Dynamic Panel Data Models

Jakub Mućk
SGH Warsaw School of Economics

The First-Difference (FD) estimator

The First-Difference (FD) estimator I

- The First-Difference (FD) estimator is an alternative estimation technique that eliminates the fixed effect as well as time invariant regressors.
- Note that

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it} \quad \text{for} \quad t = 1, \ldots T, \tag{1}$$

$$y_{it-1} = \alpha_i + \beta_1 x_{1it-1} + \ldots + \beta_k x_{kit-1} + u_{it-1} \quad \text{for} \quad t = 2, \ldots T, \tag{2}$$

and differencing both equations yields:

$$\Delta y_{it} = \beta_1 \Delta x_{1it} + \ldots + \beta_k \Delta x_{kit} + \Delta u_{it}, \tag{3}$$

where Δ is the well-known (from time series analysis) first-difference operator, i.e. $\Delta z_t = z_t - z_{t-1}$.

- The parameters in (3) can be estimated with the least squares. In the matrix form:

$$\beta^{FD} = \left(\Delta X' \Delta X\right)^{-1} \Delta X' \Delta y. \tag{4}$$

- The estimates of fixed effects can be also recovered:

$$\hat{\alpha}_i^{FD} = \bar{y}_i - \bar{x}_i \hat{\beta}^{FD}. \tag{5}$$

- If the error term in (3) is not correlated with independent variable (weak exogeneity) then the least squares estimator is unbiased and consistent.

- The above assumption is less restrictive than in standard FE model:

$$\mathbb{E}\left(\Delta u_{it}|\Delta x_{it}\right) = \mathbb{E}\left(u_{it} - u_{it-1}|x_{it} - x_{it-1}\right) = 0. \tag{6}$$

- The FE estimator is more efficient when the disturbances are not serially correlated and homoskedastic.
  - But If u_{it} is driven by random walk (autocorrelation with $\rho = 1$) then the FD estimator is more efficient.

Dynamic panel data models

- Dynamic linear panel data model:

$$y_{it} = \gamma y_{it-1} + x_{it}'\beta + u_{it}, \tag{7}$$

where

- $u_{it} = \mu_i + \varepsilon_{it}$ and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$,
- γ is the autoregressive parameter,
- y_{it-1} is the lagged dependent variable,
- x_{it} is the vector of independent variables.

- Dynamic linear panel data model:

$$y_{it} = \gamma y_{it-1} + x'_{it}\beta + u_{it}, \tag{7}$$

where

- $u_{it} = \mu_i + \varepsilon_{it}$ and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$,
- γ is the autoregressive parameter,
- y_{it-1} is the lagged dependent variable,
- x_{it} is the vector of independent variables.

- Remarks:
  - We assume that y_{it} is the stable (conditional on x_{it}) process $\implies |\gamma| < 1$. In other words, the effect of idiosyncratic shock (ε_{it}) dies out.
  - The independent variables (x_{it}) are assumed to be strictly exogenous.
  - μ_i is the individual-specific (random or fixed) effect.
  - Each observation can be written as:

$$y_{it} = \gamma^t y_{i0} + \sum_{j=0}^{t} \gamma^j \beta' x_{it-j} + \frac{1-\gamma^t}{1-\gamma}\mu_i + \sum_{j=0}^{t-1} \gamma^j u_{it-j}, \tag{8}$$

where y_{i0} is the (non-stochastic) initial value.

- The demeaning transformation used to get the within estimator *new* creates new independent variables that are correlated with the error term. As a result, the standard OLS estimator is inconsistent.
- General intuition:
  - The within estimator for the panel AR(1) model:

$$y_{it} - \bar{y}_i = (\mu_i - \mu_i) + \gamma \left(y_{it-1} - \bar{y}_{i-1} \right) + \left(\varepsilon_{it} - \bar{\varepsilon}_i \right), \tag{9}$$

    where $\bar{y}_{i-1} = 1/(T-1) \sum_{t=2}^{T} y_{it-1}$.
  - The mean of the lagged dependent variable $(\bar{y}_{i-1})$ is correlated with $\bar{\varepsilon}_i$ even if the error term is not autocorrelated. The average $\bar{\varepsilon}_i$ contains the lagged error term ε_{it-1} and, therefore, it is correlated with y_{it-1}.
- Taking the probability limit (plim) of the FE estimator (as $N \to \infty$):

$$\text{plim}\hat{\gamma}^{FE} = \gamma + \frac{\frac{1}{NT} \left(y_{it-1} - \bar{y}_{i-1} \right) \left(\varepsilon_{it} - \bar{\varepsilon}_i \right)}{\frac{1}{NT} \left(y_{it-1} - \bar{y}_{i-1} \right)^2} \tag{10}$$

  it can be observed that the correlations between the lagged dependent variable $(\bar{y}_{i-1})$ and error term will lead to inconsistency of the OLS estimator.

- **Nickell's (1981) bias**. The small T bias of the FE estimator as $N \to \infty$:

$$\text{plim}\left(\hat{\gamma}^{FE} - \gamma\right) = -\frac{(1+\gamma)}{T}\left(1 - \frac{1}{T}\frac{1-\gamma^T}{1-\gamma}\right)\left[1 - \frac{1}{T} - \frac{2\gamma}{(1-\gamma)\,T}\left(1 - \frac{1}{T}\frac{1-\gamma^T}{1-\gamma}\right)\right]^{-1} \tag{11}$$

- The bias of the FE estimator depends on T as well as γ.
- For *reasonably* large T it can be approximated:

$$\text{plim}\left(\hat{\gamma}^{FE} - \gamma\right) \approx -\frac{(1+\gamma)}{T-1} \tag{12}$$

- but when $T = 2$ then

$$\text{plim}\left(\hat{\gamma}^{FE} - \gamma\right) \approx -\frac{(1+\gamma)}{2} \tag{13}$$

- The bias in the dynamic fixed effect model is caused by elimination of the individual-specific effect from each observation. It creates a correlation of order $1/T$ between explanatory variables and error term.

- Consider the RE AR(1) model:

$$y_{it} = \gamma y_{it-1} + \varepsilon_{it} + \mu_i, \tag{14}$$

where $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ and $\mu_{it} \sim \mathcal{N}(0, \sigma^2_\mu)$.

- In the RE model, the quasi-demeaning also leads to correlation between the transformed lagged dependent variable ($\tilde{y}_{it-1} = y_{it-1} - \theta \bar{y}_{i-1}$) and the transformed error term ($\tilde{\varepsilon}_{it} = \varepsilon_{it} - \theta \bar{\varepsilon}_i$). Therefore, the RE estimates will be biased.
- For $t-1$ the dependent variable:

$$y_{it-1} = \gamma y_{it-2} + \varepsilon_{it-1} + \mu_i, \tag{15}$$

also depends on the random individual-specific effect. If so, then the assumption that the individual effects are independent of the explanatory variable (in our case also y_{it-1}) is not satisfied and

$$\mathbb{E}\left(\mu_i | y_{it-1}\right) \neq 0. \tag{16}$$

- **The FD AR(1) estimator**:

$$y_{it} - y_{it-1} = (\mu_i - \mu_i) + \gamma\left(y_{it-1} - y_{it-2}\right) + \varepsilon_{it} - \varepsilon_{it-1} \tag{17}$$

  is also biased.

- To illustrate the bias of the FD estimator it's useful to recall $y_{i,t-1}$.

$$y_{it-1} = \gamma y_{it-2} + \varepsilon_{it-1} + \mu_i + \varepsilon_{it-1}. \tag{18}$$

- In the (18) y_{it-1} depends on the error term ε_{it-1}. At the same time, in the (17) y_{it-1} is the explanatory variable and the error term, given by $\varepsilon_{it} - \varepsilon_{it-1}$, contains the lagged error term from the non-transformed model. Therefore, the lagged dependent variable is correlated with the error term also in the FD model.

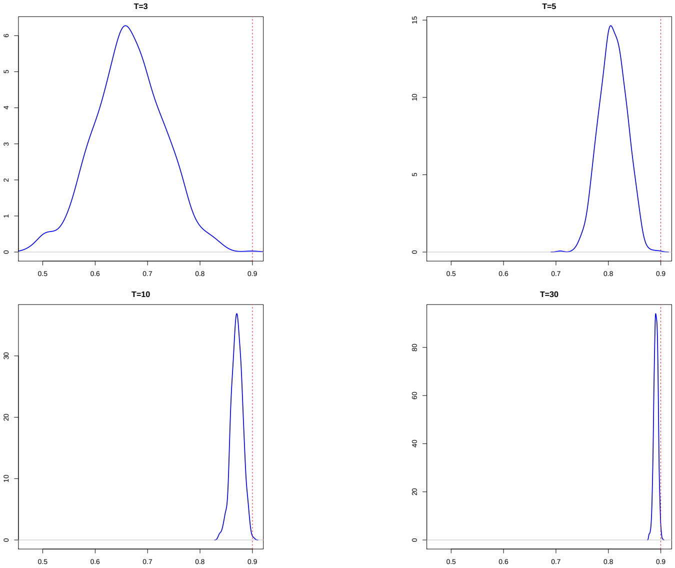
- To illustrate the magnitude of the Nickell's bias we run the MC simulations.
- Let's assume that the true DGP (data generating process) is a simple panel AR(1) process:

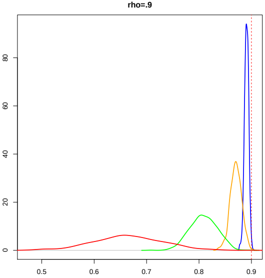
$$y_{it} = \gamma y_{it-1} + u_{it}, \tag{19}$$

where the error component is quite standard:

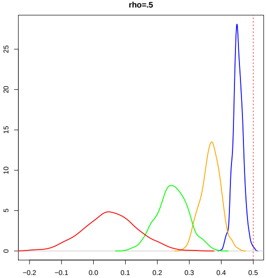
$$u_{it} = \mu_i + \varepsilon_{it} \tag{20}$$

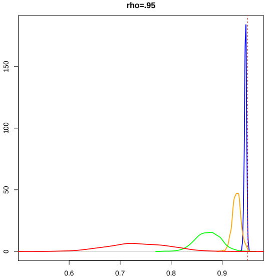
where $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$ and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.
- We will consider the FE estimator.
- The MC settings:
  - $\gamma = 0.9$ (in the second exercise also 0.5 and 0.95)
  - $T \in \{3, 5, 10, 30\}$.
  - $\sigma_\mu = 0.5$ and $\sigma_\varepsilon = 0.25$.
  - $N = 100$ (the cross-sectional dimension).
  - 1000 replications.


T=3
T=5
T=10
T=30
0.5
0.6
0.7
0.8
0.9
0
1
2
3
4
5
6
10
15
20
30
40
60
80


rho=.9
0
20
40
60
80
0.5
0.6
0.7
0.8
0.9

T=3; T=5; T=10; T=30.


rho=.5
0
5
10
15
20
25
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
0.5


rho=.95
0
50
100
150
0.6
0.7
0.8
0.9

- The standard estimators (FE, RE, FD) fail to account for dynamics in the dynamic panel data models. This is due to the fact that the lagged dependent variable becomes endogenous (correlated with error term).
- The dynamic panel data (DPD) models are designed to account for this endogeneity.
- It is important when T is relatively small $\implies$ micro data.
  - When T is large the Nickell's bias is relatively small. Which T is sufficiently large to ignore the Nickell's bias?

The Anderson and Hsiao estimator

The Anderson and Hsiao estimator I

SGH

- Anderson and Hsiao (1981) propose estimator that simply uses the IV.
- Starting point: the FD estimator:

$$\Delta y_{ti} = \gamma \Delta y_{ti-1} + \beta_1 \Delta x_{1ti} + \ldots + \beta_k \Delta x_{kti} + \Delta \varepsilon_{it}. \quad (21)$$

- Problem: Δy_{ti-1} is correlated with the error term $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$.
- Use twice lagged level of dependent variable y_{it-2} as an instrument for Δy_{ti-1}. By construction, y_{it-2} is not correlated with the error term $\Delta \varepsilon_{it}$ but is correlated with endogenous variable, i.e. Δy_{ti-1}.
- In general, one might use the twice lagged differences $\Delta y_{ti-2} = y_{ti-2} - y_{ti-3}$ as a valid instrument for endogenous variable Δy_{ti-1}. But:
  - Using y_{ti-2} as the instrumental variable $\Longrightarrow$ more data.
  - Using Δy_{ti-2} as the instrumental variable $\Longrightarrow$ larger asymptotic variance of estimator.
- The AH estimator delivers consistent but not efficient estimates of the parameters in the model. This is due to the fact that the IV doesn't exploit all the available moments conditions.
- The IV estimator also ignores the structure of the error component in the transformed model.
  - The autocorrelation in the first differences errors leads to inconsistency of the IV estimates.
- The IV estimates would be inconsistent when other regressors are correlated with the error term.

Generalized Method of Moments (GMM)

- The standard classical methods, e.g., the Maximum Likelihood (ML) method, requires a complete specification of the model that is considered to be estimated. This includes also the probability of distribution of the variable of interest.
- Contrary to the ML method, the **Generalized Method of Moments (GMM)** requires only a set moment conditions that are implied by assumption of the underlying econometric model. The GMM method is attractive when:
  - there is a variety of moment or orthogonality conditions that are deduced from the assumption of the theoretical model;
  - the economic model is complex, i.e., it's difficult to write down a tractable and applicable likelihood function,
  - to overcome the computational complexities associated with the ML estimator.

- Let's assume that a sample of T observations is drawn from the joint probability distribution:

$$f\left(w_1, w_2, \ldots, w_T, \theta_0\right) \tag{22}$$

where θ_0 is the $(q \times 1)$ vector of true parameters and w_t contains one or more endogenous and/or exogenous variables.

- Population moments condition:

$$\mathbb{E}\left[m\left(w_T, \theta_0\right)\right] = 0, \quad \text{for all } t. \tag{23}$$

- where $m(\cdot)$ is the r-dimensional vector of functions.
- Three cases:
  1. $q > r \implies$ the parameters in θ are **not identified**;
  2. $q = r \implies$ the parameters in θ are **exactly identified**;
  3. $q < r \implies$ the parameters in θ are **overidentified** and the moments conditions have to be restricted in order to deliver a unique θ in estimation. This can be done by the means of a weighting matrix (A_T).
- Estimation bases on the empirical counterpart of $\mathbb{E}\left[m\left(w_T, \theta_0\right)\right]$:

$$M_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} m\left(w_T, \theta_0\right), \tag{24}$$

where $M_T(\theta)$ is the r-dimensional vector of sample moments.

- **Linear regression:**
  - Consider the standard linear regression:
  $$y_t = x_t'\beta + \varepsilon_t \tag{25}$$
  Under the standard (classical) assumption, the population conditions is following:
  $$\mathbb{E}\left(x_t, \varepsilon_t\right) = \mathbb{E}\left[x_t, \left(y_t - x_t'\beta\right)\right] = 0 \quad \text{for } t \in 1, \ldots, T. \tag{26}$$
- **Linear regression with endogenous variables:**
  - Consider the standard linear regression with endogenous variables:
  $$y_t = x_t'\beta + \varepsilon_t \tag{27}$$
  where $\mathbb{E}(x_t, \varepsilon_t) \neq 0$.
  - The population conditions:
  $$\mathbb{E}\left(z_t, \varepsilon_t\right) = \mathbb{E}\left[z_t, \left(y_t - x_t'\beta\right)\right] = 0 \quad \text{for } t \in 1, \ldots, T \tag{28}$$
  where z_T is the set of the instrumental variables that satisfies the above orthogonality conditions.

- **The GMM estimator** of θ bases on:

$$\hat{\theta}_T = \text{argmin}_{\theta \in \Theta} \left\{ M'_T(\theta) A_T M_T(\theta) \right\}, \tag{29}$$

where A_T is a $r \times r$ positive semi-define, possibly random weighting matrix.

- We wish to choose the weighting matrix that minimizes the covariance matrix of $\hat{\theta}$.
  - This provides the efficient estimator. Other weighting matrices would lead to less efficient estimators of θ.
- **The general instrumental variable estimator (GIVE)** combines all available instruments to estimates the unknown parameters. In this case, **the number of instruments can be larger than number of parameters to estimate ($r > k$).**
- The starting point: the r population conditions:

$$\mathbb{E}\left(z_t, \varepsilon_t\right) = \mathbb{E}\left[z_t, \left(y_t - x'_t\beta\right)\right] = 0 \quad \text{for } t \in 1, \ldots, T \tag{30}$$

where z_t is the set of (r) instruments, x_t is the k-dimensional vector of regressors. The regressors are endogenous, i.e., $\mathbb{E}\left(x_t, \varepsilon_t\right) \neq 0$ while the error term is idiosyncratic, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2_\varepsilon)$. The instruments are correlated with x_t but not correlated with the error term.

The GMM and GIVE estimators II

SGH

- This implies the following sample moments:

$$M_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} z_t \left(y_t - \beta' x_t\right). \tag{31}$$

- It can be shown that the GIVE estimator is given as:

$$\hat{\beta}^{GIVE} = \left(X' P_z X\right)^{-1} X' P_z Y, \tag{32}$$

where $P_z = Z\left(Z'Z\right)^{-1} Z'$. The matrix Z collects all instruments, the matrix X stands for the regressors while Y denotes the observations of the dependent variable.

- The estimator of the variance matrix of $\hat{\beta}^{GIVE}$ is as follows:

$$Var\left(\hat{\beta}^{GIVE}\right) = \hat{\sigma}^2_{GIVE}\left(X' P_z X\right)^{-1}. \tag{33}$$

where the estimated variance of the error term bases on the variance of the residuals from the considered regression:

$$\hat{\sigma}^2_{GIVE} = \frac{1}{T-K}\hat{\varepsilon}'_{GIVE}\hat{\varepsilon}_{GIVE}. \tag{34}$$

- In analogous fashion to the basic linear models, the robust standard error (e.g. hetereoskedasticity-consistent) can be computed.

Sargan's general test for misspecification

SGH

- In the GIVE estimation we use r instruments. Are these instruments valid?
- Consider following test statistics:

$$\chi^2_{SM} = \frac{Q(\hat{\beta}^{GIVE})}{\hat{\sigma}^2_{GIVE}}, \tag{35}$$

where

$$Q(\hat{\beta}^{GIVE}) = \left(y - X\hat{\beta}^{GIVE}\right)' P_z \left(y - X\hat{\beta}^{GIVE}\right) \tag{36}$$

- Under the null the regression is correctly specified and the r instruments Z are valid instruments.
- Sargan's misspecification statistics is χ^2 distributed with $r - k$ degrees of freedom.

The Arellano Bond estimator

The Arellano Bond estimator

- Arellano and Bond (1991) suggest using a GMM approach based on all available conditions.
- Starting point: the FD estimator:

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \beta' \Delta x_{it} + \Delta \varepsilon_{it} \quad (37)$$

Valid instruments:

- [t=2 or t=1]: no instruments,
- [t=3]: the valid instrument for $\Delta y_{i2} = (y_{i2} - y_{i1})$ is y_{i1},
- [t=4]: the valid instruments for $\Delta y_{i3} = (y_{i3} - y_{i2})$ is y_{i2} as well as y_{i1},
- [t=5]: the valid instruments for $\Delta y_{i4} = (y_{i4} - y_{i3})$ is y_{i3} as well as y_{i2} and y_{i1},
- [t=6]: the valid instruments for $\Delta y_{i5} = (y_{i5} - y_{i4})$ is y_{i4} as well as y_{i3}, y_{i2} and y_{i1},
- [t=T]: the valid instruments for $\Delta y_{iT-1} = (y_{iT-1} - y_{iT-2})$ is y_{iT-2} as well as $y_{iT-3}, \ldots, y_{i1}$.

- Hence, there is a total of $(T-1)(T-2)/2$ available instruments or moment conditions for Δy_{it-1}. In general, it can be written as:

$$\mathbb{E}\left[y_{is}\left(\Delta y_{it} - \gamma \Delta y_{it-1} - \beta' \Delta x_{it}\right)\right] = 0 \quad \text{for } s = 0, \ldots, t-2 \text{ and } t = 2, \ldots, T \quad (38)$$

- Consider the following specification:

$$\Delta y_{i.} = \gamma \Delta y_{i.-1} + \Delta X_{i.}\beta + \Delta \varepsilon_{i.}, \tag{39}$$

where

$$\Delta y_{i.} = \begin{bmatrix} \Delta y_{i2} \\ \Delta y_{i3} \\ \vdots \\ \Delta y_{iT} \end{bmatrix}, \Delta y_{i.-1} = \begin{bmatrix} \Delta y_{i1} \\ \Delta y_{i2} \\ \vdots \\ \Delta y_{iT-1} \end{bmatrix}, \Delta X_{i.} = \begin{bmatrix} \Delta x'_{i2} \\ \Delta x'_{i3} \\ \vdots \\ \Delta x'_{iT} \end{bmatrix}, \Delta \varepsilon_{i.} = \begin{bmatrix} \Delta \varepsilon_{i2} \\ \Delta \varepsilon_{i3} \\ \vdots \\ \Delta \varepsilon_{iT} \end{bmatrix}.$$

- The corresponding matrix of instruments for the lagged difference:

$$W_i = \begin{bmatrix} y_{i1} & 0 & \dots & 0 \\ 0 & y_{i1}, y_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & y_{i1}, y_{i2}, \dots, y_{iT-2} \end{bmatrix},$$

- Then the moment conditions can be described as:

$$\mathbb{E}\left[W_i' \Delta \varepsilon_{i.}\right] = 0 \tag{40}$$

- Finally, the GMM estimator that takes into account the formulated moment conditions can be applied:

$$\hat{\lambda}^{GMM} = \left(G'ZS_NZ'G\right)^{-1} G'ZS_NZ'\Delta y \tag{41}$$

  where
  - $\hat{\lambda}^{GMM} = \left[\hat{\gamma}^{GMM}\hat{\beta}^{GMM}\right]'$,
  - $G = (\Delta y_{-1}, \Delta X)$,
  - $Z = (W, \Delta X)$.
  - S_N is the optimal weighting matrix.
- The matrix S_N is usually calculated from initial estimates, e.g., IV estimates.

$$S_N = \left(\sum_{i=1}^{N} Z'_{i.}\hat{e}_{i.}\hat{e}'_{i.}Z_{i.}\right)^{-1}, \tag{42}$$

  where $\hat{e}_{i.}$ stands for the residuals from the initial estimates.
- The above procedure refers to two-step GMM estimator. Alternatively, one-step estimator can be applied. One-step estimator takes into account the dynamic structure of the error term.

- The Arellano-Bond (AB) estimator is usually called difference GMM.
- The AB estimator deteriorates when:
  - y_{it} exhibits a substantial persistence, i.e., γ is close to unity.
  - the variance of unit-specific error component (σ_μ) increases relatively to the variance of the idiosyncratic error term (σ_ε).
- Note that for long panel (large T) the number of instruments increases dramatically, i.e., $r = T/(T-1)/2$.
- Consistency of the GMM estimator bases on the assumption that the transformed error term is not serially correlated, i.e., $\mathbb{E}\left(\Delta\varepsilon,_{i,t}, \Delta\varepsilon,_{i,t-2}\right) = 0$.
  - It's crucially to test whether the second-order autocorrelation is zero for all periods in the sample. Conventionally, test bases on residuals from the first difference equation.

A system GMM estimator

- Blundell and Bond (1998) propose to include additional moment restrictions.
  - These additional moment restrictions are imposed on the distribution of initial values, i.e., y_{i0}.
  - This set of restrictions is important when γ is close to unity and/or when $\sigma_\mu/\sigma_\varepsilon$ becomes large.
- Consider simply panel AR(1) without regressors. Then,

$$y_{i0} = \frac{\mu_i}{1-\gamma} + \varepsilon_{i0} \quad \text{for } i = 1, \ldots, N. \tag{43}$$

under the following assumption:

$$\mathbb{E}\left(\Delta y_{i1}\mu_i\right) = 0 \tag{44}$$

- It can be show that if the above condition is satisfied then the following $T-1$ moment conditions can be used:

$$\mathbb{E}\left[\left(y_{it} - \gamma y_{it-1}\right)\Delta y_{it-1}\right] = 0. \tag{45}$$

- Note that the system estimator combines the standard AB estimator and equation for levels (with the corresponding $T-1$ moment conditions).

- The instrument matrix:

$$Z = \begin{bmatrix} Z^{AB} & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Delta y_{iT-1} \end{bmatrix} \tag{46}$$

where Z^{AB} is the instrument matrix from the Arellano-Bond estimator.