Introductory Econometrics

Lecture 29: Panel data models

Instructor: Ma, Jun

Renmin University of China

June 7, 2023

References

1. Wooldridge, “Introductory Econometrics - A Modern Approach”, 7th edition
2. Verbeek, “A Guide to Modern Econometrics”, 5th edition

Panel data

- In panel data, individuals (households, firms, cities, ...) are observed at several points in time (days, years, ...).
- We assume that our data has the feature that there are many individuals (i.e. large cross section dimension) and we observe their information/characteristics in relative fewer number of periods (i.e. small time series dimension).
- Panel data are most useful when we suspect that the outcome variable depends on explanatory variables which are not observable but correlated with the observed explanatory variables.
- If such omitted variables are constant over time, panel data estimators allow to consistently estimate the effect of the observed explanatory variables.

- Balanced panel: each individual is observed for the same time periods.
- Unbalanced panel: individuals are observed for different time periods.
- Static panel models: lagged dependent variables are assumed not to have direct causal effects and they are not among the right-hand side explanatory variables.
- Dynamic panel models: lagged dependent variables have causal effects and are included as explanatory variables.

Linear panel model

- Consider the linear model for individual $i = 1, 2, ..., N$ who is observed at several time periods $t = 1, ..., T$. We assume N is large and T is small.
- Our explained variable Y_{it} is generated by

$$Y_{it} = \alpha_t + \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it} + \gamma_1 Z_{1,i} + \cdots + \gamma_m Z_{m,i} + c_i + U_{it}.$$

  - $X_{1,it}, ..., X_{k,it}$: time-varying explanatory variables;
  - $Z_{1,i}, ..., Z_{m,i}$: time-invariant explanatory variables;
  - α_t: time-specific effect;
  - c_i: an individual-specific effect;
  - U_{it}: an idiosyncratic error term.
- For example, Y_{it} is the output of firm i at time t; $X_{1,it}, ..., X_{k,it}$ are inputs; $Z_{1,i}, ..., Z_{m,i}$ are time-invariant characteristics such as geographic location; α_t is the time effect of the macroeconomic environment; c_i is unobserved management quality; U_{it} is the random shock.

Baseline fixed effect model

We assume that there is only one time-varying explanatory variable X_{it}

$$Y_{it} = \alpha + \beta X_{it} + c_i + U_{it}$$

and no time effect. Both c_i and U_{it} are unobserved random variables.

1. Our data

$$\{Y_{i1}, ..., Y_{iT}, X_{i1}, ..., X_{iT}\}_{i=1}^{N}$$

   are independently and identically distributed. Observations are independent across individuals but not necessarily across time.
2. The idiosyncratic error term is assumed to be exogenous:

$$\mathrm{E}[U_{it}] = \mathrm{E}[U_{it} X_{i1}] = \cdots = \mathrm{E}[U_{it} X_{iT}] = 0, \forall t = 1, ..., T,$$

3. No serial correlation along the time dimension within the individual i: $\mathrm{E}[U_{it} U_{is} \mid X_{i1}, ..., X_{iT}] = 0 \ \forall t \neq s$.
4. Homoskedasticity: $\mathrm{E}\left[U_{it}^2 \mid X_{i1}, ..., X_{iT}\right] = \sigma_U^2$ for some constant $\sigma_U^2 > 0$, $\forall t = 1, ..., T$.
5. c_i and U_{it} are uncorrelated: $\mathrm{E}[c_i] = \mathrm{E}[U_{it} c_i] = 0$, $\forall t = 1, ..., T$.

Pooled OLS

- The pooled OLS estimator ignores the panel structure of the data:

$$\hat{\beta}^{POLS} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\overline{X}\right)\left(Y_{it}-\overline{Y}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\overline{X}\right)^{2}}$$

  is inconsistent if the unobserved individual effect c_i is correlated with X_{it}.
- If c_i is uncorrelated with X_{it}, the pooled OLS is consistent but the standard error has to be adjusted, since the error terms $V_{it} = c_i + U_{it}$ are serially correlated ($\mathrm{E}\left[V_{is}V_{it}\right] = \mathrm{E}\left[c_i^2\right]$, if $s \neq t$).
- In most applications, the assumption that c_i is uncorrelated with X_{it} is not reasonable. For example, the management quality should be negatively correlated with inputs.

One-way fixed effect estimator

- Average the equation

$$Y_{it} = \alpha + \beta X_{it} + c_i + U_{it}$$

over $t = 1, \ldots, T$ to get

$$\overline{Y}_i = \alpha + \beta \overline{X}_i + c_i + \overline{U}_i$$

where $\overline{Y}_i = T^{-1} \sum_{t=1}^{T} Y_{it}$, $\overline{X}_i = T^{-1} \sum_{t=1}^{T} X_{it}$ and $\overline{U}_i = T^{-1} \sum_{t=1}^{T} U_{it}$ are averages in the time dimension.

- Subtract the "average" equation from the original equation to obtain:

$$\dot{Y}_{it} = \beta \dot{X}_{it} + \dot{U}_{it},$$

where $\dot{Y}_{it} = Y_{it} - \overline{Y}_i$, $\dot{X}_{it} = X_{it} - \overline{X}_i$ and $\dot{U}_{it} = U_{it} - \overline{U}_i$. This step is called one-way within transformation.

- We regress the "within-transformed" variables $\dot{Y}_{it}$ on $\dot{X}_{it}$ (with the panel structure ignored) without an intercept to obtain

$$\hat{\beta}^{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it} \dot{Y}_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it}^2}.$$

- Under the weak exogeneity assumption, $\dot{U}_{it}$ is uncorrelated with $\dot{X}_{it}$ for all $i = 1, \ldots, N$ for all $t = 1, \ldots, T$, the fixed effect estimator (or within estimator) is consistent and asymptotically normal (as $N \to \infty$ and T remains to be fixed) since in the regression model

$$\dot{Y}_{it} = \beta \dot{X}_{it} + \dot{U}_{it},$$

  the error term is uncorrelated with $\dot{X}_{it}$.
- The within transformation approach estimates the effects of time-varying variables only. The within transformation eliminates all the time-invariant variables.
- The fixed effect estimator is consistent even when the idiosyncratic error terms in different time periods ($s \neq t$) U_{it} and U_{is} are correlated, for the same individual.

Standard errors

- The covariance between $\dot{U}_{it}$ and $\dot{U}_{is}$ for $t \neq s$ is

$$\mathrm{E}\left[\dot{U}_{it}\dot{U}_{is}\right] = -\frac{\sigma_U^2}{T}.$$

- Similarly the variance of $\dot{U}_{it}$ is found to be

$$\mathrm{E}\left[\dot{U}_{it}^2\right] = \sigma_U^2\left(1 - \frac{1}{T}\right).$$

- Under the assumptions of no serial correlation and homoskedasticity, the asymptotic variance of $\hat{\beta}^{FE}$ can be estimated by

$$\hat{\sigma}_U^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{X}_{it}^2\right)^{-1},$$

where $\hat{\sigma}_U^2$ is a consistent estimator of σ_U^2.

- This estimator is called “fixed” effect due to a historical reason. During the years when econometricians look at only finite-sample properties, the individual-specific effects c_i’s are considered to be fixed constants that shift the intercept.
- The least squares dummy variables (LSDV) estimator is a pooled OLS estimator including a set of $N-1$ individual-specific dummy variables which identify the individuals and hence an additional $N-1$ parameters.
- Time-invariant explanatory variables are dropped because of the multicollinearity problem.
- The LSDV estimator for β (i.e., the coefficients for those time-varying explanatory variables) is numerically identical to the fixed effect estimator.

- It is possible that compared with usual pooled OLS regression, after adding individual-specific dummy variables, the R^2, as a measure of goodness-of-fit, becomes larger. This is possibly due to the reason that individual-specific (fixed) effects largely explain the variation in the explained variable.
- Occasionally, the estimated individual effects are of interest and in this case, the LSDV has the advantage that it produces estimates of $(c_1, c_2, ..., c_N)$, which are the OLS coefficients of the individual-specific dummy variables.
- For example, we may be interested in evaluating the management quality of a particular firm and comparing it with the mean.

- By using STATA to run a linear regression using the transformed data, the standard error we get is

$$\tilde{\sigma}_U^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it}^2 \right)^{-1}$$

  where $\tilde{\sigma}_U^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{U}_{it}^2 / (NT-1)$ and $\hat{U}_{it} = \dot{Y}_{it} - \hat{\beta}^{FE} \dot{X}_{it}$.
- We can show that $\tilde{\sigma}_U^2 \rightarrow_p \sigma_U^2 \left(1 - T^{-1}\right)$. Therefore the standard error is not valid and has to be corrected.
- In the presence of serial correlation and heteroskedasticity, the asymptotic variance can not be estimated by $\hat{\sigma}_U^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it}^2 \right)^{-1}$. A more complicated estimator that is robust to serial correlation and heteroskedasticity has been proposed in the literature.

Time effects and two-way fixed effect estimator

- We extend the baseline model by adding the time effect:

$$Y_{it} = \alpha_t + \beta X_{it} + c_i + U_{it}.$$

- Two different views:
  - $\{\alpha_t : t = 1, ..., T\}$ is viewed as intercepts for different periods. The model allows different time periods to have different intercepts. As regression with pooled cross section, we use time-periods dummies:

    $$Y_{it} = \alpha_1 + \alpha_2 D_t^T + \cdots + \alpha_T D_t^T + \beta X_{it} + c_i + U_{it},$$

    where $D_t^s = 1$ if $t = s$ and 0 otherwise. Then we apply the one-way within transformation or use individual-specific dummy variables.
  - The model has two-way error component $\alpha_t + c_i + U_{it}$. In this case, α_t is the unobserved time-specific effect. We use two-way within transformation to eliminate α_t from the model, as we did for c_i. The resulting estimator is usually called two-way fixed effect estimator.
  - These two approaches produce numerically identical estimates.

Two-way within transformation

- Average the equation

$$Y_{it} = \alpha_t + \beta X_{it} + c_i + U_{it} \tag{1}$$

over $t = 1, \dots, T$ to get

$$\overline{Y}_i = \overline{\alpha} + \beta \overline{X}_i + c_i + \overline{U}_i,$$

where $\overline{Y}_i = T^{-1} \sum_{t=1}^{T} Y_{it}$, $\overline{X}_i = T^{-1} \sum_{t=1}^{T} X_{it}$, $\overline{U}_i = T^{-1} \sum_{t=1}^{T} U_{it}$ and $\overline{\alpha} = T^{-1} \sum_{t=1}^{T} \alpha_t$.

- Average (1) over $i = 1, \dots, N$ to get

$$\widetilde{Y}_t = \alpha_t + \beta \widetilde{X}_t + \widetilde{c} + \widetilde{U}_t,$$

where $\widetilde{Y}_t = N^{-1} \sum_{i=1}^{N} Y_{it}$, $\widetilde{c} = N^{-1} \sum_{i=1}^{N} c_i$ and $\widetilde{U}_t = N^{-1} \sum_{i=1}^{N} U_{it}$.

- Average 1 over both $t = 1, \dots, T$ and $i = 1, \dots, N$ to get

$$\overline{Y} = \overline{\alpha} + \beta \overline{X} + \widetilde{c} + \overline{U},$$

where $\overline{Y} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{it}$, $\overline{X} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} X_{it}$ and $\overline{U} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} U_{it}$.

- Then,

$$\begin{aligned}
Y_{it} - \overline{Y}_i - \widetilde{Y}_t + \overline{Y} &= (\alpha_t + \beta X_{it} + c_i + U_{it}) - \left(\overline{\alpha} + \beta \overline{X}_i + c_i + \overline{U}_i\right) \\
&\quad - \left(\alpha_t + \beta \widetilde{X}_t + \widetilde{c} + \widetilde{U}_t\right) + \left(\overline{\alpha} + \beta \overline{X} + \widetilde{c} + \overline{U}\right) \\
&= \beta \left(X_{it} - \overline{X}_i - \widetilde{X}_t + \overline{X}\right) + \left(U_{it} - \overline{U}_i - \widetilde{U}_t + \overline{U}\right) \\
&\implies \ddot{Y}_{it} = \beta \ddot{X}_{it} + \ddot{U}_{it},
\end{aligned}$$

where $\ddot{Y}_{it} = Y_{it} - \overline{Y}_i - \widetilde{Y}_t + \overline{Y}$, $\ddot{X}_{it} = X_{it} - \overline{X}_i - \widetilde{X}_t + \overline{X}$ and $\ddot{U}_{it} = U_{it} - \overline{U}_i - \widetilde{U}_t + \overline{U}$.

- We regress the "two-way within-transformed" variables $\ddot{Y}_{it}$ on $\ddot{X}_{it}$ (with the panel structure ignored) without an intercept to obtain the two-way fixed effect estimator

$$\hat{\beta}^{TWFE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it} \ddot{Y}_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it}^2}.$$

First differencing when $T = 2$

- When $T = 2$, for individual i, write the two periods as

$$\begin{aligned} Y_{i1} &= \alpha_1 + \beta X_{i1} + c_i + U_{i1} \\ Y_{i2} &= \alpha_2 + \beta X_{i2} + c_i + U_{i2}. \end{aligned}$$

- Denote $\delta = \alpha_2 - \alpha_1$, $\Delta Y_i = Y_{i2} - Y_{i1}$, $\Delta X_i = X_{i2} - X_{i1}$ and $\Delta U_i = U_{i2} - U_{i1}$. Subtract the second equation from the first,

$$\Delta Y_i = \delta + \beta \Delta X_i + \Delta U_i.$$

- The unobserved individual effect c_i is differenced away. It is easy to see that $\mathrm{E}[\Delta U_i] = \mathrm{E}[\Delta U_i \cdot \Delta X_i] = 0$ and therefore regression of ΔY_i against ΔX_i consistently estimates β.
- It can be shown that this estimator is numerically identical to the two-way fixed effect estimator.

Baseline random effect model

- Assume the baseline model with only one time-varying explanatory variable X_{it}

$$Y_{it} = \alpha + \beta X_{it} + c_i + U_{it}$$

  and no time effect.
- c_i and U_{it} are unobserved random variables. The set of model assumptions is the same as the fixed effect model. In addition, we assume $\mathrm{E}[X_{it} c_i] = 0$.
- We can ignore the panel structure. Denote $V_{it} = c_i + U_{it}$. Since V_{it} is uncorrelated with X_{it}, the pooled OLS is consistent.
- However, the pooled OLS is inefficient. It is easy to check that the error terms are serially correlated:

$$\mathrm{Corr}[V_{it}, V_{is}] = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_U^2}, \forall s \neq t,$$

  where $\sigma_c^2 = \mathrm{E}\left[c_i^2\right]$ and $\sigma_U^2 = \mathrm{E}\left[U_{it}^2\right]$.

- The random effect estimator uses the information about the special form of serial correlation and is more efficient. It effectively uses transformed data.
- Denote $\theta^2 = \sigma_U^2 / \left(\sigma_U^2 + T\sigma_c^2\right)$. Transform the model:

$$Y_{it}^* = \alpha^* + \beta X_{it}^* + V_{it}^*,$$

  where $Y_{it}^* = Y_{it} - (1-\theta)\overline{Y}_i$, $X_{it}^* = X_{it} - (1-\theta)\overline{X}_i$, $c_i^* = \theta c_i$, $\alpha^* = \theta\alpha$, $U_{it}^* = U_{it} - (1-\theta)\overline{U}_i$ and $V_{it}^* = c_i^* + U_{it}^*$.
- Note that this transformation does not eliminate the individual effect. And the error terms are uncorrelated: $\text{Corr}\left[V_{it}^*, V_{is}^*\right] = 0$, $\forall s \neq t$.
- The random effect estimation replaces σ_U^2 and σ_c^2 (and thus, θ) by their consistent estimators. The random effect estimator $\hat{\beta}^{RE}$ is the pooled OLS estimator with estimated $\left(Y_{it}^*, X_{it}^*\right)$ ($i = 1, \ldots, n$, $t = 1, \ldots, T$).
- In applications, we add time-periods dummies and time-invariant explanatory variables. One advantage of the random effect model is that effects of time-invariant explanatory variables can be estimated.

Fixed effect versus random effect and Hausman test

- The random effect estimator is efficient, if $\mathrm{E}[X_{it} c_i] = 0$ is satisfied. The fixed effect estimator is consistent, whether $\mathrm{E}[X_{it} c_i] = 0$ is satisfied or not.
- Hausman test compares these two estimators and tests $\mathrm{H}_0 : \mathrm{E}[X_{it} c_i] = 0$ versus $\mathrm{H}_1 : \mathrm{E}[X_{it} c_i] \neq 0$.
- We can show that under H_0, $\sqrt{n}\left(\hat{\beta}^{FE} - \beta\right) \to_d \mathrm{N}\left(0, \sigma^2_{FE}\right)$ and $\sqrt{n}\left(\hat{\beta}^{RE} - \beta\right) \to_d \mathrm{N}\left(0, \sigma^2_{RE}\right)$ with $\sigma^2_{RE} < \sigma^2_{FE}$. Under H_0, both of $\hat{\beta}^{FE}$ and $\hat{\beta}^{RE}$ are consistent and asymptotically normal.
- It can be shown that under H_0,

$$\sqrt{n}\left(\hat{\beta}^{FE} - \hat{\beta}^{RE}\right) \to_d \mathrm{N}\left(0, \sigma^2_{FE} - \sigma^2_{RE}\right).$$

- The Hausman test rejects H_0 at 5% significance level if

$$\left| \frac{\hat{\beta}^{FE} - \hat{\beta}^{RE}}{\sqrt{\left(\hat{\sigma}^2_{FE} - \hat{\sigma}^2_{RE}\right)/n}} \right| > 1.96,$$

  where $\left(\hat{\sigma}^2_{FE}, \hat{\sigma}^2_{RE}\right)$ are consistent estimators of $\left(\sigma^2_{FE}, \sigma^2_{RE}\right)$.

Correlated random effect model

- The fixed effect model is more robust than the random effect model since it allows nonzero correlation between the unobserved individual effect and explanatory variables. But we cannot use this model to estimate the effects of time-invariant variables.
- Rather than using the within transformation to eliminate c_i, the correlated random effect model (Mundlak (1978)) explicitly specifies the dependence of c_i on $X_{i1}, ..., X_{iT}$:

$$c_i = \lambda_1 X_{i1} + \cdots + \lambda_T X_{iT} + \eta_i,$$

  where $\mathrm{E}\left[X_{it}\eta_i\right] = 0, \forall t$ and $(\lambda_1, ..., \lambda_T)$ are linear projection coefficients.
- Correlated random effect model assumes that the projection coefficients are all the same: $\lambda_1 = \cdots = \lambda_T = \lambda$. Then, $c_i = \lambda\left(T\overline{X}_i\right) + \eta_i$ and

$$Y_{it} = \alpha + \beta X_{it} + \lambda\left(T\overline{X}_i\right) + \eta_i + U_{it}.$$

- The model satisfies the random effect assumption: $\mathrm{E}\left[\eta_i X_{it}\right] = \mathrm{E}\left[\eta_i \overline{X}_i\right] = 0$. The correlated random effect estimator is a random effect estimator with a new time-invariant variable $T\overline{X}_i$.
- Chamberlain (1984)'s approach is more flexible: $(\lambda_1, ..., \lambda_T)$ can be different but its estimation is more sophisticated.
- One advantage of the correlated random effect model is that it produces estimates of the effects of the time-invariant variables and allows nonzero correlation between the individual effect and time-varying explanatory variables.
- However, this model rules out nonzero correlation between the individual effect and time-invariant variables. If some of the time-invariant variables are endogenous, then these estimators are inconsistent.

Hausman-Taylor's method

- Hausman-Taylor's method addresses this issue under alternative model assumptions.
- The model has k time varying variables and m time invariant variables:
$$Y_{it} = \alpha + X_{it}^\top \beta + Z_i^\top \gamma + c_i + U_{it}.$$
- Divide: $X_{it} = \begin{pmatrix} X_{1,it} \\ X_{2,it} \end{pmatrix}$ and $Z_i = \begin{pmatrix} Z_{1,i} \\ Z_{2,i} \end{pmatrix}$. $X_{1,it}$ is a vector of k_1 variables and $X_{2,it}$ is a vector of k_2 variables, $k = k_1 + k_2$. $Z_{1,i}$ is a vector of m_1 variables and $Z_{2,i}$ is a vector of m_2 variables, $m = m_1 + m_2$.
- $X_{1,it}$ and $Z_{1,i}$ are uncorrelated with c_i but $X_{2,it}$ and $Z_{2,i}$ are correlated with c_i.

- Take averages in the time dimension:

$$\overline{Y}_i - \overline{X}_i^\top \beta = \alpha + Z_i^\top \gamma + c_i + \overline{U}_i.$$

  In this group-level model, $Z_{2,i}$ is correlated with the error term $V_i = c_i + \overline{U}_i$.
- β can be consistently estimated by the fixed effect estimator $\hat{\beta}^{FE}$.
- Hausman-Taylor's method estimates the dependent variable by $\overline{Y}_i - \overline{X}_i^\top \hat{\beta}^{FE}$ and uses $\overline{X}_{1,i}$ as instruments for $Z_{2,i}$.
- For identification, we require $k_1 \geq m_2$ and $X_{1,it}$ are correlated with $Z_{2,i}$.
- Then we can apply 2SLS or GMM estimation.

Baseline dynamic panel model

- Consider the following model with $X_{it} = Y_{i(t-1)}$:

$$Y_{it} = \beta Y_{i(t-1)} + c_i + U_{it},$$

  where the initial outcome Y_{i0} is observed. We assume that the idiosyncratic errors have no serial correlation: $\mathrm{E}[U_{it}U_{is}] = 0$, $\forall s \neq t$.
- In this model, $X_{it} = Y_{i(t-1)}$ must be correlated with c_i since

$$Y_{i(t-1)} = \beta Y_{i(t-2)} + c_i + U_{i(t-1)}$$

- The assumption used by the fixed effect approach

$$\mathrm{E}[U_{it}] = \mathrm{E}[U_{it}X_{i1}] = \cdots = \mathrm{E}[U_{it}X_{iT}] = 0, \forall t = 1, ..., T,$$

  does not hold, since U_{it} is correlated with $X_{i(t+1)} = Y_{it}$.
- The within-transformed X_{it} is correlated with the within-transformed U_{it}. So the standard one-way fixed effect estimation is inconsistent.

Anderson-Hsiao estimator

- Anderson-Hsiao's approach considers the first difference:

$$\Delta Y_{it} = \beta \Delta Y_{i(t-1)} + \Delta U_{it}, t = 1, ..., T, i = 1, ..., N,$$

  where $\Delta Y_{it} = Y_{it} - Y_{i(t-1)}$, $\Delta Y_{i(t-1)} = Y_{i(t-1)} - Y_{i(t-2)}$ and $\Delta U_{it} = U_{it} - U_{i(t-1)}$.
- If the error U_{it} is uncorrelated with past Y_{it}'s: $\mathrm{E}\left[U_{it} Y_{i(t-1)}\right] = 0$, then it can be observed that $Y_{i(t-2)}$ is a valid instrument for $\Delta Y_{i(t-1)}$: $\mathrm{E}\left[Y_{i(t-2)} \Delta U_{it}\right] = 0$ and $\mathrm{E}\left[Y_{i(t-2)} \Delta Y_{i(t-1)}\right] \neq 0$. This requires that we observe data from at least three periods.
- The Anderson-Hsiao estimator is an IV estimator for the first-differenced model using $Y_{i(t-2)}$ as the instrument.
- $\Delta Y_{i(t-2)} = Y_{i(t-2)} - Y_{i(t-3)}$ is also a valid instrument, if one more time period is observed.

Arrellano-Bond estimator

- Arrellano-Bond's GMM estimator is more efficient than Anderson-Hsiao's IV estimator. It exploits all valid moment conditions the model generates.
- Suppose that $T = 4$. We know that the following moment conditions hold.
  - For $t = 2$,
  $$\mathrm{E}\left[\left(\Delta Y_{i2} - \beta \Delta Y_{i1}\right) Y_{i0}\right] = 0;$$
  - For $t = 3$,
  $$\begin{aligned} \mathrm{E}\left[\left(\Delta Y_{i3} - \beta \Delta Y_{i2}\right) Y_{i0}\right] &= 0 \\ \mathrm{E}\left[\left(\Delta Y_{i3} - \beta \Delta Y_{i2}\right) Y_{i1}\right] &= 0; \end{aligned}$$
  - For $t = 4$,
  $$\begin{aligned} \mathrm{E}\left[\left(\Delta Y_{i4} - \beta \Delta Y_{i3}\right) Y_{i0}\right] &= 0 \\ \mathrm{E}\left[\left(\Delta Y_{i4} - \beta \Delta Y_{i3}\right) Y_{i1}\right] &= 0 \\ \mathrm{E}\left[\left(\Delta Y_{i4} - \beta \Delta Y_{i3}\right) Y_{i2}\right] &= 0. \end{aligned}$$

- Arrellano-Bond's efficient GMM estimator is $\text{argmin}_b Q(b)$, where

$$Q(b) =$$
$$\begin{pmatrix} N^{-1}\sum_{i=1}^{N}(\Delta Y_{i2} - b\Delta Y_{i1})Y_{i0} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i3} - b\Delta Y_{i2})Y_{i0} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i3} - b\Delta Y_{i2})Y_{i1} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i4} - b\Delta Y_{i3})Y_{i0} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i4} - b\Delta Y_{i3})Y_{i1} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i4} - b\Delta Y_{i3})Y_{i2} \end{pmatrix}^{\top} \widehat{W}^{*} \begin{pmatrix} N^{-1}\sum_{i=1}^{N}(\Delta Y_{i2} - b\Delta Y_{i1})Y_{i0} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i3} - b\Delta Y_{i2})Y_{i0} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i3} - b\Delta Y_{i2})Y_{i1} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i4} - b\Delta Y_{i3})Y_{i0} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i4} - b\Delta Y_{i3})Y_{i1} \\ N^{-1}\sum_{i=1}^{N}(\Delta Y_{i4} - b\Delta Y_{i3})Y_{i2} \end{pmatrix}$$

and $\widehat{W}^{*}$ is the estimated efficient GMM weighting matrix.