Lecture Notes:
Econometrics II

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

Draft updated on March 3, 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

$i = 1:n$, $t = 1:T$, z_{it}

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

Then,

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

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

The FOC of this equation is:

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

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:

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

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

Under $T > 1$, we run:

$$\begin{aligned}
y_i &= x_i'\beta + \sum_{j=1}^{n} \alpha_j \mathbf{1}\{i = j\} + u_{it} \\
&= \tilde{x}_{it}'\tilde{\beta} + u_{it}
\end{aligned}$$

$$\tilde{x}_{it} = \begin{bmatrix} x_{it} \\ \mathbf{1}\{i=1\} \\ \mathbf{1}\{i=2\} \\ \vdots \\ \mathbf{1}\{i=n\} \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} \beta \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

3.2 Random Effects

We put α_i in error terms

3.3 Fixed Effects

We transform equation to get rid of α_i.

3.4 Correlated Random Effects

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