{X}_i]}_{\equiv\Omega}\right]
\end{aligned}
$$

Because we do not know Ω, the RE-GLS estimator is infeasible.

If indeed we have:

$$
\begin{aligned}
\Omega &= \mathbb{E}[v_i v_i'|\tilde{X}_i] \\
&= \mathbb{E}[(\alpha_i\mathbf{1}_i + u_i)(\alpha_i\mathbf{1}_i + u_i)'|\tilde{X}_i] \\
&= \mathbb{E}[\alpha_i^2]\mathbf{1}_i\mathbf{1}_i' + \mathbb{E}[u_i u_i'|\tilde{X}_i]
\end{aligned}
$$

which implies homoskedasticity.

A feasible version replaces Ω with an estimator $\hat{\Omega}_i$. Assuming homoskedasticity of the original errors:

$$\mathbb{E}[u_i u_i' | \tilde{X}_i, \tilde{\alpha}_i] = \sigma_u^2 I_T$$
$$\mathbb{E}[\tilde{\alpha}_i^2 | \tilde{x}_i] = \sigma_\alpha^2$$

We obtain: $\Omega = \sigma_\alpha^2 \mathbf{1}_i \mathbf{1}_i' + \sigma_u^2 I_T$.

Hence, the motivation for using GLS is different than under a cross-sectional regression with heteroskedasticity. We use GLS because of the autocorrelation in v_{it} induced by the presence of time variant α_i.

3.3 Fixed Effects

In the econometrics literature if the stochastic structure of α_i is treated as unknown and possibly correlated with x_{it}, then α_i is called a **fixed effect**.

Correlation between α_i and x_{it} will cause both pooled and random effect estimators ro be biased.

We transform equation to get rid of α_i: $y_{it} = \alpha_i + x_{it}'\beta + u_{it}$. This is due to the classic problems of omitted variables bias and endogeneity.

The presence of the unstructured individual effect α_i means that it is not possible to identify β under a simple projection assumption such as $\mathbb{E}[u_{it} x_{it}] = 0$. It turns out that a sufficient condition for identification is the following.

Definition 3.3.1 (Strictly exogenous).

A regressor x_{it} is said to be strictly exogenous if $\mathbb{E}[x_{it} u_{is}] = 0, \forall t, s = 1, \cdots, T$.

3.3.1 Within Transformation

If we leave the relationship between α_i and x_{it} fully unstructured, then the only way to consistently estimate the coefficient β is by an estimator which is invariant to α_i.

Define the mean of a variable for a given individual as

$$\overline{y}_i = \frac{1}{T} \sum_t y_{it}$$
$$\overline{x}_i = \frac{1}{T} \sum_t x_{it}$$
$$\overline{u}_i = \frac{1}{T} \sum_t u_{it}$$

Then,

$$(y_{it} - \overline{y}_i) = (x_{it} - \overline{x}_i)'\beta + (u_{it} - \overline{u}_i)$$
$$\ddot{y}_{it} = \ddot{x}_{it}'\beta + \ddot{u}_{it}$$

Denote the time-averages method by $\hat{\beta}_{FE-W}$, the fixed effect estimator is consistent and asymptotically normal.

$$\hat{\beta}_{FE-W} = \left(\sum_i \sum_t \ddot{x}_{it} \ddot{x}_{it}' \right)^{-1} \sum_i \sum_t \ddot{x}_{it} \ddot{y}_{it}$$
$$= \beta + \left(\sum_i \sum_t \ddot{x}_{it} \ddot{x}_{it}' \right)^{-1} \sum_i \sum_t \ddot{x}_{it} \ddot{u}_{it}$$

$$
\begin{aligned}
&\xrightarrow{p} \beta + \mathbb{E}\left[\sum_t \ddot{x}_{it}\ddot{x}_{it}'\right]^{-1} \mathbb{E}\left[\sum_t \ddot{x}_{it}\ddot{u}_{it}\right] \\
\text{where } \mathbb{E}\left[\sum_t \ddot{x}_{it}\ddot{u}_{it}\right] &= \sum_t \mathbb{E}\left[\ddot{x}_{it}\ddot{u}_{it}\right] \\
\mathbb{E}\left[\ddot{x}_{it}\ddot{u}_{it}\right] &= \mathbb{E}\left[\left(x_{it} - \frac{1}{T}\sum_t x_{it}\right)\left(u_{it} - \frac{1}{T}\sum_t u_{it}\right)'\right] \\
&= 0 \quad \text{if } u_{it} \perp\!\!\!\perp x_{is}, \forall t, s = 1, \cdots, T.
\end{aligned}
$$

3.3.2 First Difference Transformation

$$
\begin{aligned}
y_{it} - y_{i,t-1} &= (x_{it} - x_{i,t-1})'\beta + (u_{it} - u_{i,t-1}) \\
\Delta y_{it} &= \Delta x_{it}'\beta + \Delta u_{it}, i = 1 \cdots n, t = 2 \cdots T.
\end{aligned}
$$

Denote the first difference method by $\hat{\beta}_{FE-FD}$, the fixed effect estimator is consistent and asymptotically normal.

$$
\begin{aligned}
\hat{\beta}_{FE-FD} &= \left(\sum_i \sum_t \Delta x_{it}\Delta x_{it}'\right)^{-1} \sum_i \sum_t \Delta x_{it}\Delta y_{it} \\
&= \beta + \left(\frac{1}{n}\sum_i \sum_t \Delta x_{it}\Delta x_{it}'\right)^{-1} \frac{1}{n}\sum_i \sum_t \Delta x_{it}\Delta u_{it} \\
&\xrightarrow{p} \beta + \mathbb{E}\left[\sum_t \Delta x_{it}\Delta x_{it}'\right]^{-1} \mathbb{E}\left[\sum_t \Delta x_{it}\Delta u_{it}\right] \\
\text{where } \mathbb{E}\left[\sum_t \Delta x_{it}\Delta u_{it}\right] &= \sum_t \mathbb{E}\left[\Delta x_{it}\Delta u_{it}\right] \\
\mathbb{E}\left[\Delta x_{it}\Delta u_{it}\right] &= \mathbb{E}\left[(x_{it} - x_{i,t-1})(u_{it} - u_{i,t-1})'\right] \\
&= 0 \quad \text{if } x_{it} \perp\!\!\!\perp (u_{it}, u_{i,t-1}), \forall t.
\end{aligned}
$$

Note.

The FD method is not as strong as the within method, because it only requires that the variable is uncorrelated with the error term in the same period and the previous period.

If there is a correlation between the error term in current period and two periods ago, there is a problem of feedback loop, which we will imply the correlated random effect model.

Take x_{it} fo which $\overline{x_i} = x_{it}, \forall i, t$.

Theorem 3.3.1 (Hausman-Test)**.**

$\mathcal{H}_0$: $\hat{\beta}_{RE,pop} = \hat{\beta}_{FE-W,pop} \Leftrightarrow$ We should use $\hat{\beta}_{RE}$.

We define:

$$
T_{Hausman} = n\Big(\hat{\beta}_{FE} - \hat{\beta}_{RE}\Big)'\Big(A\mathbb{V}[\hat{\beta}_{FE}] - A\mathbb{V}[\hat{\beta}_{RE}]\Big)^{-1}\Big(\hat{\beta}_{FE} - \hat{\beta}_{RE}\Big) \to \chi_k^2
$$

Note.

To sum up, the FE estimators work under arbitrary correlation between the unobserved heterogeneity α_i and covariates X_i , but they cannot deal with time-constant regressors and their consistency is paid for by an efficiency loss relative to RE estimators.

Most importantly, their consistency requires strict exogeneity, a much stronger assumption than contemporaneous exogeneity of covariates and error terms.

3.3.3 FE-IV Estimation

1. Contemperaneous exogeneity: $\mathbb{E}[x_{it}u_{it}] = 0, \forall t$.
2. Strict exogeneity: $\mathbb{E}[x_{it}u_{is}] = 0, \forall t, s$.
3. Sequential exogeneity: $\mathbb{E}[x_{it}u_{is}] = 0, \forall t, s \geq t$.

Definition 3.3.2 (Predetermined variables(Or Sequantial Exogeneity)).

Predetermined variables are variables that were determined prior to the current period. In econometric models this implies that the current period error term is uncorrelated with current and lagged values of the predetermined variable but may be correlated with future values. This is a weaker restriction than strict exogeneity, which requires the variable to be uncorrelated with past, present, and future shocks.

Still assume that we have a standard model:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta + u_{it} \\ &= \alpha_i + \beta_1 y_{i,t-1} + \tilde{x}_{it}'\beta_{-1} + u_{it} \\ \Rightarrow \Delta y_{it} &= \Delta\alpha_i + \Delta x_{it}'\beta + \Delta u_{it} \end{aligned}$$

Definition 3.3.3 (Anderson and Hsiao(1981)).

FE-IV: Use $y_{i,t-2}$ as the IV for $\Delta y_{i,t-1}$.

Under sequential exogeneity, instrument-exogeneity is satisied:

$$\mathbb{E}[y_{is}\Delta u_{it}] = 0, \forall s \leq t-2.$$

Using similar reasoning, other approaches use sequential exogeneity to circumvent FE meth- ods altogether rather than to save their consistency. For example, Blundell and Bond (1998) start from the original specification:

$$y_{it} = x_{it}'\beta + \alpha_i + u_{it},$$

where correlation between α_i an x_{it} is suspected to be due to $y_{i,t-1}$, contained in x_{it}.

Definition 3.3.4 (Blundell and Bond(1998)).

$$\begin{aligned} y_{it} &= \alpha_i + \beta_1 y_{i,t-1} + u_{it} \\ &= \beta_1 y_{i,t-1} + (u_{it} + \alpha_i) \end{aligned}$$

17.38 Anderson-Hsiao Estimator

Anderson and Hsiao (1982) made an important breakthrough by showing that a simple instrumental variables estimator is consistent for the parameters of (17.81).

The method first eliminates the individual effect u_i by first-differencing (17.81) for $t \geq p+1$

$$\Delta Y_{it} = \alpha_1 \Delta Y_{i,t-1} + \alpha_2 \Delta Y_{i,t-2} + \cdots + \alpha_p \Delta Y_{i,t-p} + \Delta X_{it}' \beta + \Delta \varepsilon_{it}. \tag{17.87}$$

This eliminates the individual effect u_i. The challenge is that first-differencing induces correlation between ΔY_{it-1} and $\Delta \varepsilon_{it}$:

$$\mathbb{E}\left[\Delta Y_{i,t-1} \Delta \varepsilon_{it}\right] = \mathbb{E}\left[\left(Y_{i,t-1} - Y_{i,t-2}\right)\left(\varepsilon_{it} - \varepsilon_{it-1}\right)\right] = -\sigma_\varepsilon^2.$$

The other regressors are not correlated with $\Delta \varepsilon_{it}$. For $s > 1$, $\mathbb{E}\left[\Delta Y_{it-s} \Delta \varepsilon_{it}\right] = 0$, and when X_{it} is strictly exogenous $\mathbb{E}\left[\Delta X_{it} \Delta \varepsilon_{it}\right] = 0$.

The correlation between ΔY_{it-1} and $\Delta \varepsilon_{it}$ is endogeneity. One solution to endogeneity is to use an instrument. Anderson-Hsiao pointed out that Y_{it-2} is a valid instrument because it is correlated with $\Delta Y_{i,t-1}$ yet uncorrelated with $\Delta \varepsilon_{it}$.

$$\mathbb{E}\left[Y_{i,t-2} \Delta \varepsilon_{it}\right] = \mathbb{E}\left[Y_{i,t-2} \varepsilon_{it}\right] - \mathbb{E}\left[Y_{i,t-2} \varepsilon_{it-1}\right] = 0. \tag{17.88}$$

The Anderson-Hsiao estimator is IV using $Y_{i,t-2}$ as an instrument for $\Delta Y_{i,t-1}$. Equivalently, this is IV using the instruments $(Y_{i,t-2}, ..., Y_{i,t-p-1})$ for $(\Delta Y_{i,t-1}, ..., \Delta Y_{i,t-p})$. The estimator requires $T \geq p+2$.

To show that this estimator is consistent, for simplicity assume we have a balanced panel with $T = 3$, $p = 1$, and no regressors. In this case the Anderson-Hsiao IV estimator is

$$\widehat{\alpha}_{\mathrm{iv}} = \left(\sum_{i=1}^N Y_{i1} \Delta Y_{i2}\right)^{-1} \left(\sum_{i=1}^N Y_{i1} \Delta Y_{i3}\right) = \alpha + \left(\sum_{i=1}^N Y_{i1} \Delta Y_{i2}\right)^{-1} \left(\sum_{i=1}^N Y_{i1} \Delta \varepsilon_{i3}\right).$$

Under the assumption that ε_{it} is serially uncorrelated, (17.88) shows that $\mathbb{E}\left[Y_{i1} \Delta \varepsilon_{i3}\right] = 0$. In general, $\mathbb{E}\left[Y_{i1} \Delta Y_{i2}\right] \neq 0$. As $N \to \infty$

$$\widehat{\alpha}_{\mathrm{iv}} \underset{p}{\longrightarrow} \alpha - \frac{\mathbb{E}\left[Y_{i1} \Delta \varepsilon_{i3}\right]}{\mathbb{E}\left[Y_{i1} \Delta Y_{i2}\right]} = \alpha.$$

Thus the IV estimator is consistent for α.

The Anderson-Hsiao IV estimator relies on two critical assumptions. First, the validity of the instrument (uncorrelatedness with the equation error) relies on the assumption that the dynamics are correctly specified so that ε_{it} is serially uncorrelated. For example, many applications use an AR(1). If instead the true model is an AR(2) then Y_{it-2} is not a valid instrument and the IV estimates will be biased. Second, the relevance of the instrument (correlatedness with the endogenous regressor) requires $\mathbb{E}\left[Y_{i1} \Delta Y_{i2}\right] \neq 0$. This turns out to be problematic and is explored further in Section 17.40. These considerations suggest that the validity and accuracy of the estimator are likely to be sensitive to these unknown features.

Figure 3.1: Anderson and Hsiao(1981)

Use $\Delta y_{i,t-1}$ as the IV for $y_{i,t-1}$

Lecture 4.

Time Series

4.1 Univariate Time Series

We have a sample $\{w_i\}_{i=1}^n$, with $w_i = (y_i, x_i')'$,

$\{w_{it}\}_{i=1:n,t=1:T}$.

Now, we look at $\{w_t\}_{i=1}^T$, usually written as y_t, is univariate time series data.

In the cross-sectional context, we average over i to get

$$\mathbb{E}[u_i] = \int u_i f_u(u_i) du_i.$$

Under tiem series data, we also think y_t as a RV. without i.i.d. assumption, we generallt have T realizations of different and mutually dependent variables.

$$\begin{aligned}
\mathbb{E}[y_t] &= \int y_t f_{y_t}(y_t) dy_t = \mu_t, \\
\mathbb{V}[y_t] &= \mathbb{E}\Big[(y_t - \mu_t)^2\Big] = \gamma_{0,t}, \\
\mathrm{Cov}(y_t, y_{t-h}) &= \mathbb{E}\Big[(y_t - \mu_t)(y_{t-h} - \mu_{t-h})\Big] = \gamma_{h,t}.
\end{aligned}$$

Definition 4.1.1 (Weak Stationarity).

y_t is a weakly stationary process if

1. $\mu_t = \mu$ for all t,
2. $\gamma_{h,t} = \gamma_h$ for all t.

autocovariance function (ACF): $\{\gamma_0, \gamma_1, \cdots\}$ autocorrelation function: $\{\rho_0, \rho_1, \cdots\}$, where $\rho_h = \frac{\gamma_h}{\gamma_0}$.

Appendix

Recommended Resources

Books

[1] James H. Stock and Mark W. Watson. *Introduction to Econometrics*. 4th ed. New York: Pearson, 2003

[2] Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. 7th ed. Cengage Learning, 2020

[3] Bruce E. Hansen. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2022

[4] Fumio Hayashi. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2000

[5] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Massachusetts: The MIT Press, 2010

[6] Joshua Chan et al. *Bayesian Econometric Methods*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2019

[7] Badi H. Baltagi. *Econometric Analysis of Panel Data*. 6th ed. Cham, Switzerland: Springer, 2021

[8] James D. Hamilton. *Time Series Analysis*. Princeton, New Jersey: Princeton University Press, 1994. ISBN: 9780691042893

[9] Takeshi Amemiya. *Advanced Econometrics*. Cambridge, MA: Harvard University Press, 1985

Others

[10] Roger Bowden. “The Theory of Parametric Identification”. In: *Econometrica* 41.6 (1973), pp. 1069–1074. DOI: 10.2307/1914036

[11] Robert I. Jennrich. “Asymptotic Properties of Non-linear Least Squares Estimators”. In: *The Annals of Mathematical Statistics* 40.2 (1969), pp. 633–643. DOI: 10.1214/aoms/1177697731

[12] Michael P. Keane. “A Note on Identification in the Multinomial Probit Model”. In: *Journal of Business & Economic Statistics* 10.2 (1992), pp. 193–200. DOI: 10.1080/07350015.1992.10509906

[13] Thomas J. Rothenberg. “Identification in Parametric Models”. In: *Econometrica* 39.3 (1971), pp. 577–591. DOI: 10.2307/1913267

[14] George Tauchen. “Diagnostic Testing and Evaluation of Maximum Likelihood Models”. In: *Journal of Econometrics* 30 (1985), pp. 415–443. DOI: 10.1016/0304-4076(85)90149-6

[15] Abraham Wald. “Note on the Consistency of the Maximum Likelihood Estimate”. In: *The Annals of Mathematical Statistics* 20.4 (1949), pp. 595–601. DOI: 10.1214/aoms/1177729952

[16] Halbert White. “Maximum Likelihood Estimation of Misspecified Models”. In: *Econometrica* 50.1 (1982), pp. 1–25. DOI: 10.2307/1912526