Dynamic Linear Panel Data Models

230347 Advanced Microeconometrics
Tilburg University

Christoph Walsh

Introduction

- In this lecture we will allow y_{it} to also be a function of y_{it-1} in the linear model.
- We will show why we can no longer estimate the model with fixed effects (unless $T \to \infty$).
- We will discuss how to estimate dynamic linear models in different ways:
  - Anderson-Hsiao First Difference IV
  - Arellano-Bond Difference GMM
  - Blundell-Bond System GMM

Model

- Throughout this section we will be interested in estimating $(\rho, \boldsymbol{\beta})$ from:

$$y_{it} = \rho y_{it-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$$

  where

$$\mathbb{E}\left[\alpha_i\right] = \mathbb{E}\left[\varepsilon_{it}\right] = \mathbb{E}\left[\alpha_i \varepsilon_{it}\right] = 0$$

- Here y_{it-1} is predetermined: it is independent of the current disturbance ε_{it} but is influenced by ε_{it-1}.
- Samples will have "large N and small T".
- For demonstration purposes, we will often drop the covariates $\boldsymbol{x}_{it}$ to simplify notation.

Nickell (1981) Bias

- Consider the following model without covariates:

$$y_{it} = \rho y_{it-1} + \alpha_i + \varepsilon_{it}, \qquad \text{where } \varepsilon_{it} \text{ is iid over } t \text{ and } |\rho| < 1$$

- If we have $T \geq 3$ time periods, we can apply the within transformation to remove the α_i:

$$y_{it} - \bar{y}_i = \rho \left(y_{it-1} - \bar{y}_{i-1} \right) + \varepsilon_{it} - \bar{\varepsilon}_i$$

  where $\bar{y}_i = \frac{1}{T-1} \sum_{t=2}^{T} y_{it}$ and $\bar{y}_{i-1} = \frac{1}{T-1} \sum_{t=1}^{T-1} y_{it}$

- The regressor $y_{it-1} - \bar{y}_{i-1}$ is correlated with the error $\varepsilon_{it} - \bar{\varepsilon}_i$ as:
  - $\bar{\varepsilon}_i$ contains ε_{it-1} which is correlated with y_{it-1}.
  - $\bar{y}_{i-1}$ contains y_{it} which is correlated with ε_{it}.

Nickell (1981) Bias

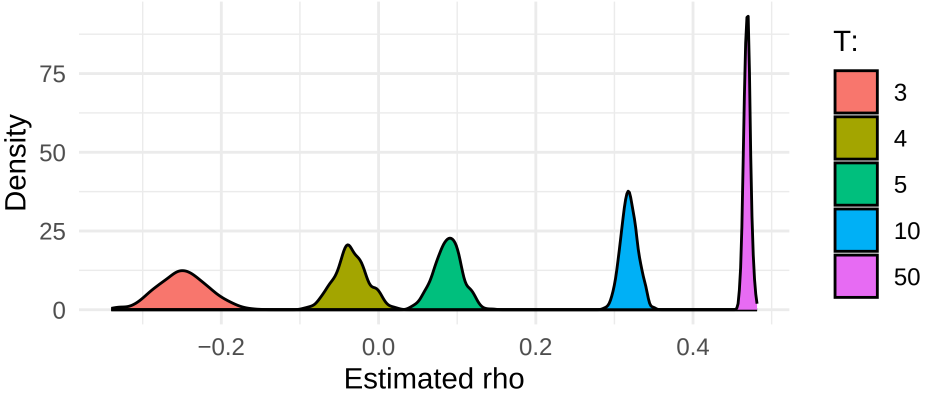
- This correlation creates a bias that does not vanish as $N \to \infty$ if T is fixed, so the FE estimator is inconsistent.
- This bias as $N \to \infty$ is approximately $-(1+\rho)/(T-2)$ for reasonably large values of T.
  - If $T \to \infty$, the bias goes to zero.
  - For $T = 3$, the bias is exactly $-\frac{1}{2}(1+\rho)$.
- If we can calculate the bias analytically, why not just correct the estimate ex-post?
  - Kiviet (1995) shows how to do this, but it does not work for unbalanced panels, nor for the possibility of other endogenous regressors.
  - We will study different methods that do not require a correction.

Demonstrating Nickell Bias with Simulations

If we generate 500 datasets with sample size $N = 1000$ and various T according to the model:

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

with $\varepsilon_{it} \sim \mathcal{N}(0, 1)$, $\alpha_i \sim \mathcal{N}(0, 1)$ and $\rho = 0.5$, and estimate ρ via fixed effects we get the following densities for the estimated ρs:



First Differencing: Anderson and Hsiao (1982)

- If we take first differences of the model $y_{it} = \rho y_{it-1} + \alpha_i + \varepsilon_{it}$, we get:

$$y_{it} - y_{it-1} = \rho \left(y_{it-1} - y_{it-2} \right) + \varepsilon_{it} - \varepsilon_{it-1} \qquad \text{for } t = 3, \ldots, T$$

- $y_{it-1} - y_{it-2}$ will still be correlated with $\varepsilon_{it} - \varepsilon_{it-1}$, and OLS estimates of ρ will be biased.
- y_{it-2}, however, is correlated with $y_{it-1} - y_{it-2}$, but not correlated with $\varepsilon_{it} - \varepsilon_{it-1}$.
  - Therefore, according to the model, y_{it-2} is a valid instrument for Δy_{it-1}.
  - $y_{it-2} - y_{it-3}$ would also be a valid instrument, but you would lose a time period for every individual.

First Differencing: Anderson and Hsiao (1982)

- The $(T-2) \times 1$ instrument matrix would then be:

$$\boldsymbol{Z}_i = (y_{i1}, y_{i2}, \ldots, y_{iT-2})'$$

- For example, for $T = 4$, for each i, the moments are:

$$\mathbb{E}\left[\boldsymbol{Z}_i' \Delta \boldsymbol{\varepsilon}_i\right] = \mathbb{E}\left[\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix}' \begin{pmatrix} \varepsilon_{i3} - \varepsilon_{i2} \\ \varepsilon_{i4} - \varepsilon_{i3} \end{pmatrix}\right] = 0$$

- If we were concerned that ε_{it} was serially correlated, you could use further lags instead as instruments. However, this would result in more dropped time periods.

Arellano and Bond (1991)

- While the Andserson-Hsiao estimator is consistent, it is not efficient, as it does not take into account of all available moment conditions.
- For the same reason that y_{it-2} is a valid instrument for Δy_{it-1}, y_{it-3} is also a valid instrument.
- We can continue adding instruments this way, so $y_{i1}, y_{i2}, \ldots, y_{it-2}$ are all valid instruments for Δy_{it-1}
- But, adding further lags in the Anderson-Hsiao approach would result in more dropped time periods.
- For example, if $T = 4$ and we use the 2nd and 3rd lag as instruments, we only have a complete set of instruments for $t = 4$:

$$\boldsymbol{Z}_i = \begin{pmatrix} y_{i1} & \cdot \\ y_{i2} & y_{i1} \end{pmatrix}$$

Holtz-Eakin et al. (1988)

- Holtz-Eakin et al. (1988) replaced the "dots" with zeros in the instrument matrix as each column would still be orthogonal to the first-differenced errors, assuming $\mathbb{E}\left[y_{it-2}\Delta\varepsilon_{it}\right] = 0$.
- Constructing the instrument matrix this way in the $T = 4$ case means we can use the $t = 3$ observations as well:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{Z}_i'\Delta\boldsymbol{\varepsilon}_i\right] &= \mathbb{E}\left[\begin{pmatrix} y_{i1} & y_{i2} \\ 0 & y_{i1} \end{pmatrix}\begin{pmatrix} \Delta\varepsilon_{i3} \\ \Delta\varepsilon_{i4} \end{pmatrix}\right] \\
&= \mathbb{E}\left[\begin{pmatrix} y_{i1}\Delta\varepsilon_{i3} + y_{i2}\Delta\varepsilon_{i4} \\ 0 + y_{i1}\Delta\varepsilon_{i4} \end{pmatrix}\right] \\
&= \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}$$

Arellano-Bond Instrument Matrix

- Arellano-Bond construct a slightly different instrument matrix which adds additional moments:

$$\boldsymbol{Z}_i'\Delta\boldsymbol{\varepsilon}_i = \begin{pmatrix} y_{i1} & 0 \\ 0 & y_{i1} \\ 0 & y_{i2} \end{pmatrix} \begin{pmatrix} \Delta\varepsilon_{i3} \\ \Delta\varepsilon_{i4} \end{pmatrix} = \begin{pmatrix} y_{i1}\Delta\varepsilon_{i3} \\ y_{i1}\Delta\varepsilon_{i4} \\ y_{i2}\Delta\varepsilon_{i4} \end{pmatrix}$$

- For $T = 6$, the instrument matrix would be:

$$\boldsymbol{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & y_{i4} \end{pmatrix}$$

GMM Estimation With One-Step Weight Matrix

- ρ is estimated by minimizing the GMM objective using an initial weight matrix $\boldsymbol{W}_1$:

$$\widehat{\rho}_1 = \arg\min_{\rho} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{m}_i(\rho) \right)' \boldsymbol{W}_1^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{m}_i(\rho) \right)$$

where:

$$\boldsymbol{m}_i(\rho) = \boldsymbol{Z}_i' \Delta \boldsymbol{\varepsilon}_i(\rho) = \boldsymbol{Z}_i' \left(\Delta \boldsymbol{y}_i - \rho \Delta \boldsymbol{y}_{i-1} \right) = \boldsymbol{Z}_i' \begin{pmatrix} \Delta y_{i3} - \rho \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} - \rho \Delta y_{iT-1} \end{pmatrix}$$

One-Step Weight Matrix

Software packages often use the following one-step weight matrix:

$$\boldsymbol{W}_1 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{H} \boldsymbol{Z}_i$$

where $\boldsymbol{H}$ is $\boldsymbol{D}'\boldsymbol{D}$, and where $\boldsymbol{D}$ is the $(T-2) \times (T-1)$ first difference operator:

$$\boldsymbol{D} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \quad \boldsymbol{H} = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ 0 & 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{pmatrix}$$

GMM Estimation With Two-Step Weight Matrix

- The second-step weight matrix is formed using the residuals from the first step estimate $\widehat{\rho}_1$:

$$\widehat{\boldsymbol{W}}_2 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i' \left[\Delta\boldsymbol{\varepsilon}_i\left(\widehat{\rho}_1\right)\right]\left[\Delta\boldsymbol{\varepsilon}_i\left(\widehat{\rho}_1\right)\right]' \boldsymbol{Z}_i$$

- Using this weight matrix in the same objective gives the two-step estimate $\widehat{\rho}_2$:

$$\widehat{\rho}_2 = \arg\min_{\rho} \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{m}_i\left(\rho\right)\right)' \widehat{\boldsymbol{W}}_2^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{m}_i\left(\rho\right)\right)$$

Closed-Form Solution

- We can actually solve for $\widehat{\rho}_k$, $k = 1, 2$ by taking first-order conditions of the GMM objective function with respect to ρ.
- The objective function is:

$$Q_k(\rho) = \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i' \left(\Delta \boldsymbol{y}_i - \rho \Delta \boldsymbol{y}_{i,-1}\right)\right]' \boldsymbol{W}_k^{-1} \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i' \left(\Delta \boldsymbol{y}_i - \rho \Delta \boldsymbol{y}_{i,-1}\right)\right]$$

- The first-order condition is:

$$\frac{\partial Q(\rho)}{\partial \rho} = -2\left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_{i,-1}\right]' \boldsymbol{W}_k^{-1} \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_i' \left(\Delta \boldsymbol{y}_i - \rho \Delta \boldsymbol{y}_{i,-1}\right)\right] = 0$$

- Solving for ρ yields the k-th step estimator for ρ:

$$\widehat{\rho}_k = \frac{\left(\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_{i,-1}\right)' \boldsymbol{W}_k^{-1} \left(\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_i\right)}{\left(\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_{i,-1}\right)' \boldsymbol{W}_k^{-1} \left(\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_{i,-1}\right)}$$

Sargan/Hansen Test for Over-Identification Restrictions

- In the model $\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \varepsilon_{it}$, with $T = 4$ we have 3 moments to identify one parameter.
- In general, there will be more moments than unknown parameters.
- When the model is overidentified, $\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \widehat{\boldsymbol{\varepsilon}}_i \neq \mathbf{0}$.
- The Sargan Test Statistic is:

$$J = \left[\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \widehat{\boldsymbol{\varepsilon}}_i \right]' \boldsymbol{W}_2^{-1} \left[\sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \widehat{\boldsymbol{\varepsilon}}_i \right] \sim \chi^2_{p-K-1}$$

  where p is the number of instruments and K is the number of variables in $\boldsymbol{x}_{it}$ (zero in this example).
- A low p-value indicates that the instruments may not be valid.

Test for Autocorrelation

- Arellano and Bond (1991) also propose a test for second-order serial correlation for the disturbances in the first-differenced equation.
- This is important, because consistency of the GMM estimator relies on $\mathbb{E}\left[\Delta\varepsilon_{it}\Delta\varepsilon_{it-2}\right] = 0$.
- A rejection of the test indicates that there may be serial correlation.
- You should check that the p-value for 2nd-order serial correlation is large.

System GMM: Blundell and Bond (1998)

- Consider the model:
$$y_{it} = \rho y_{it-1} + \alpha_i + \varepsilon_{it}$$
with $\rho \in (0,1)$, $\mathbb{E}[\alpha_i] = \mathbb{E}[\varepsilon_{it}] = \mathbb{E}[\alpha_i \varepsilon_{it}] = 0$ and $T = 3$.
- There is only one orthogonality condition, $\mathbb{E}[y_{i1}\Delta\varepsilon_{i3}] = 0$, so ρ is just identified.
- Subtracting y_{i1} from both sides of the model at $t = 2$ gives the first stage of this IV regression:
$$\Delta y_{i2} = (\rho - 1)\, y_{i1} + \alpha_i + \varepsilon_{it}$$
- Since we expect $\mathbb{E}[y_{i1}\alpha_i] > 0$, the coefficient $(\rho - 1)$ will be biased upwards towards zero.
- In general, the lagged values of y_{it} may be weak instruments for Δy_{it} if ρ is close to 1.

System GMM: Blundell and Bond (1998)

- With ρ close to 1, lagged *changes* may be more predictive of current *levels* than past levels on current changes.
- In the difference GMM approach, we use lagged levels of y_{it} as instruments for equations in differences.
- The system GMM approach uses lagged differences of y_{it} as instruments for equations in levels, *in addition to* lagged levels of y_{it} as instruments for equations in differences.
- Doing this assumes a stationarity restriction on the initial conditions.

System GMM: Blundell and Bond (1998)

- The additional moment condition is $\mathbb{E}\left[\Delta y_{it-1}\left(\alpha_i + \varepsilon_{it}\right)\right] = 0$.
- If $T = 3$, the 2 moment conditions are:

$$\mathbb{E}\left[\left(y_{i2} - y_{i1}\right)\left(\alpha_i + \varepsilon_{i3}\right)\right] = 0 \qquad \mathbb{E}\left[\left(\varepsilon_{i3} - \varepsilon_{i2}\right) y_{i1}\right] = 0$$

- If $T = 4$, the 6 moment conditions are:

$$\begin{aligned}
\mathbb{E}\left[\left(y_{i2} - y_{i1}\right)\left(\alpha_i + \varepsilon_{i3}\right)\right] &= 0 & \qquad \mathbb{E}\left[\left(\varepsilon_{i3} - \varepsilon_{i2}\right) y_{i1}\right] &= 0 \\
\mathbb{E}\left[\left(y_{i2} - y_{i1}\right)\left(\alpha_i + \varepsilon_{i4}\right)\right] &= 0 & \qquad \mathbb{E}\left[\left(\varepsilon_{i4} - \varepsilon_{i3}\right) y_{i1}\right] &= 0 \\
\mathbb{E}\left[\left(y_{i3} - y_{i2}\right)\left(\alpha_i + \varepsilon_{i4}\right)\right] &= 0 & \qquad \mathbb{E}\left[\left(\varepsilon_{i4} - \varepsilon_{i3}\right) y_{i2}\right] &= 0
\end{aligned}$$

System GMM: Blundell and Bond (1998)

- Taking a closer look at $\mathbb{E}\left[\Delta y_{it-1}\left(\alpha_i + \varepsilon_{it}\right)\right] = 0$.
- Using $\Delta y_{it-1} = (\rho - 1)\, y_{it-2} + \alpha_i + \varepsilon_{it-1}$:

$$\mathbb{E}\left[\left((\rho - 1)\, y_{it-2} + \alpha_i + \varepsilon_{it-1}\right)\left(\alpha_i + \varepsilon_{it}\right)\right] = 0$$

- Since the ε_{it} are assumed not to be serially correlated, and $\mathbb{E}\left[\alpha_i \varepsilon_{it}\right] = 0$, this simplifies to:

$$\mathbb{E}\left[\left((\rho - 1)\, y_{it-2} + \alpha_i\right)\alpha_i\right] = 0$$

- Assuming $|\rho| < 1$, rewriting this:

$$\mathbb{E}\left[\left(y_{it-2} - \frac{\alpha_i}{1 - \rho}\right)\alpha_i\right] = 0$$

- In the pure autoregressive case, $\alpha_i / (1 - \rho)$ is the steady-state value of y_{it}.
    - The moment condition is that deviations from the steady-state must be uncorrelated with the level of the steady state $\alpha_i / (1 - \rho)$.

Suggested Reading

- Baltagi, chapter 8
- Croissant and Millo, chapter 7
- Roodman (2009) "How to do `xtabond2`: An Introduction to difference and system GMM in Stata"

References

Anderson, T. W. and C. Hsiao (1982): "Formulation and estimation of dynamic models using panel data," *Journal of econometrics*, 18, 47–82.

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, 58, 277–297.

Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 87, 115–143.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988): "Estimating vector autoregressions with panel data," *Econometrica*, 1371–1395.

Kiviet, J. F. (1995): "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models," *Journal of econometrics*, 68, 53–78.

Nickell, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica*, 1417–1426.

Roodman, D. (2009): "How to do xtabond2: An introduction to difference and system GMM in Stata," *The stata journal*, 9, 86–136.