Lecture Notes:
Econometrics II

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

Draft updated on April 2, 2025

These lecture notes were taken in the course *Econometrics II* taught by **Marko Mlikota** at Graduate Institute 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 assumptions

As we 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

We say that there's endogeneity in the linear model

$$y_i = x_i'\beta + u_i$$

if β is the parameter of interest and

$$\mathbb{E}[x_i u_i] \neq 0.$$

This is a core problem in econometrics and largely differentiates the field from statistics.

Endogeneity implies that the least squares estimator is inconsistent for the structural parameter. Indeed, under i.i.d. sampling, least squares is consistent for the projection coefficient.

$$\hat{\beta} \xrightarrow{p} \beta + \left(\mathbb{E}[XX']\right)^{-1}\mathbb{E}[XU] \neq \beta$$

The inconsistency of least squares is typically referred to as **endogeneity bias** or**estimation bias** due to endogeneity.

Commonly, there are three reasons for endogeneity:

1. Measurement error: x_i is measured with error.

Suppose our true Regression is: $y_i = x_i^{*\prime}\beta + \varepsilon_i$, $\mathbb{E}[x_i^* \varepsilon_i] = 0$, β is the structural parameter. But, $x_i^{*\prime}$ is not observed. Instead, we observe: $x_i = x_i^* + v_i$, where v_i is the measurement error, independent of x_i^* and ε_i: $\mathbb{E}[x_i^* v_i'] = 0$, $\mathbb{E}[v_i \varepsilon_i] = 0$ [1].

The model $x_i = x_i^* + v_i$ with x_i^* and v_i uncorrelated, and $\mathbb{E}[v_i] = 0$ is known as the **classical measurement error model**. This means that x_i is a noisy but unbiased estimate of x_i^*. By substitution we can express y_i as a function of the observed variable x_i.

$$y_i = x_i^{*\prime}\beta + \varepsilon_i = (x_i - v_i)'\beta + \varepsilon_i = x_i'\beta + u_i$$

where $u_i = \varepsilon_i - v_i'\beta$.

This means that (y_i, x_i) satisfy the linear equation $y_i = x_i'\beta + u_i$ with an error u_i. But this error is not a projection error.

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

if $\mathbb{E}[v_i v_i'] \neq 0$ and $\beta \neq 0$.

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: The most prominent cause of endogeneity are omitted variables(OVs).

   Suppose the true regression is: $y_i = x_i'\beta + w_i'\delta + \varepsilon_i$, where exogeneity holds: $\mathbb{E}[x_i\varepsilon_i] = 0$, $\mathbb{E}[w_i\varepsilon_i] = 0$.

   If we omit w_i and instead estimates:

$$y_i = x_i'\beta + u_i$$

   where $u_i = w_i'\delta + \varepsilon_i$, then in this misspecified model, exogeneity is only given if x_i and w_i are uncorrelated, since:

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

   Since $\hat{\beta} - \beta \xrightarrow{p} \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i u_i]$, we can assess the sign and size of the asymptotic bias based ont the signs of correlation between x_i and w_i.

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.

To consistently estimate β, we require additional assumptions. One type of information which is commonly used in economics is the **instruments**.

Definition 1.2.1 (Instrumental Variable).

We take $z_i \in \mathbb{R}^r$ as an instrumental variable if:

$$\begin{aligned}
\mathbb{E}[z_i u_i] &= 0 \\
\mathbb{E}[z_i x_i] &\neq 0
\end{aligned}$$

[1] This is an example of a latent variable model, where "latent" refers to an unobserved structural variable.

$$\mathbb{E}[z_i z_i'] > 0$$
$$\text{rank}\Big(\mathbb{E}[z_i x_i']\Big) = k \leq r \quad ^{2}$$

[2] We say that the model is just-identified if $k = r$ and over-identified if $k < r$.

1.2.1 Instrumental Variables and 2SLS

Then, we have the 2SLS method:

Definition 1.2.2 (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: $y_i = \hat{x}_i'\beta + u_i^*$.

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y \\
&= \left((P_Z X)'P_Z X\right)^{-1}(P_Z X)'Y \\
&= (X'P_Z X)^{-1}X'P_Z Y \\
&= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \beta + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'u \\
&= \beta + (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'u \quad ^{3} \\
&= \beta + (Z'X)^{-1}Z'u \\
&\xrightarrow{p} \beta.
\end{aligned}$$

[3] To compute $\hat{\beta}_{2SLS}$, we need $Z'Z$ to be full rank, which requires us to have more observations than IVs.

Ideally, z_i should be as highly correlated with x_i as possible, but uncorrelated with u_i. To see this, we find hte variance of $\hat{\beta}_{2SLS}$

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

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

$$\begin{aligned}
\mathbb{V}\left[\hat{\beta}_{OLS}\right]^{-1} - \mathbb{V}\left[\hat{\beta}_{2SLS}\right]^{-1} &= (\sigma^2)^{-1}X'X - (\sigma^2)^{-1}X'P_Z X \\
&= (\sigma^2)^{-1}X'(I - P_Z)X \\
&= (\sigma^{-2})X'M_Z X \\
&= \sigma^{-2}(\underbrace{M_Z X}_{\hat{E}})'M_Z X \\
&= \sigma^{-2}SSR_{1SLS} > 0.
\end{aligned}$$

This means that the variance of 2SLS estimator is larger than that of the OLS.

By the usual arguments, the asymptotic analysis reveals that:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{p} \mathcal{N}(0, V_{2SLS})$$

where

$$V_{2SLS} = Q_{XZ}^{-1} X'Z \left(Z'Z\right)^{-1} Z'UU'Z \left(Z'Z\right)^{-1} \left(X'Z\right)' Q_{XZ}^{-1}$$

where $Q_{XZ} = \left(Z'X\right) \left(Z'Z\right)^{-1} \left(X'Z\right)$

As usual, we can estimate it by replacing u_i with $\hat{u}_i$ and expectation operators with population means. Thereby, it's important to note that $u_i \neq u_i^*$, and to obtain $\hat{u}_i$, we don't use $\hat{x}_i$,but x_i:

$$\hat{u}_i = y_i - x_i'\hat{\beta}_{2SLS}$$

Under homoskedasticity, $V_{2SLS} = \sigma^2 Q_{XZ}^{-1}$, which we estimate using $\hat{\sigma}^2 = \frac{1}{n}\sum_i u_i^2$.

1.2.2 Weak Identification in IV Models

If the the correlation between x_i and z_i is weak, then we say it's a **weak instrument**. Under weak IVs, the finite sample distribution of $\hat{\beta}_{2SLS}$ may not assemble the asymptotic property.

In absence of an asymptotic distribution, we can conduct inference using its numerical approximation via bootstrapping. Or alternatively, we can construct a confidence set for β using the Following the procedure of Anderson and Rubin (1949).

The method is based on the idea that, for $\beta = \beta_0$, the auxiliary regression $y_i - x_i'\beta = \delta z_i + v_i$ should yield $\delta = 0$, because $y_i - x_i'\beta_0 = u_i$ and u_i is uncorrelated with z_i.

Theorem 1.2.1 (Anderson-Rubin Method).

For a given β_0, we get:

$$\sqrt{n}\hat{\delta}(\beta_0) = \sqrt{n}\left(Z'Z\right)^{-1} Z'(Y - X\beta_0) = \left(Z'Z\right)^{-1}\sqrt{n}Z'U \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_u^2}{\mathbb{E}(z_i^2)}\right)$$

which allows us to test $\mathcal{H}_0 : \delta = 0$. For many βs, test: $\mathcal{H}_0 : \delta(\beta) = 0$, e.g. using t-test.

$$T_t = \frac{\hat{\delta}(\beta_0)}{se(\hat{\delta}(\beta_0))} = \frac{\hat{\delta}_0}{\sqrt{\hat{\sigma}_u^2/Z'Z}} \xrightarrow{d} \mathcal{N}(0,1)$$

The 90% CI for β is the set of βs at which $\delta(\beta) = 0$ cannot be rejected at 90% confidence level. A confidence set for β is given by taking all β_0 such that $\mathcal{H}_0 : \delta = 0$ cannot be rejected.

Remark (About Anderson-Rubin (AR) Test). [a]

Consider our model

$$\begin{aligned} y &= X\beta + u, \\ X &= Z\Pi + v, \end{aligned}$$

where X is one-dimensional and test for hypothesis $H_0 : \beta = \beta_0$. Under the null, vector $y - X\beta$ is equal to the error u_t and is uncorrelated with Z (due to exogeneity of instruments). The suggested statistics is:

$$AR(\beta_0) = \frac{(y - X\beta)'P_Z(y - X\beta)}{(y - X\beta)'M_Z(y - X\beta)/(T - k)}.$$

here $P_Z = Z(Z'Z)^{-1}Z', M_Z = I - P_z$.

The distribution of AR does not depend on μ asymptotically $AR \to \chi_k^2/k$. The formula may remind you of the J-test for over-identifying restrictions. It would be a J-test if one were to plugs in $\hat{\beta}_{TSLS}$. In a more general situation of more than one endogenous variable and/or included exogenous regressors AR statistic is F-statistic testing that all coeffcients on Z are zero in the regression of $y - \beta_0 X$ on Z and W.

Note, that one tests all coefficients β simultaneously (as a set) in a case of more than one endogenous regressor. AR confidence set One can construct a confidence set robust towards weak instruments based on the AR test by inverting it. That is, by finding all β which are not rejected by the data. In this case, it is the set :

$$CI = \{\beta_0 : AR(\beta_0) < \chi^2_{k,1-\alpha}\}.$$

The nice thing about this procedure is that solving for the confidence set is equivalent to solving a quadratic inequality. This confidence set can be empty with positive probability (caution!).

[a]Retrieved from MIT14.384 Time Series Analysis, Fall 2007 Professor Anna Mikusheva, Lecture 7-8, https://ocw.mit.edu/courses/14-384-time-series-analysis-fall-2013/365cba34145fa204731e9df202d4771e_MIT14_384F13_lec7and8.pdf

Lecture 2.

Causal Inference

Rubin (1975[11]) and Holland (1986[12]) made up the aphorism[1]:

> *"No causation without manipulation"*

Not everybody agrees with this point of view.

In our lectuere, we'll define causal effects using the potential outcomes framework (Neyman, 1923[13]; Rubin, 1974[14]).

2.1 Potential Outcomes Framework

In this framework, an experiment, or at least a thought experiment, has a treatment, and we are interested in its effect on an outcome or multiple outcomes. Sometimes, the treatment is also called an intervention or a manipulation.

Firstly, we consider an experiment with n units indexed by $i = 1, 2, \cdots, n$. We focus on a treatment with two levels:

$$d_i = \begin{cases} 0 & \text{control} \\ 1 & \text{treatment} \end{cases}$$

We seek to identify the causal effect of treatment d_i on some outcome y_i. For each i, the outcome od interest y_i has two versions:

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

This notation emphasizes that y_{di} is the realization of the outcome y_i that would materialize if unit i received treatment $d_i = d$.

Neyman (1923[13]) first used this notation. It seems intuitive but has some hidden assumptions. Rubin (1980[15]) made the following clarifications on the hidden assumptions.

Assumption 2.1.1 (No interference).

Unit i's potential outcomes do not depend on other units' treatments. This is sometimes called the no-interference assumption.

Assumption 2.1.2 (Consistency).

There are no other versions of the treatment. Equivalently, we require that the treatment levels be well-defined, or have no ambiguity at least for the outcome of interest. This is sometimes called the consistency assumption.

The causal effect of the treatment on the i-th unit is then defined as:

$$\Delta_i = y_{1i} - y_{0i}$$

These potential outcomes are constants at the level of unit i.

Remark (Problem of causal inference).

The fundamental problem in causal inference is that only one treatment can be assigned to a given individual, and so only one of y_{0i} and y_{1i} can be observed. Thus Δ_i can never be observed.

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

Rubin (1980[15]) called the Assumptions 2.1.1 and 2.1.2 above together the *Stable Unit Treatment Value Assumption (SUTVA)*.

The observed outcome of unit i is a function of the potential outcomes and the treatment indicator, we can write:

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

In principle, by virtue of being (discrete) RVs, both d_i and y_i each have a distribution function, which, together with their possible realizations, defines various moments. However, their unconditional probabilities and moments at the level of unit i is not of interest. Only the conditional probabilities of y_i given d_i is of interest.

Remark (Rubin (2005[16])).

Under SUTVA, Rubin (2005) called the $n \times 2$ matrix of potential outcomes the Science Table:

i	y_{1i}	y_{0i}
1	y_{11}	y_{01}
2	y_{12}	y_{02}
$\vdots$	$\vdots$	$\vdots$
n	y_{1n}	y_{0n}

Due to the fundamental contributions of Neyman and Rubin to statistical causal inference, the potential outcomes framework is sometimes referred to as the Neyman Model, the Neyman-Rubin Model, or the Rubin Causal Model. Causal effects are functions of the Science Table. Inferring individual causal effects

$$\tau_i = y_{1i} - y_{0i}, \quad (i = 1, \ldots, n)$$

is fundamentally challenging because we can only observe either y_{1i} or y_{0i}, for each unit i, that is, we can observe only half of the Science Table.

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

Now, although Δ_i itself is unobservable, we can (perhaps remarkably) use randomized experiments to learn certain properties of it. The expectations $\mathbb{E}[y_{0i}]$ and $\mathbb{E}[y_{1i}]$ denote the average potential outcomes across unit i in population.

In particular, large randomized experiments let us recover the **Average Treatment Effect (ATE)**:

$$\text{ATE} = \mathbb{E}[y_{1i} - y_{0i}] = \mathbb{E}[y_{1i}] - \mathbb{E}[y_{0i}]$$

For a population, 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.

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

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

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

2.1.1 Identification of Causal Effects

Now, suppose we observe treatments and outcomes over a random sample n from the overall population, $\{d_i, y_i\}_{i=1}^n = \{d_i, y_{d_i i}\}_{i=1}^n$, as either $y_I = y_{0i}$, or $y_i = y_{1i}$.

Let $n_w = |\{i : d_i = w\}|$ be the size of sets of units in our sample who received and did not receive treatment, respectively. This means that: while we observe a sample of size n of d_i and y_i from the overall population, we are observing a sample of size n_0 of realizations of y_{0i} from the non-treated subpopulation and a sample of size n_1 of realizations of y_{1i} from the treated subpopulation.

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

Based on this data, we can use the analogy principle to consistently estimate the first term in the ATT formula and the second term in the ATC formula:

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

Without further assumptions, we cannot identify the remaining terms. Firstly, we cannot observe $\mathbb{E}[y_{0i}|d_i = 1]$ and $\mathbb{E}[y_{1i}|d_i = 0]$ because we do not observe y_{0i} for treated units, and we do not observe y_{1i} for non-treated units. Secondly, we can not observe $\mathbb{E}[y_{1i}]$ and $\mathbb{E}[y_{0i}]$ because both N_1 and N_0 are random samples from the overall population. As a result, the ATE is in general not identified from our data!

We can define the the difference-in-means estimator as:

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

We define the 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, $y_i = \beta_0 + \beta_1 d_i + u_i$,

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

$$= \beta_1 + \mathbb{E}[u_i|d_i = 1] - \mathbb{E}[u_i|d_i = 0]$$

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

Lecture 3.

Panel Data Analysis

Economists traditionally use the term **panel data** to refer to data structures consisting of observa- tions on individuals for multiple time periods. There are several distinct advantages of panel data relative to cross-section data:

1. Possibility of controlling for unobserved time-invariant endogeneity without the use of instrumental variables
2. Possibility of allowing for broader forms of heterogeneity
3. Modeling dynamic relationships and effects

It's typical to index observations by both the individual i and the time period, t, thus y_{it} denotes a variable for indivual i in time t, where $n = 1, \cdots, N$, $t = 1, \cdots, T$.

Definition 3.0.1 (Balanced and Unbalanced Panel Data[2]).

When observations are available on all individuals for the same time periods we say that the panel is **balanced**. In this case there are an equal number T of observations for each individual and the total number of observations is $n = NT$.

When different time periods are available for the individuals in the sample we say that the panel is **unbalanced**. This is the most common type of panel data set. It does not pose a problem for applications but does make the notation cumbersome and also complicates computer programming.

3.1 Incidental Parameters Problem

3.1.1 Pooled OLS Estimation

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:

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

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:

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

The FOC of this equation is:

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

Hence $\hat{\beta}_{OLS}$ is consistent provided that x_{it} and u_{it} are contemperaneously uncorrelated, as $\mathbb{E}[x_{it}u_{it}] = 0, \forall t$. The regressors are allowed to be correlated with the past, and future u_{it}. This occurs when there's feedback loop by which $y_{i,t-1}$ affects x_{it}.

In this proof, we show that either $N \to \infty$ or $T \to \infty$ is sufficient for consistency of $\hat{\beta}_{POLS}$. However, most panel data applications have a large n and small T dimension, so standrad panel data features T fixed and $n \to \infty$.

3.1.2 Asymptotic Normality

From the analysis of consistency, we know that:

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

Hence:

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

The above model is homogeneous, 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 fully 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:

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

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^*:

$$\hat{\beta}^* = \left(\sum_i \sum_t x^*_{it} x^{*'}_{it}\right) \sum_i \sum_t x^*_{it} y_{it}.$$

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

As with pooled OLS, a random effects analysis puts α_i into the error term. In fact, random effects analysis imposes more assumptions than those needed for pooled OLS: **strict exogeneity** in addition to orthogonality between α_i and x_{it}.

3.2.1 Basic Assumptions and POLS

Stating the assumption in terms of conditional means, we have:

Assumption 3.2.1 (Random Effect).

(a) $\mathbb{E}[u_{it}|X_i, \alpha_i] = 0, \forall t$.

(b) $\mathbb{E}[\alpha_i|X_i] = \mathbb{E}[\alpha_i] = 0$.

where $X_i = (x_{i1}, \cdots, x_{iT})$.

Assumption 3.2.1(a) is the strict exogeneity condition and Assumption 3.2.1(b) is is how we will state the orthogonality.

Remark (Why Strict Ecogeneity?[3]).

Why do we maintain Assumption 3.2.1(a) when it is more restrictive than needed for a pooled OLS analysis? Because the random eꀀects approach exploits the serial correlation in the composite error, $v_{it} = \alpha_i + u_{it}$, in a generalized least squares (GLS) framework. In order to ensure that feasible GLS is consistent, we need some form of strict exoge- neity between the explanatory variables and the composite error.

Under this assumption, we can write:

$$\begin{aligned} y_{it} &= x'_{it}\beta + v_{it} \\ \mathbb{E}[v_{it}|X_i] &= 0, t = 1, \cdots, T \end{aligned}$$

The conditions shows that our model satisfies the GLS assumption, which confirms that we can apply GLS methods that account for the particular error structure $v_{it} = \alpha_i + u_{it}$.

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

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

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

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

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: $y_{it} = \alpha + x'_{it}\beta + u_{it} = \tilde{x}'_{it}\tilde{\beta} + v_{it}$. $\hat{\tilde{\beta}}$ 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.

3.2.2 From POLS to GLS

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.

Assumption 3.2.2 (Random Effect).

$\operatorname{rank}\mathbb{E}\left[X'_i\Omega^{-1}X_i\right] = K$

We know that both GLS and feasible GLS estimator would be consistent under Asusmption 3.2.1 and 3.2.2. A general FGLS analysis, using an unrestricted variance estimator Ω, is consistent and asymptotically normal as $N \to \infty$.

But, we won't exploit the unobserved effects structure v_{it}. A standard random e�ects analysis adds assumptions on the idiosyncratic errors that give Ω a special form. The first assumption is that the idiosyncratic errors u_{it} have a constant unconditional variance across t:

Assumption 3.2.3 (RE-Homoskedasticity).
$\mathbb{E}[u_{it}^2] = \sigma_u^2, \forall t$

The second assumption is that the idiosyncratic errors are serially uncorrelated:

Assumption 3.2.4 (RE-Serial Uncorrelated).
$\mathbb{E}[u_{it}u_{is}] = 0, \forall t \neq s$

Under these two assumptions, we can derive the variances and covariances of the elements of v_i. Given the error structure the natural estimator for β is GLS. The GLS eimator for β is:

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

where $\Omega^{-\frac{1}{2}}y_i = \Omega^{-\frac{1}{2}}\tilde{X}_i'\tilde{\beta} + \Omega^{-\frac{1}{2}}v_i$.

$$\begin{aligned}
\Omega = \mathbb{E}[v_iv_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} \\
&= \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} \\
&= \begin{bmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_u^2 + \sigma_\alpha^2 \end{bmatrix} \\
&= \sigma_\alpha^2 \mathbf{1}_i\mathbf{1}_i' + \sigma_u^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_u^2, \forall i.
\end{aligned}$$

where I is an identity matrix of dimention T_i. Under the assumption $\mathbb{E}[u_{it}x_{is}] = 0$, 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 conditional variance of $\hat{\tilde{\beta}}_{RE-GLS}$ is:

$$
\mathbb{V}\Big[\hat{\tilde{\beta}}_{RE-GLS}\Big] = \Big(\sum_i \tilde{X}_i'\Omega^{-1}\tilde{X}_i\Big)^{-1}\sigma_u^2
$$

The asymptotic variance of $\hat{\tilde{\beta}}_{RE-GLS}$ is:

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

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

If indeed we have:

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

which implies homoskedasticity.

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

$$
\begin{aligned}
\mathbb{E}[u_iu_i'|\tilde{X}_i, \tilde{\alpha}_i] &= \sigma_u^2 I_T \\
\mathbb{E}[\tilde{\alpha}_i^2|\tilde{x}_i] &= \sigma_\alpha^2
\end{aligned}
$$

We obtain: $\hat{\Omega} = \hat{\sigma}_\alpha^2\mathbf{1}_i\mathbf{1}_i' + \hat{\sigma}_u^2 I_T$, a $T \times T$ matrix that we assume to be positive definite. In a panel data context, the FGLS estimator that uses this variance matrix is what is known as the **random effects estimator**.

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.

3.2.3 Comparing POLS and GLS

Now, let's compare the $\hat{\beta}_{RE-GLS}$ with the pooled estimator $\hat{\beta}_{POLS}$.

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,

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

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

$$\begin{aligned}\hat{\beta}_{FE-W} &= \left(\sum_i\sum_t \ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\sum_i\sum_t \ddot{x}_{it}\ddot{y}_{it}\\ &= \beta + \left(\sum_i\sum_t \ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\sum_i\sum_t \ddot{x}_{it}\ddot{u}_{it}\\ &\xrightarrow{p} \beta + \mathbb{E}\left[\sum_t \ddot{x}_{it}\ddot{x}_{it}'\right]^{-1}\mathbb{E}\left[\sum_t \ddot{x}_{it}\ddot{u}_{it}\right]\end{aligned}$$

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

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:

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

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)).

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

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

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 $N \to \infty$

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

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)

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.

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

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] Peng Ding. *A First Course in Causal Inference*. 2023. arXiv: 2305.18793 [stat.ME]. URL: https://arxiv.org/abs/2305.18793 (p. 6)

[2] Bruce E. Hansen. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2022 (p. 10)

[3] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Massachusetts: The MIT Press, 2010 (p. 13)

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

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

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

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

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

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

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

Others

[11] Donald B. Rubin. "Bayesian Inference for Causality: The Importance of Randomization". In: *The Annals of Statistics* 3.1 (1975), pp. 121–131. DOI: 10.1214/aos/1176343238 (p. 6)

[12] Paul W. Holland. "Statistics and Causal Inference(with discussion)". In: *Journal of the American Statistical Association* 81.396 (1986), pp. 945–960. DOI: 10.1080/01621459.1986.10478373 (p. 6)

[13] Jerzy Neyman. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9". In: *Statistical Science* 5.4 (1923), pp. 465–472 (p. 6)

[14] Donald B. Rubin. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66.5 (1974), pp. 688–701. DOI: 10.1037/h0037350 (p. 6)

[15] Donald B. Rubin. "Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization Test" by D. Basu". In: *Journal of the American Statistical Association* 75.371 (1980), pp. 591–593. DOI: 10.1080/01621459.1980.10477410 (pp. 6, 7)

[16] Donald B. Rubin. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions". In: *Journal of the American Statistical Association* 100.469 (2005), pp. 322–331. DOI: 10.1198/016214504000001880 (p. 7)

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

[18] 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

[19] 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

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

[21] 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

[22] 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

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