Advanced Dynamic Panel Data Methods

Methods to Analyze a Simple Dynamic Panel Data Model

Sebastian Kripfganz

University of Exeter Business School, Department of Economics, Exeter, UK
www.kripfganz.de @Kripfganz

Universidad de Salamanca

July 17–19, 2023

University of Exeter Business School

Universidad de Salamanca

Panel AR(1) Model

- Let us initially consider a simple autoregressive panel model of order 1 – panel AR(1) – with error components structure:

$$y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$$

  with homoskedastic and serially uncorrelated ε_{it}.
- Assuming that the process is dynamically stable – i.e., $|\lambda| < 1$ – and started in the infinite past, we can iterate the process backwards to obtain

$$y_{it} = \frac{\alpha_i}{1-\lambda} + \sum_{s=0}^{\infty} \lambda^s \varepsilon_{i,t-s}$$

  - The properties of the process are the same if it started at time period $t = 1$ with a draw from the stationary distribution:

$$y_{i1} = \frac{\alpha_i}{1-\lambda} + \frac{1}{\sqrt{1-\lambda^2}}\varepsilon_{i1}$$

Estimation Methods

- There are three leading approaches to obtain consistent estimates of λ, given that $y_{i,t-1}$ is predetermined with respect to ε_{it} and

$$Cov(y_{i,t-1}, \alpha_i + \varepsilon_{it}) = Cov(y_{i,t-1}, \alpha_i) = \frac{\sigma_\alpha^2}{(1-\lambda)^2} \neq 0$$

  - Bias-corrected (BC) estimation, utilizing the known bias expression for the FE estimator.
  - (Quasi-)maximum likelihood (QML) estimation, making further assumptions about the initial observations.
  - Instrumental-variables (IV) and generalized method of moments (GMM) estimation, exploiting knowledge about the model dynamics to construct valid instruments.

Least-Squares Estimation

- Many estimators can be regarded as special cases of general estimation methods.
- The FE estimator is the pooled OLS estimator for the transformed model

$$\bar{\Delta} y_{it} = \lambda \bar{\Delta} y_{i,t-1} + \bar{\Delta} \varepsilon_{it}$$

where

$$\bar{\Delta} y_{it} = y_{it} - \bar{y}_i, \quad \bar{y}_i = \frac{1}{T-1} \sum_{t=2}^{T} y_{it}$$

$$\bar{\Delta} y_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}, \quad \bar{y}_{i,-1} = \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1}$$

$$\bar{\Delta} \varepsilon_{it} = \varepsilon_{it} - \bar{\varepsilon}_i, \quad \bar{\varepsilon}_i = \frac{1}{T-1} \sum_{t=2}^{T} \varepsilon_{it}$$

Least-Squares Estimation

- The FE estimator minimizes the sum of squared (transformed) residuals:

$$\hat{\lambda}_{FE} = \arg\min_{\hat{\lambda}} \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} (\bar{\Delta} y_{it} - \hat{\lambda} \bar{\Delta} y_{i,t-1})^2$$

- The first-order condition can be written as

$$\frac{1}{N} \sum_{i=1}^{N} m_{FE,i}(\hat{\lambda}_{FE}) = 0$$

where

$$m_{FE,i}(\lambda) = \frac{1}{T-1} \sum_{t=2}^{T} \bar{\Delta} y_{i,t-1} \underbrace{(\bar{\Delta} y_{it} - \lambda \bar{\Delta} y_{i,t-1})}_{\bar{\Delta} \varepsilon_{it}}$$

Least-Squares Estimation

- It is often convenient to use more compact vector notation. By stacking all time-series observations above each other, the untransformed regression model for unit i can be written as

$$\underbrace{\mathbf{y}_i}_{\begin{pmatrix} y_{i2} \\ y_{i3} \\ \vdots \\ y_{iT} \end{pmatrix}} = \lambda \underbrace{\mathbf{y}_{i,-1}}_{\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,T-1} \end{pmatrix}} + \alpha_i \underbrace{\boldsymbol{\iota}_{T-1}}_{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}} + \underbrace{\boldsymbol{\varepsilon}_i}_{\begin{pmatrix} \varepsilon_{i2} \\ \varepsilon_{i3} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}}$$

- Similarly, the transformed model becomes

$$\bar{\Delta}\mathbf{y}_i = \lambda \bar{\Delta}\mathbf{y}_{i,-1} + \bar{\Delta}\boldsymbol{\varepsilon}_i$$

Least-Squares Estimation

- Even more compactly, we can stack all observations above each other:

$$\underbrace{\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}}_{\mathbf{y}} = \lambda \underbrace{\begin{pmatrix} \mathbf{y}_{1,-1} \\ \mathbf{y}_{2,-1} \\ \vdots \\ \mathbf{y}_{N,-1} \end{pmatrix}}_{\mathbf{y}_{-1}} + \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix}}_{\boldsymbol{\alpha}} \otimes \boldsymbol{\iota}_{T-1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

where $\otimes$ denotes the Kronecker product, such that

$$\boldsymbol{\alpha} \otimes \boldsymbol{\iota}_{T-1} = \begin{pmatrix} \alpha_1 \boldsymbol{\iota}_{T-1} \\ \alpha_2 \boldsymbol{\iota}_{T-1} \\ \vdots \\ \alpha_N \boldsymbol{\iota}_{T-1} \end{pmatrix}$$

Least-Squares Estimation

- The FE estimator as the solution to the first-order condition $m_{FE}(\hat{\lambda}_{FE}) = \frac{1}{N}\sum_{i=1}^{N} m_{FE,i}(\hat{\lambda}_{FE}) = 0$ can then be written as

$$\hat{\lambda}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}(\bar{\Delta} y_{i,t-1})^2\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\bar{\Delta} y_{i,t-1}\bar{\Delta} y_{it}$$
$$= \left(\sum_{i=1}^{N}\bar{\Delta}\mathbf{y}'_{i,-1}\bar{\Delta}\mathbf{y}_{i,-1}\right)^{-1}\sum_{i=1}^{N}\bar{\Delta}\mathbf{y}'_{i,-1}\bar{\Delta}\mathbf{y}_{i}$$
$$= (\bar{\Delta}\mathbf{y}'_{-1}\bar{\Delta}\mathbf{y}_{-1})^{-1}\bar{\Delta}\mathbf{y}'_{-1}\bar{\Delta}\mathbf{y}$$

- The estimator is only defined for $T \geq 3$:
  - For each unit, one effective observation is lost due to the lagged dependent variable and another degree of freedom is absorbed by removing the unit-specific mean.

Least-Squares Estimation

- Let the residual maker matrix that partials out the unit-specific intercepts (in line with the Frisch-Waugh-Lovell theorem) be

$$\mathbf{M}_\iota = \underbrace{\mathbf{I}_{T-1}}_{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}} - \underbrace{\boldsymbol{\iota}_{T-1}(\boldsymbol{\iota}'_{T-1}\boldsymbol{\iota}_{T-1})^{-1}\boldsymbol{\iota}'_{T-1}}_{\frac{1}{T-1}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & \vdots \\ \vdots & & \ddots & 1 \\ 1 & \cdots & 1 & 1 \end{pmatrix}}$$

  such that $(\boldsymbol{\iota}'_{T-1}\boldsymbol{\iota}_{T-1})^{-1}\boldsymbol{\iota}'_{T-1}\mathbf{y}_i = \frac{1}{T-1}\sum_{t=2}^{T} y_{it}$ and thus $\mathbf{M}_\iota \mathbf{y}_i = \mathbf{y}_i - \bar{y}_i \boldsymbol{\iota}_{T-1} = \bar{\Delta}\mathbf{y}_i$.
  - $\mathbf{M}_\iota$ is idempotent – i.e., $\mathbf{M}_\iota\mathbf{M}_\iota = \mathbf{M}_\iota$ – and symmetric – i.e, $\mathbf{M}'_\iota = \mathbf{M}_\iota$

Least-Squares Estimation

- Because $\mathbf{M}_\iota \boldsymbol{\iota}_{T-1} = \mathbf{0}$, pre-multiplying the regression model with matrix $\mathbf{M}_\iota$ eliminates the unit-specific intercept α_i. The transformed regression model can thus be written as

$$\mathbf{M}_\iota \mathbf{y}_i = \lambda \mathbf{M}_\iota \mathbf{y}_{i,-1} + \mathbf{M}_\iota \boldsymbol{\varepsilon}_i$$
$$(\mathbf{I}_N \otimes \mathbf{M}_\iota)\mathbf{y} = \lambda (\mathbf{I}_N \otimes \mathbf{M}_\iota)\mathbf{y}_{-1} + (\mathbf{I}_N \otimes \mathbf{M}_\iota)\boldsymbol{\varepsilon}$$

and the FE estimator as

$$\hat{\lambda}_{FE} = \left(\sum_{i=1}^{N} \mathbf{y}'_{i,-1} \mathbf{M}_\iota \mathbf{y}_{i,-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{y}'_{i,-1} \mathbf{M}_\iota \mathbf{y}_i$$
$$= (\mathbf{y}'_{-1}(\mathbf{I}_N \otimes \mathbf{M}_\iota)\mathbf{y}_{-1})^{-1} \mathbf{y}'_{-1}(\mathbf{I}_N \otimes \mathbf{M}_\iota)\mathbf{y}$$

Least-Squares Estimation

- Perhaps surprisingly, the FE estimator can also be obtained from the first-differenced model:

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta \varepsilon_{it}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$, and $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$.

- In compact notation, the first-differenced model is

$$\underbrace{\begin{pmatrix} \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}}_{\Delta \mathbf{y}_i} = \lambda \underbrace{\begin{pmatrix} \Delta y_{i2} \\ \Delta y_{i3} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix}}_{\Delta \mathbf{y}_{i,-1}} + \underbrace{\begin{pmatrix} \Delta \varepsilon_{i3} \\ \Delta \varepsilon_{i4} \\ \vdots \\ \Delta \varepsilon_{iT} \end{pmatrix}}_{\Delta \boldsymbol{\varepsilon}_i}$$

Least-Squares Estimation

- It turns out that

$$\mathbf{M}_{\iota} = \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}$$

for the first-difference transformation matrix

$$\mathbf{D} = (\mathbf{0}, \mathbf{I}_{T-2}) - (\mathbf{I}_{T-2}, \mathbf{0}) = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

such that, $\mathbf{D}\mathbf{y}_i = \Delta\mathbf{y}_i$.

- As highlighted by Bun and Kiviet (2006), it holds more generally that

$$\mathbf{M}_{\iota} = \mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}\mathbf{K}$$

for any $(T-2) \times (T-1)$ transformation matrix $\mathbf{K}$ with $\text{rk}(\mathbf{K}) = T-2$ and $\mathbf{K}\boldsymbol{\iota}_{T-1} = \mathbf{0}$.

Least-Squares Estimation

- Consequently, the FE estimator can be written as

$$\hat{\lambda}_{FE} = \left(\sum_{i=1}^{N} \mathbf{y}'_{i,-1} \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1} \mathbf{D}\mathbf{y}_{i,-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{y}'_{i,-1} \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1} \mathbf{D}\mathbf{y}_{i}$$

  - When ε_{it} is homoskedastic and serially uncorrelated, this is a generalized least squares (GLS) estimator because $Var(\Delta\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2 \mathbf{D}\mathbf{D}'$, which reflects the fact that $\Delta\varepsilon_{it}$ has first-order serial correlation:

$$Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{is}) = \begin{cases} 2\sigma_\varepsilon^2 & , s = t \\ -\sigma_\epsilon^2 & , |t-s| = 1 \\ 0 & , |t-s| \geq 2 \end{cases}$$

Maximum Likelihood Estimation

- Minimizing the sum of squared residuals is equivalent to maximizing the log-likelihood function:

$$\hat{\lambda}_{FE} = \arg\max_{\hat{\lambda}} \ln \mathcal{L}(\hat{\lambda}, \hat{\sigma}_{\varepsilon}^2(\hat{\lambda}))$$

where

$$\ln \mathcal{L}(\hat{\lambda}, \hat{\sigma}_{\varepsilon}^2(\hat{\lambda})) = -\frac{N(T-1)}{2} \ln(2\pi\hat{\sigma}_{\varepsilon}^2(\hat{\lambda})) - \frac{1}{2\hat{\sigma}_{\varepsilon}^2(\hat{\lambda})} \sum_{i=1}^{N} \sum_{t=2}^{T} (\bar{\Delta} y_{it} - \hat{\lambda}\bar{\Delta} y_{i,t-1})^2$$

after concentrating out the variance parameter by using its closed-form solution:

$$\hat{\sigma}_{\varepsilon}^2(\hat{\lambda}) = \frac{1}{N(T-2)} \sum_{i=1}^{N} \sum_{t=2}^{T} (\bar{\Delta} y_{it} - \hat{\lambda}\bar{\Delta} y_{i,t-1})^2$$

Maximum Likelihood Estimation

- The first-order condition is the same as before:

$$\frac{1}{N}\sum_{i=1}^{N} m_{FE,i}(\hat{\lambda}_{FE}) = 0$$

- The likelihood function is derived from the joint multivariate normal distribution of $\bar{\Delta}\varepsilon_{i2}, \bar{\Delta}\varepsilon_{i3}, \ldots, \bar{\Delta}\varepsilon_{iT}$, conditional on the transformed initial observations $\bar{\Delta}y_{i1}$.
  - In general, numerical methods are used to maximize the log-likelihood function – e.g., the Newton-Raphson algorithm.
  - If the errors are not normally distributed, the same likelihood function can be used but the estimator is called a quasi-ML estimator.
  - The inconsistency of the FE estimator results from the fact that $\bar{\Delta}y_{i1}$ is not exogenous with respect to the joint distribution of the idiosyncratic errors.

Method of Moments Estimation

- Using the first-order condition as the starting point, the FE estimator can equivalently be characterized as a method of moments (MM) estimator, solving

$$\hat{\lambda}_{FE} = \arg\min_{\hat{\lambda}} \left(\frac{1}{N} \sum_{i=1}^{N} m_{FE,i}(\hat{\lambda}) \right)^2$$

  - This quadratic objective function attains its minimum again at $\frac{1}{N}\sum_{i=1}^{N} m_{FE,i}(\hat{\lambda}_{FE}) = 0$.
  - The FE estimator is biased/inconsistent because $E[m_{FE,i}(\lambda)] \neq 0$, due to

$$E[\bar{\Delta} y_{i,t-1} \bar{\Delta} \varepsilon_{it}] = E\left[\left(y_{i,t-1} - \frac{1}{T-1} \sum_{s=2}^{T} y_{i,s-1} \right) \left(\varepsilon_{it} - \frac{1}{T-1} \sum_{s=2}^{T} \varepsilon_{is} \right) \right] \neq 0$$

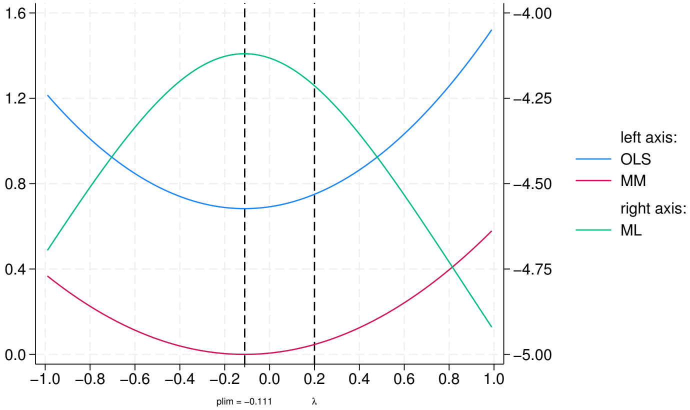
Instrumental-Variables Estimation

- A special case of an MM estimator is the simple IV estimator with instrument z_{it}.
  - A valid instrument must satisfy the moment condition $E\left[\frac{1}{T-1}\sum_{t=2}^{T} z_{it}\bar{\Delta}\varepsilon_{it}\right] = 0$ (or a similar moment condition for a different transformation or no transformation of the error term $\alpha_i + \varepsilon_{it}$).
  - The FE estimator can be characterized as an IV estimator with invalid instrument $z_{it} = \bar{\Delta}y_{i,t-1}$ – i.e., the transformed lagged dependent variable is instrumented by itself – such that

$$m_{FE,i}(\lambda) = \frac{1}{T-1}\sum_{t=2}^{T} z_{it}\bar{\Delta}\varepsilon_{it}$$

Equivalence of Estimation Methods

- FE objective functions for the pooled OLS, MM, and ML methods:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.2$



Method of Moments Estimation

- More generally, a consistent MM estimator solves

$$\hat{\boldsymbol{\theta}}_{MM} = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}(\hat{\boldsymbol{\theta}})' \, \mathbf{m}(\hat{\boldsymbol{\theta}})$$

given a $K \times 1$ vector of parameters $\boldsymbol{\theta}$ and an $L \times 1$ vector of moment functions

$$\mathbf{m}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\boldsymbol{\theta}})$$

satisfying the moment conditions $E[\mathbf{m}_i(\boldsymbol{\theta})] = \mathbf{0}$.

  - If the moment functions $\mathbf{m}(\hat{\boldsymbol{\theta}})$ are nonlinear in the parameters, a closed-form solution does not exist, and numerical optimization methods are needed – e.g., the Gauss-Newton algorithm.

Generalized Method of Moments Estimation

- If there are more (nonredundant) moment equations than parameters – i.e., $L > K$ – the estimator is overidentified.
- A GMM estimator introduces a weighting matrix $\mathbf{W}(\hat{\boldsymbol{\theta}}_1)$, possibly depending on a preliminary estimator $\hat{\boldsymbol{\theta}}_1$:

$$\hat{\boldsymbol{\theta}}_{GMM} = \hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}(\hat{\boldsymbol{\theta}}_1)) = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}(\hat{\boldsymbol{\theta}})' \, \mathbf{W}(\hat{\boldsymbol{\theta}}_1) \, \mathbf{m}(\hat{\boldsymbol{\theta}})$$

  - When $L > K$, some or all of the moment functions evaluated at the minimizer $\hat{\boldsymbol{\theta}}_{GMM}$ will be nonzero. The Hansen (1982) test of the overidentifying restrictions can be used as a model misspecification test, assuming that $\hat{\boldsymbol{\theta}}_{GMM}$ is an efficient estimator:

$$J = N \, \mathbf{m}(\hat{\boldsymbol{\theta}}_{GMM})' \, \mathbf{W}(\hat{\boldsymbol{\theta}}_1) \, \mathbf{m}(\hat{\boldsymbol{\theta}}_{GMM}) \xrightarrow{d} \chi^2(L-K)$$

- In the just-identified case – i.e., $L = K$ – the weighting matrix is irrelevant, $\hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}(\hat{\boldsymbol{\theta}}_1)) = \hat{\boldsymbol{\theta}}_{MM}$, and $J = 0$.

Generalized Method of Moments Estimation

- While the GMM estimator is consistent for any weighting matrix if the moment conditions $E[\mathbf{m}_i(\boldsymbol{\theta})] = \mathbf{0}$ are satisfied, a two-step procedure is needed for efficiency.
  1. Estimate $\boldsymbol{\theta}$ with any consistent estimator $\hat{\boldsymbol{\theta}}_1$, which is typically a one-step GMM estimator $\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W})$ with fixed initial weighting matrix $\mathbf{W}$, which might depend on the data but not on unknown parameters.
  2. The efficient two-step estimator $\hat{\beta}_{GMM}(\mathbf{W}(\hat{\boldsymbol{\theta}}_1))$ uses an estimate of the moment functions' inverse variance-covariance matrix from the one-step estimator:

$$\mathbf{W}(\hat{\boldsymbol{\theta}}_1) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\boldsymbol{\theta}}_1)\, \mathbf{m}_i(\hat{\boldsymbol{\theta}}_1)' \right)^{-1}$$

  This is the optimal weighting matrix in the sense that it minimizes the asymptotic variance of the estimator.

Generalized Method of Moments Estimation

- While the choice of the initial estimator (or initial weighting matrix) is irrelevant asymptotically, it affects the finite-sample results. To remove the dependence on this initial choice, one can repeatedly update the weighting matrix and estimator:

$$\hat{\boldsymbol{\theta}}_j = \hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}(\hat{\boldsymbol{\theta}}_{j-1})) = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}(\hat{\boldsymbol{\theta}})' \, \mathbf{W}(\hat{\boldsymbol{\theta}}_{j-1}) \, \mathbf{m}(\hat{\boldsymbol{\theta}})$$

- The iterated GMM estimator (Hansen, Heaton, and Yaron, 1996) is obtained upon convergence as the fixed point of this updating rule:

$$\hat{\boldsymbol{\theta}}_{iGMM} = \hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}(\hat{\boldsymbol{\theta}}_{iGMM})) = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}(\hat{\boldsymbol{\theta}})' \, \mathbf{W}(\hat{\boldsymbol{\theta}}_{iGMM}) \, \mathbf{m}(\hat{\boldsymbol{\theta}})$$

Generalized Method of Moments Estimation

- Another alternative is the continuously-updating GMM estimator (Hansen, Heaton, and Yaron, 1996), which evaluates the optimal weighting matrix jointly with the moment functions:

$$\hat{\boldsymbol{\theta}}_{cuGMM} = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}(\hat{\boldsymbol{\theta}})' \; \mathbf{W}(\hat{\boldsymbol{\theta}}) \; \mathbf{m}(\hat{\boldsymbol{\theta}})$$

  where

$$\mathbf{W}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\boldsymbol{\theta}}) \; \mathbf{m}_i(\hat{\boldsymbol{\theta}})' \right)^{-1}$$

  - This estimator is generally computationally more demanding because its first-order condition is nonlinear in the parameters $\hat{\boldsymbol{\theta}}$ even if the moment functions $\mathbf{m}(\hat{\boldsymbol{\theta}})$ are linear in the parameters. Consequently, there is no closed-form solution and numerical methods need to be employed.

Dynamic Panel Bias

- Assuming $E[y_{i1}\varepsilon_{it}] = 0$ for all $t \geq 2$, Nickell (1981) obtained an expression for the downward bias – more precisely, the inconsistency – of the FE estimator:

$$\operatorname*{plim}_{N\to\infty} (\hat{\lambda}_{FE} - \lambda) = -\frac{\frac{1+\lambda}{T-2}\left(1 - \frac{1-\lambda^{T-1}}{(1-\lambda)(T-1)}\right)}{1 - \frac{2\lambda}{(1-\lambda)(T-2)}\left(1 - \frac{1-\lambda^{T-1}}{(1-\lambda)(T-1)}\right)}$$

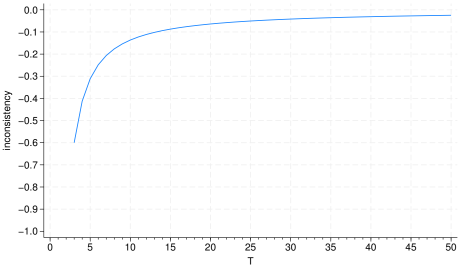
  - A simple approximation is

$$\operatorname*{plim}_{N\to\infty} (\hat{\lambda}_{FE} - \lambda) \approx -\frac{1+\lambda}{T-1}$$

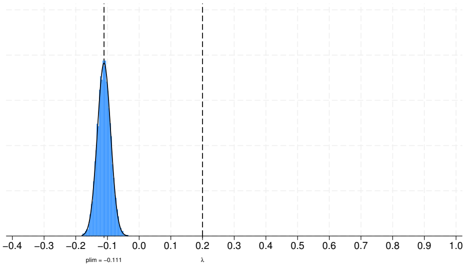
  - Note that $\operatorname{plim}_{N\to\infty}(\hat{\lambda}_{FE} - \lambda) \to 0$ as $T \to \infty$ (or as $\lambda \to -1$).

Dynamic Panel Bias

- Inconsistency/bias of the dynamic FE estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.2$ (low persistence)



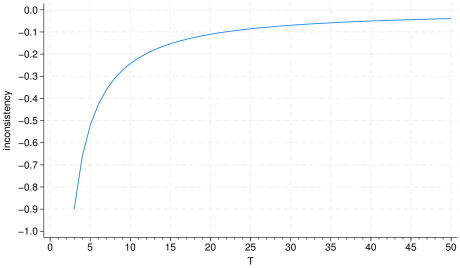
(a) inconsistency for varying T



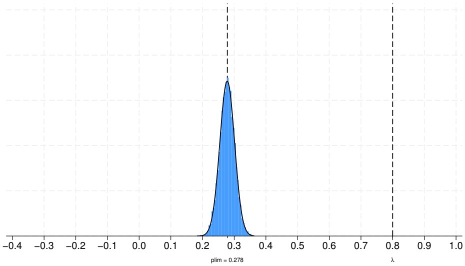
(b) bias for $N = 600$, $T = 5$; $\varepsilon_{it} \sim \mathcal{N}(0,1)$; 10,000 replications

Dynamic Panel Bias

- Inconsistency/bias of the dynamic FE estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.8$ (high persistence)



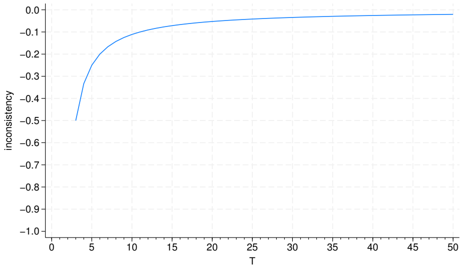
(a) inconsistency for varying T



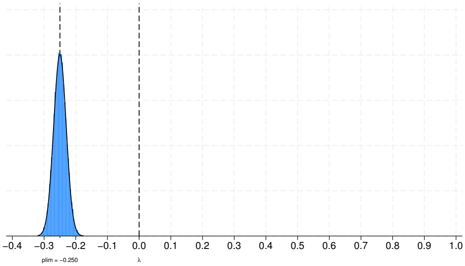
(b) bias for $N = 600$, $T = 5$; $\varepsilon_{it} \sim \mathcal{N}(0, 1)$; 10,000 replications

Dynamic Panel Bias

- Inconsistency/bias of the dynamic FE estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0$ (no persistence)



(a) inconsistency for varying T



(b) bias for $N = 600$, $T = 5$; $\varepsilon_{it} \sim \mathcal{N}(0, 1)$; 10,000 replications

Additive Bias Correction

- Based on the closed-form expression

$$b_T(\lambda) = \operatorname*{plim}_{N\to\infty} (\hat{\lambda}_{FE} - \lambda)$$

(or another suitable bias approximation), an additive bias correction can be applied to the FE estimator:

$$\hat{\lambda}_{aBCFE} = \hat{\lambda}_{FE} - b_T(\hat{\lambda}_0)$$

with a bias estimate obtained from an initial consistent estimator $\hat{\lambda}_0$.

  - Kiviet (1995) operationalized this idea based on a refined bias approximation, also allowing for additional strictly exogenous regressors $\mathbf{x}_{it}$.
  - A downside of this approach is its reliance on an initial consistent estimator, and the necessity to obtain standard errors with bootstrap methods.

Iterative Bias Correction

- To avoid reliance on an initial consistent estimator, we can consider the iterations

$$\hat{\lambda}_j = \hat{\lambda}_{FE} - b_T(\hat{\lambda}_{j-1})$$

  starting with an initial guess $\hat{\lambda}_0 \in (-1, 1)$, which could be $\hat{\lambda}_{FE}$ or any other reasonable value.

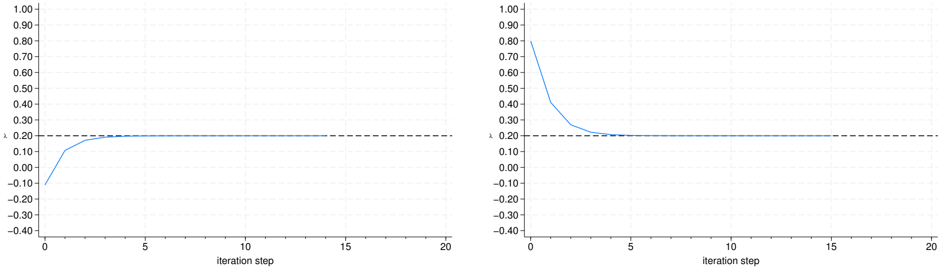
- The iterative bias-corrected estimator is the fixed point of this updating rule obtained upon convergence:

$$\hat{\lambda}_{iBCFE} = \hat{\lambda}_{FE} - b_T(\hat{\lambda}_{iBCFE})$$

  - The estimator proposed by Bun and Carree (2005) is based on this idea, again extended to accommodate strictly exogenous regressors $\mathbf{x}_{it}$.

Iterative Bias Correction

- Numerical illustration (proof of concept) for the iterative bias correction, where $\hat{\lambda}_{FE}$ is replaced by its probability limit instead of using actual data.
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\lambda = 0.2$; $T = 5$
  - $\hat{\lambda}_j = (\text{plim}_{N\to\infty} \hat{\lambda}_{FE}) - b_T(\hat{\lambda}_{j-1})$
  - Convergence criterion: $|\hat{\lambda}_j - \hat{\lambda}_{j-1}| < 10^{-7}$



(a) $\hat{\lambda}_0 = \text{plim}_{N\to\infty} \hat{\lambda}_{FE}$ (b) $\hat{\lambda}_0 = 0.8$

Bias-Corrected Method of Moments Estimation

- Recall that the bias of the FE estimator arises because $E[m_{FE,i}(\lambda)] = d_T(\lambda, \sigma_\varepsilon^2) \neq 0$, where

$$d_T(\lambda, \sigma_\varepsilon^2) = -\frac{\sigma_\varepsilon^2}{(1-\lambda)(T-1)} \left(1 - \frac{1-\lambda^{T-1}}{(1-\lambda)(T-1)}\right)$$

- An alternative approach is to directly correct the moment functions:

$$m_{BCFE,i}(\lambda, \sigma_\varepsilon^2) = m_{FE,i}(\lambda) - d_T(\lambda, \sigma_\varepsilon^2)$$

such that $E[m_{BCFE,i}(\lambda)] = 0$

Bias-Corrected Method of Moments Estimation

- The bias-corrected MM estimator of Breitung, Kripfganz, and Hayakawa (2022) then solves

$$\hat{\lambda}_{BCFE} = \arg\min_{\hat{\lambda}} \left(\frac{1}{N} \sum_{i=1}^{N} m_{BCFE,i}(\hat{\lambda}, \hat{\sigma}_{\varepsilon}^2(\hat{\lambda})) \right)^2$$

  - The estimator is equivalent to the iterative bias-corrected estimator of Bun and Carree (2005) and the adjusted profile likelihood estimator of Dhaene and Jochmans (2016).
  - The bias formula can be adjusted for higher-order autoregressive models:

$$y_{it} = \sum_{j=1}^{p} \lambda_j y_{i,t-j} + \alpha_i + \varepsilon_{it}$$

Bias-Corrected Method of Moments Estimation

- Extending the bias-corrected MM approach to models with strictly exogenous regressors $\mathbf{x}_{it}$ is straightforward by simply adding the relevant moment functions:

$$\mathbf{m}_{BCFE,i}(\boldsymbol{\theta}, \sigma_\varepsilon^2) = \mathbf{m}_{FE,i}(\boldsymbol{\theta}) - \begin{pmatrix} d_T(\lambda, \sigma_\varepsilon^2) \\ \mathbf{0} \end{pmatrix}$$

where $\boldsymbol{\theta} = (\lambda, \beta')'$ and

$$\mathbf{m}_{FE,i}(\boldsymbol{\theta}) = \frac{1}{T-1} \sum_{t=2}^{T} \begin{pmatrix} \bar{\Delta} y_{i,t-1} \\ \bar{\Delta} \mathbf{x}_{it} \end{pmatrix} \underbrace{(\bar{\Delta} y_{it} - \lambda \bar{\Delta} y_{i,t-1} - \bar{\Delta} \mathbf{x}'_{it} \beta)}_{\bar{\Delta} \varepsilon_{it}}$$

  - Standard errors are easy to compute using well established asymptotic results for method of moments estimators.

Bias-Corrected Method of Moments Estimation

- Moreover, a RE version is readily constructed assuming that the regressors $\mathbf{x}_{it}$ are uncorrelated with the unit-specific error component α_i:

$$\mathbf{m}_{BCRE,i}(\boldsymbol{\theta}, \sigma_\varepsilon^2) = \mathbf{m}_{RE,i}(\boldsymbol{\theta}) - \begin{pmatrix} d_T(\lambda, \sigma_\varepsilon^2) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

where

$$\mathbf{m}_{RE,i}(\boldsymbol{\theta}) = \frac{1}{T-1} \sum_{t=2}^{T} \begin{pmatrix} \bar{\Delta} y_{i,t-1} \\ \bar{\Delta} \mathbf{x}_{it} \\ \mathbf{x}_{it} \end{pmatrix} \underbrace{(y_{it} - \lambda y_{i,t-1} - \mathbf{x}_{it}' \beta)}_{\alpha_i + \varepsilon_{it}}$$

Bias-Corrected Method of Moments Estimation

- The bias-corrected RE estimator is overidentified. There are $1 + 2K_x$ moment equations for $1 + K_x$ parameters. The estimator solves

$$\hat{\boldsymbol{\theta}}_{BCRE}(\mathbf{W}(\hat{\boldsymbol{\theta}}_1)) = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbf{m}_{BCRE}(\hat{\boldsymbol{\theta}})' \ \mathbf{W}(\hat{\boldsymbol{\theta}}_1) \ \mathbf{m}_{BCRE}(\hat{\boldsymbol{\theta}})$$

  - Two-step, iterated, or continuously-updating GMM can be used for efficient estimation.
  - The initial consistent (but inefficient) estimator $\hat{\boldsymbol{\theta}}_1$ is usually the one-step GMM estimator with initial weighting matrix $\mathbf{W} = \left(\frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \mathbf{z}_{it}\mathbf{z}_{it}' \right)^{-1}$, where

$$\mathbf{z}_{it} = \begin{pmatrix} \bar{\Delta} y_{i,t-1} \\ \bar{\Delta}\mathbf{x}_{it} \\ \mathbf{x}_{it} \end{pmatrix}$$

Bias-Corrected Method of Moments Estimation

- The K_x overidentifying restrictions imposed by the additional RE moment conditions $E[\mathbf{x}_{it}(\alpha_i + \varepsilon_{it})] = 0$ can be evaluated with the Hansen (1982) test:

$$J = N\ \mathbf{m}_{BCRE}(\hat{\boldsymbol{\theta}}_{BCRE})'\ \mathbf{W}(\hat{\boldsymbol{\theta}}_1)\ \mathbf{m}_{BCRE}(\hat{\boldsymbol{\theta}}_{BCRE}) \xrightarrow{d} \chi^2(K_x)$$

- $\hat{\boldsymbol{\theta}}_{BCRE}$ is efficient if the RE assumption holds, but otherwise inconsistent. $\hat{\boldsymbol{\theta}}_{BCFE}$ is consistent in both cases. This motivates a test for a statistically significant difference between both estimators following the Hausman (1978) principle:

$$H = (\hat{\boldsymbol{\theta}}_{BCFE} - \hat{\boldsymbol{\theta}}_{BCRE})'(Var(\hat{\boldsymbol{\theta}}_{BCFE} - \hat{\boldsymbol{\theta}}_{BCRE}))^{-1}(\hat{\boldsymbol{\theta}}_{BCFE} - \hat{\boldsymbol{\theta}}_{BCRE}) \xrightarrow{d} \chi^2(K_x)$$

- Both tests are asymptotically equivalent.

Quasi-Maximum Likelihood Estimation

- Recall that the FE estimator could be equivalently obtained from the model in deviations from within-group means or the model in first differences.
- Based on the latter, to account for the correlation of the initial observations Δy_{i2} with $\Delta \varepsilon_{i3}$, an auxiliary equation can be added to the model:

$$\Delta y_{i2} = \psi + \nu_{i2}$$

with $Var(\nu_{i2}) = \omega \sigma_\varepsilon^2$ and

$$Cov(\nu_{i2}, \Delta \varepsilon_{it}) = \begin{cases} -\sigma_\varepsilon^2 & , t = 3 \\ 0 & , t > 3 \end{cases}$$

Quasi-Maximum Likelihood Estimation

- An unconditional log-likelihood function $\mathcal{L}(\hat{\lambda}, \hat{\sigma}_\varepsilon^2, \hat{\psi}, \hat{\omega})$ can be constructed based on the joint multivariate (normal) distribution of $\nu_{i2}, \Delta\varepsilon_{i3}, \ldots, \Delta\varepsilon_{iT}$, where ψ and ω are auxiliary parameters to be estimated.
  - This results in the unconditional FE-QML estimator suggested by Hsiao, Pesaran, and Tahmiscioglu (2002), who also extend it to the model with additional strictly exogenous regressors $\mathbf{x}_{it}$.
  - The restriction $\hat{\psi} = 0$ could be employed when it is assumed that the initial observations are draws from the stationary distribution.
  - Following a similar idea, an unconditional RE-QML estimator was proposed by Bhargava and Sargan (1983).

Quasi-Maximum Likelihood Estimation

- Breitung, Kripfganz, and Hayakawa (2022) demonstrate that the unconditional log-likelihood function can be nicely written in terms of deviations from within-group means:

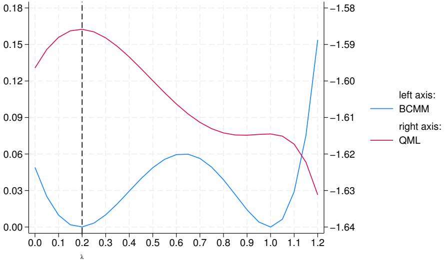
$$\ln \mathcal{L}(\hat{\lambda}, \sigma_\varepsilon^2, \hat{\psi}, \hat{\omega})) = -\frac{N(T-1)}{2} \ln(2\pi\hat{\sigma}_\varepsilon^2) - \frac{1}{2\hat{\sigma}_\varepsilon^2} \sum_{i=1}^{N} \sum_{t=2}^{T} (\bar{\Delta} y_{it} - \hat{\lambda} \bar{\Delta} y_{i,t-1})^2$$

$$-\frac{N}{2} \ln(1+(T-1)(\hat{\omega}-1)) - \frac{T-1}{2\hat{\sigma}_\varepsilon^2 (1+(T-1)(\hat{\omega}-1))} \sum_{i=1}^{N} ((1-\hat{\lambda}) y_{i1} + \hat{\psi} - (\bar{y}_i - \hat{\lambda} \bar{y}_{i,-1}))^2$$

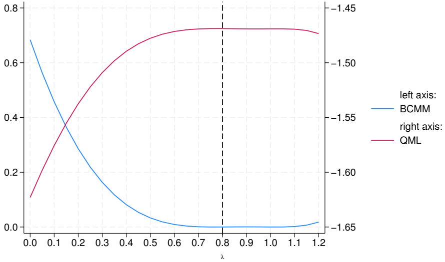
  - The first two terms equal the conditional log-likelihood function. Adding the last two terms effectively serves as a bias correction.

Bias-Corrected Estimation Methods

- Objective functions of the bias-corrected MM estimator and the QML estimator:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$; $T = 5$



(a) $\lambda = 0.2$ (low persistence)



(b) $\lambda = 0.8$ (high persistence)

Instrumental-Variables Estimation

- The FD estimator is biased/inconsistent because $\Delta y_{i,t-1}$ is correlated with $\Delta\varepsilon_{it}$ in the first-differenced regression model

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta\varepsilon_{it}$$

- If the (untransformed) idiosyncratic error term ε_{it} is serially uncorrelated, then $Cov(\Delta y_{i,t-s}, \Delta\varepsilon_{it}) = Cov(y_{i,t-s}, \Delta\varepsilon_{it}) = 0$ for $s \geq 2$.
  - To be precise, $Cov(\Delta y_{i,t-s}, \Delta\varepsilon_{it}) = 0$ relies on the initial-observations assumption

$$E\left[\left(y_{i1} - \frac{\alpha}{1-\lambda}\right)\Delta\varepsilon_{it}\right] = 0$$

for $t \geq 3$, while $Cov(y_{i,t-s}, \Delta\varepsilon_{it}) = 0$ requires the joint assumption that $E[y_{i1}\varepsilon_{it}] = 0$ and $E[\alpha_i\varepsilon_{it}] = 0$ for $t \geq 2$. The latter implies the former but not vice versa.

Instrumental-Variables Estimation

- Anderson and Hsiao (1981) thus suggest to construct a simple (just-identified) IV estimator for the first-differenced model with either instrument $z_{it} = \Delta y_{i,t-2}$ or $z_{it} = y_{i,t-2}$.
  - Using $z_{it} = \Delta y_{i,t-2}$ loses a further observation per unit and requires $T \geq 4$, while $z_{it} = y_{i,t-2}$ still only requires $T \geq 3$.
- The strength of the instruments varies with the persistence of the process:

$$Cov(z_{it}, \Delta y_{i,t-1}) = \begin{cases} -\frac{1-\lambda}{1+\lambda}\sigma_\varepsilon^2 & , z_{it} = \Delta y_{i,t-2} \\ -\frac{1}{1+\lambda}\sigma_\varepsilon^2 & , z_{it} = y_{i,t-2} \end{cases}$$

  assuming that the process is dynamically stable and reached stationarity.
  - The strength of either instrument diminishes as $\lambda \to 1$, but more severely for $z_{it} = \Delta y_{i,t-2}$.
  - The observation that $Cov(y_{i,t-2}, \Delta y_{i,t-1}) \to -\frac{1}{2}\sigma_\epsilon^2$ as $\lambda \to 1$ suggests that $z_{it} = y_{i,t-2}$ retains its strength. However, this is misleading.

Instrumental-Variables Estimation

- With

$$Var(z_{it}) = \begin{cases} \frac{2}{1+\lambda}\sigma_\varepsilon^2 & , z_{it} = \Delta y_{i,t-2} \\ \frac{1}{(1-\lambda)^2}\sigma_\alpha^2 + \frac{1}{1-\lambda^2}\sigma_\varepsilon^2 & , z_{it} = y_{i,t-2} \end{cases}$$

we find for the first-stage regression

$$\Delta y_{i,t-1} = \pi z_{it} + \nu_{it}$$

that

$$\pi = \frac{Cov(z_{it}, \Delta y_{i,t-1})}{Var(z_{it})} = \begin{cases} -\frac{1-\lambda}{2} & , z_{it} = \Delta y_{i,t-2} \\ -(1-\lambda)\frac{\frac{1-\lambda}{1+\lambda}\sigma_\varepsilon^2}{\sigma_\alpha^2 + \frac{1-\lambda}{1+\lambda}\sigma_\varepsilon^2} & , z_{it} = y_{i,t-2} \end{cases}$$

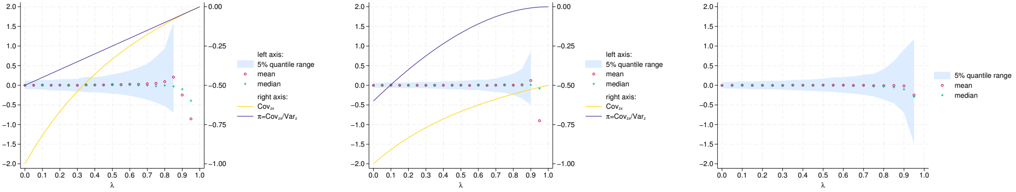
such that $\pi \to 0$ as $\lambda \to 1$ for either instrument.

Instrumental-Variables Estimation

- Instead of choosing between the two instruments, they can also be combined with a two-stage least squares (2SLS) estimator – i.e., $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$
  - Because $\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$, an equivalent set of instruments is $\mathbf{z}_{it} = (y_{i,t-2}, y_{i,t-3})$
  - However, this does not solve the problem of instrument weakness when $\lambda \to 1$.

Instrumental-Variables Estimation

- Instrument strength and distribution of $\hat{\lambda} - \lambda$ for the simple IV and the 2SLS estimator based on 1,001 replications:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \in \{-1, 0, 1\}$, and $\lambda \in \{0, 0.05, 0.1, \ldots, 0.9, 0.95\}$; $N = 600$, $T = 5$
  - Stationary initial observations: $y_{i1} = \frac{\alpha_i}{1-\lambda} + \nu_{i1}$, where $\nu_{i1} \sim \mathcal{N}\left(0, \frac{1}{1-\lambda^2}\right)$



(a) $z_{it} = \Delta y_{i,t-2}$ (b) $z_{it} = y_{i,t-2}$ (c) $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$

Instrumental-Variables Estimation

- In the previous data-generating process (DGP), $E[y_{it}|\alpha_i] \rightarrow \infty$ and $Var(y_{it}) \rightarrow \infty$ as $\lambda \rightarrow 1$. A potential remedy for this dependence on λ is to assume a restricted DGP:

$$y_{it} = \lambda y_{i,t-1} + \underbrace{(1-\lambda)\tilde{\alpha}_i}_{\alpha_i} + \underbrace{\sqrt{1-\lambda^2}\tilde{\varepsilon}_{it}}_{\varepsilon_{it}}$$

  such that $E[y_{it}|\alpha_i] = \tilde{\alpha}_i$ and $Var(y_{it}) = \sigma^2_{\tilde{\alpha}} + \sigma^2_{\tilde{\varepsilon}}$

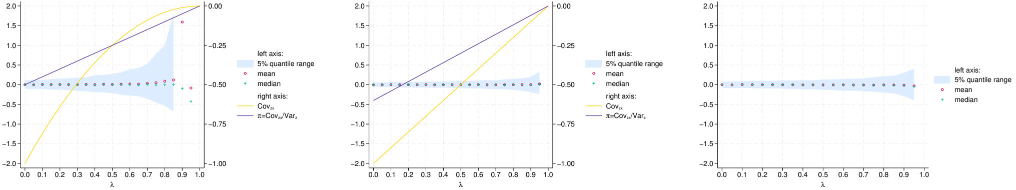
  - Other properties of the model still depend on λ, such as $Var(\Delta y_{it}) = 2(1-\lambda)\sigma^2_{\tilde{\varepsilon}}$ and

$$Cov(z_{it}, \Delta y_{i,t-1}) = \begin{cases} -(1-\lambda)^2\sigma^2_{\tilde{\varepsilon}} & , z_{it} = \Delta y_{i,t-2} \\ -(1-\lambda)\sigma^2_{\tilde{\varepsilon}} & , z_{it} = y_{i,t-2} \end{cases}$$

$$\pi = \frac{Cov(z_{it}, \Delta y_{i,t-1})}{Var(z_{it})} = \begin{cases} -\frac{1-\lambda}{2} & , z_{it} = \Delta y_{i,t-2} \\ -(1-\lambda)\frac{\sigma^2_{\tilde{\varepsilon}}}{\sigma^2_{\tilde{\alpha}}+\sigma^2_{\tilde{\varepsilon}}} & , z_{it} = y_{i,t-2} \end{cases}$$

Instrumental-Variables Estimation

- Instrument strength and distribution of $\hat{\lambda} - \lambda$ for the simple IV and the 2SLS estimator based on 1,001 replications:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0, 1-\lambda^2)$, $\alpha_i \in \{-(1-\lambda), 0, 1-\lambda\}$, and $\lambda \in \{0, 0.05, 0.1, \ldots, 0.9, 0.95\}$; $N = 600$, $T = 5$
  - Stationary initial observations: $y_{i1} = \frac{\alpha_i}{1-\lambda} + \nu_{i1}$, where $\nu_{i1} \sim \mathcal{N}(0, 1)$



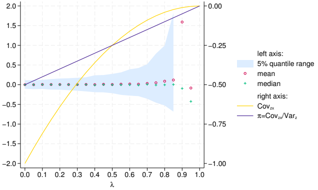
(a) $z_{it} = \Delta y_{i,t-2}$ (b) $z_{it} = y_{i,t-2}$ (c) $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$

Instrumental-Variables Estimation

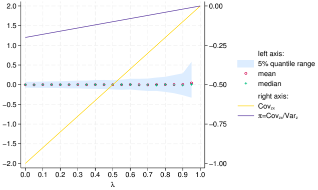
- Other properties of the DGP affect the performance of the estimators as well:
  - As highlighted by Blundell and Bond (1998), a higher variance ratio $\frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$ further reduces the strength of the instrument $z_{it} = y_{i,t-2}$.
  - Nonstationary initial observations might have non-trivial effects on the estimators.

Instrumental-Variables Estimation

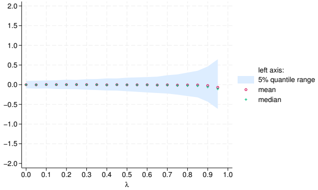
- Instrument strength and distribution of $\hat{\lambda} - \lambda$ for the simple IV and the 2SLS estimator based on 1,001 replications:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0, 1 - \lambda^2)$, $\alpha_i \in \{-2(1-\lambda), 0, 2(1-\lambda)\}$, and $\lambda \in \{0, 0.05, 0.1, \ldots, 0.9, 0.95\}$; $N = 600$, $T = 5$
  - Stationary initial observations: $y_{i1} = \frac{\alpha_i}{1-\lambda} + \nu_{i1}$, where $\nu_{i1} \sim \mathcal{N}(0,1)$



(a) $z_{it} = \Delta y_{i,t-2}$



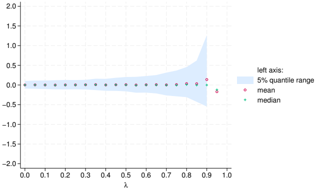
(b) $z_{it} = y_{i,t-2}$



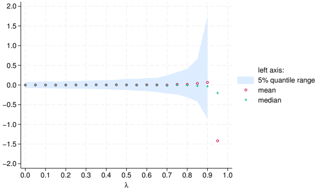
(c) $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$

Instrumental-Variables Estimation

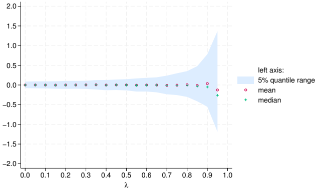
- Instrument strength and distribution of $\hat{\lambda} - \lambda$ for the simple IV and the 2SLS estimator based on 1,001 replications:
  - $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathcal{N}(0, 1-\lambda^2)$, $\alpha_i \in \{-(1-\lambda), 0, 1-\lambda\}$, and $\lambda \in \{0, 0.05, 0.1, \ldots, 0.9, 0.95\}$; $N = 600$, $T = 5$
  - Initial observations: $y_{i1} = E[y_{it}|\alpha_i] = \frac{\alpha_i}{1-\lambda}$ (mean stationary but not covariance stationary)



(a) $z_{it} = \Delta y_{i,t-2}$



(b) $z_{it} = y_{i,t-2}$



(c) $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$

Serial-Correlation Tests

- A crucial assumption for all estimators considered so far is that the idiosyncratic error component ε_{it} is serially uncorrelated. Testing the validity of this assumption is essential.
- Arellano and Bond (1991) suggest a test of the null hypothesis $H_0 : Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{i,t-s}) = 0$ for some $s \geq 1$.
  - If ε_{it} is indeed serially uncorrelated, we expect to reject the null hypothesis for order $s = 1$ but not to reject it for any higher order $s \geq 2$.
  - In the empirical practice, the test for no serial correlation of the first-differenced errors at order $s = 2$ is routinely applied.
- As an extension, Yamagata (2008) considers the null hypothesis $H_0 : Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{i,t-s}) = 0$ for all $2 \leq s \leq q$. This is a joint test of no serial correlation from order 2 up to some order q.

Serial-Correlation Tests

- Both tests are special cases of the portmanteau test proposed by Jochmans (2020), which allows for nonstationary alternatives, where $Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{i,t-s})$ might be a function of t.
  - The portmanteau test can be more powerful if T is (very) small, but quickly loses power when T becomes moderately large.
- All of these tests are applicable after BC-MM, QML, and IV estimation.
- Evidence of serial correlation in ε_{it} might be an indication of omitted variables. Estimating a higher-order autoregressive model could be a remedy.

Serial-Correlation Tests

- If the model is overidentified and the overidentifying restriction is implied by the absence of serial correlation, a conventional overidentification test can be used to assess this assumption.
  - For the IV estimator, both $z_{it} = \Delta y_{i,t-2}$ and $z_{it} = y_{i,t-2}$ become invalid instruments when there is second-order serial correlation in $\Delta\varepsilon_{it}$ (which would be implied by first-order serial correlation in ε_{it}).
  - However, the linear combination $y_{i,t-2} - \Delta y_{i,t-2} = y_{i,t-3}$ remains a valid instrument (as long as there is no higher-order serial correlation). Since the 2SLS estimator with both instruments $\mathbf{z}_{it} = (\Delta y_{i,t-2}, y_{i,t-2})$ is equivalent to the 2SLS estimator with instruments $\mathbf{z}_{it} = (y_{i,t-2}, y_{i,t-3})$, testing the resulting overidentifying restriction is informative about the absence of serial correlation.
  - Effectively, such a test contrasts the overidentified 2SLS estimator with a just-identified estimator using only the instrument $z_{it} = y_{i,t-3}$, which is valid both under the null and the alternative hypothesis.

Interim Conclusion

- In the simple panel AR(1) model (or its extension with strictly exogenous regressors), the bias/inconsistency of the FE estimator can be addressed effectively by bias correcting the estimator, or by adjusting the objective function of an MM or QML estimator.
- A simple IV or 2SLS estimator is less effective and might suffer from weak instruments, especially when the process is very persistent. However, this approach might be beneficial when other predetermined or endogenous regressors are added to the model.
- All estimators seen so far crucially depend on the assumption of serially uncorrelated idiosyncratic errors. The BC and QML estimators also typically require (time-series) homoskedasticity.
  - A more flexible approach is often desirable.