Introduction to Econometrics II: Recitation 6[*]

Seung-hun Lee[†]

February 26th, 2020

1 Panel Data

1.1 Random Effects

Consider the following data generating process

$$y_{it} = x'_{it}\beta_1 + \alpha_i + u_{it}, \quad (i = 1, .., n \text{ and } t = 1, .., T)$$

where we assume that $E(\alpha_i) = 0$, $E(\alpha_i u_{it}) = 0$, u_{it} is IID across i and t and independent of x_{it}. On top of this, a key assumption that separates random effects from fixed effects is α_i and x_{it} is now **uncorrelated**. The major consequence of this assumption is that even the POLS estimator becomes consistent. (We also still assume that strict exogeneity holds.)

To show this, I apply an OLS procedure on the data generating process above, while writing $v_{it} = \alpha_i + u_{it}$. This gets us

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}y_{it}$$

which can alternatively be written as

$$\hat{\beta}_{POLS} - \beta_1 = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}v_{it}$$

We need to work with the asymptotics of $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}v_{it}$. Since

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}v_{it} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}(\alpha_i + u_{it})$$

$$= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}\alpha_i + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}u_{it}$$

$$\xrightarrow{p} E(x_{it}\alpha_i) + E(x_{it}u_{it}) = 0 + 0 = 0$$

The last summation terms converge in probability to 0 now. The reason POLS in fixed effects setting was inconsistent was that this term converges in probability something nonzero. Therefore, POLS is now consistent as well.

However, this is not really the best we can do. The reason is that the composite error from different time periods are serially correlated. This can be seen by

$$cov(v_{it}, v_{is}) = cov(\alpha_i + u_{it}, \alpha_i + u_{is})$$

$$= E[(\alpha_i + u_{it})(\alpha_i + u_{is})] - E[\alpha_i + u_{it}]E[\alpha_i + u_{is}]$$

$$= E[\alpha_i^2 + \alpha_i u_{it} + \alpha_i u_{is} + u_{it}u_{is}] - (0+0)(0+0)$$

$$= E[\alpha_i^2] + E[\alpha_i u_{it}] + E[\alpha_i u_{is}] + E[u_{it}u_{is}]$$

$$= var(\alpha_i)$$

If it is the case that $var(\alpha_i) = 0$, then α_i is a degenerate variable centered at 0 - effectively there is no individual fixed effect. Except for this case, the composite error terms would be serially correlated.

The estimation method that is the most efficient for this case would be the GLS. The key to constructing a GLS is to find a variance-covariance matrix for $\mathbf{v}_i$. What we can do is to stack the T observations for each individual and work with

$$\mathbf{y}_i = \mathbf{X}_i\beta_1 + \mathbf{v}_i$$

where $\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ ... \\ y_{iT} \end{pmatrix}$, $\mathbf{X}_i = \begin{pmatrix} x'_{i1} \\ ... \\ x'_{iT} \end{pmatrix}$, $\mathbf{v}_i = \begin{pmatrix} v_{i1} \\ ... \\ v_{iT} \end{pmatrix} = \begin{pmatrix} \alpha_i + u_{i1} \\ ... \\ \alpha_i + u_{iT} \end{pmatrix}$. To make algebra more straightforward, I will define 1_T as a T-dimensional column vector of 1's. Then,

$$\mathbf{v}_i = 1_T\alpha_i + \mathbf{u}_i$$

Then the variance can be calculated by the $E[(X - E(X))(X - E(X))']$ formula. Thus,

$$
\begin{aligned}
var(v_i) &= E[(\mathbf{v}_i - E(\mathbf{v}_i))(\mathbf{v}_i - E(\mathbf{v}_i))'] \\
&= E[\mathbf{v}_i\mathbf{v}_i'] \ (\because E(\mathbf{v}_i) = 0) \\
&= E\left[1_T\alpha_i^2 1_T' + \mathbf{u}_i\alpha_i 1_T' + 1_T\alpha_i\mathbf{u}_i + \mathbf{u}_i\mathbf{u}_i'\right] \\
&= 1_T 1_T' var(\alpha_i) + E[\mathbf{u}_i\mathbf{u}_i']
\end{aligned}
$$

Suppose that u_{it} is homoskedastic, then $E[\mathbf{u}_i\mathbf{u}_i']$ becomes $\sigma_u^2 I_T$, the result derived in class.

The GLS estimator can be derived by the following: Note that V is a symmetric matrix and positive semidefinite. Then $V = Q\Omega Q'$ by spectral decomposition theorem[1] and $V^{-1/2}$ can be defined (It would be $Q\Omega^{-1/2}Q'$). The trick now is to pre-multiply $V^{-1/2}$ to get

$$
V^{-1/2}\mathbf{y}_i = V^{-1/2}\mathbf{X}_i\beta_1 + V^{-1/2}\mathbf{v}_i
$$

and the resulting GLS estimator would be

$$
\left(\sum_{i=1}^{n}(V^{-1/2}\mathbf{X}_i)'(V^{-1/2}\mathbf{X}_i)\right)^{-1}\sum_{i=1}^{n}(V^{-1/2}\mathbf{X}_i)'(V^{-1/2}\mathbf{y}_i) = \left(\sum_{i=1}^{n}\mathbf{X}_i'V^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i'V^{-1}\mathbf{y}_i
$$

Comment 1.1 (So when do we use fixed effects vs. random effects?). *One way of figuring out when to use random effect is through a Hausmann Test. Random effects can be considered as a special case of fixed effects. In this test, the null hypothesis would be random effects and the alternative would be fixed effects. Note that the ingredients we have here are similar to the Hausmann Test in the endogeneity test. Random effects estimators would only be consistent if H_0 is true whereas fixed effects are consistent in both H_0 and H_1. If you are planning to use the random effects estimator (and believe that the random effects assumptions are appropriate in your context), the Hausmann test can be used as a specification check for this assumption and provide evidence to support your approach.*

---

[1]It is a theorem that states that if a matrix is symmetric and positive semidefinite, there exists real eigenvalues and real orthogonal eigenvectors. You can look at Gockenbach(2010) for more details.

2 Dynamic Panel Data

Now consider a setting where

$$y_{it} = \rho y_{it-1} + \alpha_i + u_{it}$$

where $t = 0, 1, ..., T$ and $i = 1, .., n$. Additionally, assume that $E[\alpha_i] = 0$, $E[u_{it}] = 0$, $E[\alpha_i u_{it}] = 0 \; \forall i, t$ and $E[u_{it} u_{is}] = 0$ for $t \neq s$. Also, the initial observation y_{i0} satisfies $E[y_{i0} u_{it}] = 0 \; (t \geq 1)$ and that $E|u_{it}|^{2+\delta} \leq M < \infty (\delta > 0)$. Also note that observations are independent across i.

The difference between this and the previous panel estimation is that the lagged dependent variable enters as the regressor. In this case, all of the previous methods - random effects and fixed effects - are inconsistent. I will show this for POLS. Let $v_{it} = \alpha_i + u_{it}$. The OLS estimates would be

$$\hat{\rho} = \left(\sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1}^2 \right)^{-1} \sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1} y_{it}$$

or

$$\hat{\rho} - \rho = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1}^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1} (\alpha_i + u_{it})$$

The problem is that since $y_{it-1} = \rho y_{it-2} + \alpha_i + u_{it-1}$, the $\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1} \alpha_i$ term does not converge in probability to 0.

Even if we get rid of α_i by using the first difference at $T = 2$, ρ estimates are still inconsistent. To show this, start with

$$\Delta y_{i2} = \rho \Delta y_{i1} + \Delta u_{i2}$$

We can now show that the regressor and the error term are correlated, since

$$
\begin{aligned}
cov(\Delta y_{i1}, \Delta u_{i2}) &= E[(y_{i1} - y_{i0})(u_{i2} - u_{i1})] \\
&= E[y_{i1} u_{i2}] - E[y_{i1} u_{i1}] - E[y_{i0} u_{i2}] + E[y_{i0} u_{i1}]
\end{aligned}
$$

Because $E[y_{i1} u_{i1}]$ contains terms from the 1st period, this contains term that is nonzero.

Even for the within estimator, which is written as

$$y_{it} - \frac{1}{T} \sum_{i=1}^{T} y_{it} = \rho \left(y_{it-1} - \frac{1}{T} \sum_{i=1}^{T} y_{it-1} \right) + u_{it} - \frac{1}{T} \sum_{i=1}^{T} u_{it}$$

The regressor contains $y_{i0}, .., y_{iT-1}$ and residuals contain $u_{i1}, .., u_{iT}$. There are overlapping time periods, implying that the regressor becomes endogenous.

2.1 Anderson-Hsiao Estimator

In estimating

$$y_{it} = \rho y_{it-1} + \alpha_i + u_{it}$$

Anderson and Hsiao (1982) proposed the following IV approach. First difference the above equation and obtain

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$$

where possible values of i remain the same but $t = 2, ..., T$. The suggested IV is that Δy_{it-1} be instrumented with y_{it-2}. We can check that this is a valid IV.

- **Relevancy**: Note that $\Delta y_{it-1} = y_{it-1} - y_{it-2}$. So this term contains y_{it-2} and thus relevancy is satisfied.

- **Exogeneity**: Note that

$$\begin{aligned}
cov(y_{it-2}, \Delta u_{it}) &= E[y_{it-2}, u_{it} - u_{it-1}] \\
&= E[y_{it-2}u_{it}] - E[y_{it-2}u_{it-1}] \\
&= 0 - 0 = 0
\end{aligned}$$

The last line is justified as follows. Since we are assuming $E[y_{i0}u_{it}] = 0$ for $t \geq 1$, we can expand to

$$\begin{aligned}
E[y_{i1}u_{it}] \ (t \geq 2) &= E[(\rho y_{i0} + \alpha_i + u_{i1})u_{it}] \\
&= \rho E[y_{i0}u_{it}] + E[\alpha_i u_{it}] + E[u_{i1}u_{it}] \\
&= \rho \times 0 + 0 + 0 = 0
\end{aligned}$$

Likewise,

$$\begin{aligned}
E[y_{i2}u_{it}] \ (t \geq 3) &= E[(\rho y_{i1} + \alpha_i + u_{i2})u_{it}] \\
&= \rho E[y_{i1}u_{it}] + E[\alpha_i u_{it}] + E[u_{i2}u_{it}] = 0
\end{aligned}$$

Therefore, we can generalize to $E[y_{is}u_{it}] = 0$ for $t \geq s + 1$.

We can generalize using matrix notation. Define

$$\Delta \mathbf{y}_i = \rho \Delta \mathbf{y}_{i,-1} + \Delta \mathbf{u}_i$$

where $\Delta\mathbf{y}_i = \begin{pmatrix} \Delta y_{i2} \\ ... \\ \Delta y_{iT} \end{pmatrix}$, $\Delta\mathbf{y}_{i,-1} = \begin{pmatrix} \Delta y_{i1} \\ ... \\ \Delta y_{iT-1} \end{pmatrix}$ and $\Delta\mathbf{u}_i = \begin{pmatrix} \Delta u_{i2} \\ ... \\ \Delta u_{iT} \end{pmatrix}$. The matrix of instruments Z_i would be

$$Z_i = \begin{pmatrix} y_{i0} & 0 & 0 & ... \\ 0 & y_{i1} & 0 & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & y_{iT-2} \end{pmatrix} \in \mathbb{R}^{(T-1)\times(T-1)}$$

This is a just identified case in the sense that the number of IV is equal to the number of endogenous variables. So we solve the sample analogue of

$$E[Z_i'\Delta\mathbf{u}_i] = 0 \iff E[Z_i'(\Delta\mathbf{y}_i - \rho\Delta\mathbf{y}_{i,-1})] = 0$$

$$\implies \frac{1}{n}\sum_{i=1}^{n}(Z_i'\Delta\mathbf{y}_i - \rho Z_i'\Delta\mathbf{y}_{i,-1}) = 0$$

$$\iff \rho\frac{1}{n}\sum_{i=1}^{n}Z_i'\Delta\mathbf{y}_{i,-1} = \frac{1}{n}\sum_{i=1}^{n}\rho Z_i'\Delta\mathbf{y}_i$$

$$\iff \hat{\rho} = \left(\sum_{i=1}^{n}Z_i'\Delta\mathbf{y}_{i,-1}\right)^{-1}\sum_{i=1}^{n}Z_i'\Delta\mathbf{y}_i$$

2.2 Arellano-Bond Estimator

2.2.1 Without Regressors

This approach is similar to that of Anderson-Hsiao in the sense that we start with the first differentiation. The difference is in the instruments used. Arellano and Bond (1991) suggests that to instrument for Δy_{it-1}, we use $y_{i0}, ..., y_{it-2}$ as instruments. To see why they are valid

- **Relevancy:** It should be clear why y_{it-2} is relevant. As for others, since $y_{it-1} = \rho y_{it-2} + u_{it-1}$ and $y_{it-2} = \rho y_{it-3} + u_{it-2}$ We can write recursively that

$$y_{it-1} = \rho^2 y_{it-3} + \rho u_{it-2} + u_{it-1}$$

... and so on. Therefore, we can verify relevancy.

- **Exogeneity:** Note that $cov(y_{is}, \Delta u_{it})$ for any $s < t$ is 0, as we have shown above. So exogeneity holds as well.

The generalized approach using matrix will be similar except for the instrument matrix Z_i.

$$Z_i = \begin{pmatrix} y_{i0} & 0 & 0 & ... \\ 0 & (y_{i0}, y_{i1}) & 0 & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & (y_{i0}, ..., y_{iT-2}) \end{pmatrix} \in \mathbb{R}^{(T-1) \times \frac{T(T-1)}{2}}$$

Unlike Anderson-Hsiao estimator, this is an overidentified case. So we would need to use a GMM criterion with a weight matrix W_n. This would result in

$$\hat{\rho} = \arg \min_{\rho} \left\{ n \times \frac{1}{n} \sum_{i=1}^{n} (Z_i' \Delta \mathbf{y}_i - \rho Z_i' \Delta \mathbf{y}_{i,-1})' W_n \frac{1}{n} \sum_{i=1}^{n} (Z_i' \Delta \mathbf{y}_i - \rho Z_i' \Delta \mathbf{y}_{i,-1}) \right\}$$

$$\implies \left[\left(\sum_{i=1}^{n} \Delta \mathbf{y}_{i,-1}' Z_i \right) W_n \left(\sum_{i=1}^{n} Z_i' \Delta \mathbf{y}_{i,-1} \right) \right]^{-1} \left(\sum_{i=1}^{n} \Delta \mathbf{y}_{i,-1}' Z_i \right) W_n \left(\sum_{i=1}^{n} Z_i' \Delta \mathbf{y}_i \right)$$

2.2.2 Selecting Weighting Matrix W_n

The remaining question is to now select a matrix W_n that would lead to the lowest variance possible. If $g(Z_i, \rho)$ is the moment condition, the following would qualify as the most efficient weighting matrix.

$$W_n = E[g(Z_i, \rho)g(Z_i, \rho)']^{-1}$$

In our context, $g(Z_i, \rho)$ would be equivalent to $Z_i' \Delta u_i$. So we would need to find a sample analogue of

$$E[Z_i' \Delta \mathbf{u}_i \Delta \mathbf{u}_i' Z_i] \implies \frac{1}{n} \sum_{i=1}^{n} Z_i \Delta \mathbf{u}_i \Delta \mathbf{u}_i' Z_i$$

For further simplification, we can assume two types of settings

- u_{it} is **IID and homoskedastic**: In such case, where $E[u_{it}^2] = \sigma_u^2$, we can write

$$E[\Delta \mathbf{u}_i \Delta \mathbf{u}_i] = E \begin{bmatrix} \Delta u_{i2}^2 & \Delta u_{i2} \Delta u_{i3} & ... & \Delta u_{i2} \Delta u_{iT} \\ \Delta u_{i2} \Delta u_{i3} & \Delta u_{i3}^2 & ... & \Delta u_{ie} \Delta u_{iT} \\ ... & ... & ... & ... \\ \Delta u_{i2} \Delta u_{iT} & \Delta u_{i3} \Delta u_{iT} & ... & \Delta u_{iT}^2 \end{bmatrix} = \begin{bmatrix} 2\sigma_u^2 & -\sigma_u^2 & ... & 0 \\ -\sigma_u^2 & 2\sigma_u^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 2\sigma_u^2 \end{bmatrix}$$

Define matrix $H \in \mathbb{R}^{(T-1) \times (T-1)}$ to have 2's in the diagonal elements, -1's in the immediate off-diagonals, and 0 everywhere else. Then $E[\Delta \mathbf{u}_i \Delta \mathbf{u}_i'] = \sigma_u^2 H$. This implies

that the weighting matrix that we are looking for is

$$E[Z_i'\Delta\mathbf{u}_i\Delta\mathbf{u}_i'Z_i]^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n}Z_iHZ_i\right)^{-1}$$

where σ_u^2 is taken out since scaling W_n by a scalar does not affect the value of the esti-
mator. However, note that it may affect the test statistics for the overidentification test
we will conduct later.

- u_{it} is **heteroskedastic**: For this, we take an approach similar to the two-step GMM
  estimator we did some weeks ago. The optimal weighting matrix in this case would be

$$W_n = \left(\frac{1}{n}\sum_{i=1}^{n}Z_i\Delta\tilde{\mathbf{u}}_i\Delta\tilde{\mathbf{u}}_i'Z_i\right)^{-1}$$

  where $\Delta\tilde{\mathbf{u}}_i = \Delta\mathbf{y}_i - \tilde{\rho}\Delta\mathbf{y}_{i,-1}$, a residual from the preliminary estimator $\tilde{\rho}$. The pre-
  liminary estimator could be either from $W_n = I_{T(T-1)/2}$ or from the one-step GMM
  estimator that we derived earlier.

2.2.3   Overidentification Test

Because we are using more moment conditions that the number of endogenous variables,
this is when we could test for an overidentification restriction. Suppose that W_n is the effi-
cient weighting matrix. Then, similar to the GMM overidentification test, we are testing

$$H_0 : E[g(Z_i,\rho)] = 0, \quad H_1 : E[g(Z_i,\rho)] \neq 0$$

We can construct the following test statistic (Sargan-Hansen overidentification statistic)

$$J = n\bar{g}_n(\hat{\rho})'W_n\bar{g}_n(\hat{\rho})$$

Under H_0, J has a limiting distribution $\chi^2_{\left(\frac{T(T-1)}{2}-1\right)}$. We lose one degree of freedom since we
have used one estimator for ρ.

2.2.4 With Regressors

Now we generalize further by including regressors $x_{it} \in \mathbb{R}^k$. We can write the data generating process as

$$y_{it} = \rho y_{it-1} + \beta' x_{it} + \alpha_i + u_{it}$$

There are two types of assumption we can put on the regressors

Assumption 2.1 (Predetermined). *We say that the regressors are **predetermined** if*

$$E[u_{it}|x_{i0}, ..., x_{it}] = 0 \text{ for each } t = 0, ..., T$$

Assumption 2.2 (Strictly Exogenous). *We say that the regressors are **strictly exogenous** if*

$$E[u_{it}|x_{i0}, ..., x_{iT}] = 0 \text{ for all } t = 0, ..., T$$

Since there is an unobservable term α_i which we allow to be correlated with x_{it} and y_{it-1}, one thing we can do is to take the first difference

$$\Delta y_{it} = \rho \Delta y_{it-1} + \beta' \Delta x_{it} + \Delta u_{it}$$

However, note that the equation is endogenous because of the two reasons

- Δy_{it-1} is endogenous as

$$\begin{aligned} E[\Delta y_{it-1} \Delta u_{it}] &= E[(y_{it-1} y_{it-2})(u_{it} - u_{it-1})] \\ &= E[y_{it-1} u_{it}] - E[y_{it-1} u_{it-1}] - E[y_{it-2} u_{it}] + E[y_{it-2} u_{it-1}] \end{aligned}$$

Since we still assume $E[y_{i0} u_{it}] = 0$ for $t \geq 1$, (and thus $E[y_{it} u_{is}] = 0$ for $s \geq t+1$), we can still get rid of other terms except $E[y_{it-1} u_{it-1}]$. The remaining term has a nonzero correlation, so Δy_{it-1} becomes endogenous

- Δx_{it} is endogenous under predetermined case since

$$\begin{aligned} E[\Delta x_{it} \Delta u_{it}] &= E[(x_{it} - x_{it-1})(u_{it} - u_{it-1})] \\ &= E[x_{it} u_{it}] - E[x_{it} u_{it-1}] - E[x_{it-1} u_{it}] + E[x_{it-1} u_{it-1}] \end{aligned}$$

While other terms are 0 by predetermined assumption, this assumption is silent about $E[x_{it} u_{it-1}]$. This is where the nonzero correlation is from.

Given the two possible assumptions on the regressors, there are two sets of IV/GMM estimation that we can use.

- **Predetermined Regressors**: Rewrite the differenced equation and stack in the vector form for each individual by writing

$$\Delta \mathbf{y}_i = \Delta \mathbf{w}_i \delta + \Delta \mathbf{u}_i$$

where $\Delta \mathbf{w}_i = [\Delta \mathbf{y}_{i,-1}, \Delta \mathbf{x}_i] \in \mathbb{R}^{(T-1)\times(k+1)}$, $\Delta \mathbf{x}_i = \begin{pmatrix} x_{i2} \\ ... \\ x_{iT} \end{pmatrix}$, $\delta = \begin{pmatrix} \rho \\ \beta' \end{pmatrix}$ and the other two variables are the same as before. For notational convenience, I define

$$\mathbf{y}_{it} = \begin{pmatrix} y_{i0} \\ ... \\ y_{it} \end{pmatrix}, \mathbf{x}_{it} = \begin{pmatrix} x'_{i0} \\ ... \\ x'_{it} \end{pmatrix}$$

The possible instruments for the endogenous variable Δx_{it} is any regressor up to period $t-1$, or $\mathbf{x}'_{it-1}$. The exogeneity is satisfied since the error term is $u_{it} - u_{it-1}$, regressors $x_{i1}, ..., x_{it-1}$ is not correlated to this by predeterminedness assumption. Since $\mathbf{x}'_{it-1}$ includes x_{it-1}, the IV's are relevant as well. Thus, we can write

$$Z_i = \begin{pmatrix} y_{i0}, \mathbf{x}'_{i1} & 0 & 0 & ... & 0 \\ 0 & \mathbf{y}'_{i1}, \mathbf{x}'_{i2} & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \mathbf{y}'_{iT-2}, \mathbf{x}'_{iT-1} \end{pmatrix}$$

Then we take a GMM approach by using the moment condition $E[Z'_i \Delta \mathbf{u}_i] = 0$, which results in the following estimator

$$\hat{\delta} = \left[\left(\sum_{i=1}^{n} \Delta \mathbf{w}'_i Z_i \right) W_n \left(\sum_{i=1}^{n} Z'_i \Delta \mathbf{w}_i \right) \right]^{-1} \left(\sum_{i=1}^{n} \Delta \mathbf{w}'_i Z_i \right) W_n \left(\sum_{i=1}^{n} Z'_i \Delta \mathbf{y}_i \right)$$

- **Strictly exogenous**: The difference between predetermined and strictly exogenous case is that the entire elements in $\mathbf{x}'_{iT}$ is a valid IV. Strict exogeneity prevents any correlation between error terms of all time periods and there are elements in $\mathbf{x}'_{iT}$ that are included

in Δy_{it} for each t. So we can write the Z_i matrix as

$$Z_i = \begin{pmatrix} y_{i0}, \mathbf{x}'_{iT} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{y}'_{i1}, \mathbf{x}'_{iT} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{y}'_{iT-2}, \mathbf{x}'_{iT} \end{pmatrix}$$

We still work with the same moment condition $E[Z'_i \Delta \mathbf{u}_i] = 0$. The resulting estimator for δ shares the same expression as in the predetermined case, but with different matrix for Z_i.