TA Session 7: Dynamic Linear Panel Data

Shuowen Chen[1]

EC708: PhD Econometrics I (Spring 2020)

[1] Parts of the materials are borrowed from Blundell and Bond (1998), teaching and presentation slides by Manuel Arellano, Victor Chernozhukov and Iván Fernández-Val. I thank Iván Fernández-Val for helpful suggestions.

Outline

- Lagged Dependent Variable Model
  - Bias of FE (Nickell's Bias), RE and FD
  - Incidental Parameter Problem
- GMM Estimation of Dynamic Panel
  - Anderson–Hsiao (1981)
  - Arellano–Bond (1991)
  - Blundell–Bond (1998)
- High–Dimensional GMM Bias
  - Fixed Effects: Many Nuisance Parameters
  - Arellano–Bond: Many Moments
  - Bias Correction Methods:
    - Panel Splitting (Jackknife)
    - Analytical Bias Correction (ABC) for FE

Dynamic Linear Panel $AR(1)$

$$y_{i,t} = \rho y_{i,t-1} + v_{i,t}, \quad i = 1, \ldots, N; t = 1, \ldots, T$$

- $y_{i,t-1}$: lagged dependent variable
- $v_{i,t} = \alpha_i + u_{i,t}$, $u_{i,t} \sim \mathcal{N}(0, \sigma_u^2)$
- ρ: autoregressive parameter ($|\rho| < 1$)

Recursive substitution leads to

$$y_{i,t} = \rho^t y_{i,0} + \frac{1-\rho^t}{1-\rho}\alpha_i + \sum_{j=0}^{t-1} \rho^j u_{i,t-j}$$

Remarks:

- $y_{i,0}$: initial values
- $|\rho| < 1$ means effect of idiosyncratic shock ($u_{i,t}$) dies out
- Can add exogenous covariates on the RHS, but by FWL theorem we can project them out
- Drop one period of data for regression due to $AR(1)$ set up

Fixed–Effects in Dynamic Linear Panel

- Within transformation:

$$y_{i,t} - \overline{y}_i = \rho(y_{i,t-1} - \overline{y}_{i-1}) + (u_{i,t} - \overline{u}_i),$$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$ and $\overline{y}_{i-1} = \frac{1}{T-1}\sum_{t=1}^{T-1} y_{i,t}$

- $\overline{y}_{i-1}$ is correlated with $\overline{u}_i$ because the latter contains $u_{i,t-1}$
- Why is this a problem? Inconsistency of $\widehat{\rho}_{FE}$, which is essentially OLS applied to the within transformed regression

$$\widehat{\rho}_{FE} = \rho + \frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i-1})(u_{i,t} - \overline{u}_i)}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i-1})^2}$$

- We consider short panel (small T) asymptotics ($N \to \infty$)
- For long panel asymptotics ($N \to \infty$, $T \to \infty$), refer to Fernández-Val and Weidner (2018)

FE in Dynamic Linear Panel: Nickell's Bias (1981)

- Fix T and set $N \to \infty$,

$$\widehat{\rho}_{FE} - \rho \to -\frac{1+\rho}{T}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)\left[1 - \frac{1}{T} - \frac{2\rho}{(1-\rho)T}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)\right]^{-1}$$

- When $T = 2$:

$$\widehat{\rho}_{FE} - \rho \approx -\frac{1+\rho}{2}$$

- When T is large (long panel):

$$\widehat{\rho}_{FE} - \rho \approx -\frac{1+\rho}{T-1}$$

- Within transformation creates a correlation between regressors and error
- In short panel ($N \to \infty$, T fixed) a consistency problem, in long panel ($N \to \infty$, $T \to \infty$) a bias problem

Random Effects in Dynamic Linear Panel

$$y_{i,t} = \rho y_{i,t-1} + \nu_{i,t}$$

- $\nu_{i,t} = \alpha_i + u_{i,t}$, $u_{i,t} \sim \mathcal{N}(0, \sigma_u^2)$,
- RE assumption: $\alpha_i | y_{i,1} \ldots, y_{i,T} \sim \mathcal{N}(0, \sigma_\alpha^2)$
- Recall quasi–demean in the linear panel TA session:

$$y_{i,t} - \lambda \overline{y}_i = \rho(y_{i,t-1} - \lambda \overline{y}_{i-1}) + (\nu_{i,t} - \lambda \overline{\nu}_i),$$

where $\lambda = 1 - \sqrt{1/\left[1 + T\left(\sigma_\alpha/\sigma_u\right)^2\right]}$
- $y_{i,t-1} - \lambda \overline{y}_{i-1}$ is correlated with $\nu_{i,t} - \lambda \overline{\nu}_i$
- Why? $y_{i,t-1}$ also depends on α_i, which means

$$\mathbb{E}(\alpha_i | y_{i,t-1}) \neq 0$$

Violates the RE assumption

First Difference in Dynamic Linear Panel

$$y_{i,t} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + (u_{i,t} - u_{i,t-1})$$

- $y_{i,t-1}$ depends on $u_{i,t-1}$
- Implies $(y_{i,t-1} - y_{i,t-2})$ is correlated with $(u_{i,t} - u_{i,t-1})$
- $\widehat{\rho}_{FD}$ is biased
- Side note: in linear panel, fixed-effect and first-difference estimators coincide when $T = 2$

Taking stock: When lagged dependent variable is included as a regressor, FE, RE and FD fail to account for the endogeneity it brings. Worsen if data is persistent. We will introduce other estimators.

Incidental Parameter Problem and Fixed Effects

- MLE isn't efficient all the time ▸ Superefficient Estimator
- MLE isn't consistent (Neyman and Scott, 1948) when model features incidental parameters: nuisance parameters whose dimension grows with the sample size ▸ Neyman–Scott Example
- Fixed effects is consistent in linear panel because within transformation differences out α_i; don't need to deal with α_i
- This technique doesn't work for nonlinear panel, so fixed effects is generally inconsistent in nonlinear panel as N grows and T fixed[2]
- Nonlinear panel uses MLE, α_i treated as nuisance parameters
- Small T leads to noisy $\widehat{\alpha}_i$, which contaminates estimators of interest: incidental parameters leads to inconsistency
- Becomes a bias problem if having large T, will explore this soon

▸ Semiparametric Nonlinear Panel ▸ Some Examples ▸ Nonlinear panel MLE

[2] Dynamic Linear Panel serves as one example. Nonlinear panel will be covered in more details in EC709/711.

Anderson–Hsiao: First–Differenced IV

- The problem of FE, RE and FD is due to endogeneity, so naturally use an IV method
- Consider the first-differenced regression:

$$\Delta y_{i,t} = \rho \Delta y_{i,t-1} + \Delta u_{i,t}$$

- FD is problematic because $\Delta y_{i,t-1}$ is correlated with $\Delta u_{i,t}$
- Remedy: use $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$
  1. Relevance: $\Delta y_{i,t-1} := y_{i,t-1} - y_{i,t-2}$
  2. Exogeneity: $\mathbb{E}(y_{i,t-2} \Delta u_{i,t}) = 0$
- Estimator is consistent but inefficient: doesn't exploit all the moment conditions (Motivates Arellano–Bond estimator)
- If $\Delta u_{i,t}$ is autocorrelated, inconsistent

Arellano–Bond

- Recall the first-differenced regression:

$$\Delta y_{i,t} = \rho \Delta y_{i,t-1} + \Delta u_{i,t}$$

- What are the valid instruments for each period?
  - $t = 1, 2$: no instruments[3]
  - $t = 3$: $\Delta y_{i,2} = y_{i,2} - y_{i,1}$; IV is $y_{i,1}$
  - $t = 4$: $\Delta y_{i,3} = y_{i,3} - y_{i,2}$: IVs are $y_{i,2}$ and $y_{i,1}$. Convince yourself by using the recursive substitution on $\Delta y_{i,3}$
  - $t = T$: $\Delta y_{i,T-1} = y_{i,T-1} - y_{i,T-2}$: IVs are $y_{i,T-2}, \ldots, y_{i,1}$
- There are in total $\frac{(T-1)(T-2)}{2}$ IVs and hence moment conditions

$$\mathbb{E}[(\Delta y_{i,t} - \rho \Delta y_{i,t-1}) y_{i,s}] = 0, \quad t = 3, \ldots, T; s = 1, \ldots, t-2$$

- Estimate by two–step GMM ▸ Details

[3]When $t = 1$, regression is infeasible as it has $y_{i,-1}$. This is relevant to the timing assumption in Blundell and Bond (1998), who include additional moments at $t = 2$ by imposing assumptions on the distribution of initial values.

Remarks on Arellano–Bond

- Also called first–differenced GMM estimator
- Poor performance if ρ close to unity (persistent dynamics) or relative variance of α_i increases (weak instruments). Blundell and Bond (1998) address this issue ▸ $T = 3$ Case
- For large T moments increase a lot, incurs high–dimensional bias (many instruments). More discussion later
- Consistency predicated on transformed error not serially correlated $\mathbb{E}(\Delta u_{i,t}\Delta u_{i,t-2}) = 0$. Can test this based on residuals of the first-differenced equation.

Blundell–Bond: Introduction

$$y_{i,t} = \rho y_{i,t-1} + \underbrace{\alpha_i + u_{i,t}}_{\nu_{i,t}} \quad i = 1, \ldots, N, t = 1, \ldots, T$$

- Blundell and Bond (1998) has $t = 2, \ldots, T$, I change notations and corresponding timings for consistency of the slides
- $\mathbb{E}(\alpha_i) = 0, \mathbb{E}(u_{i,t}) = 0, \mathbb{E}(\alpha_i u_{i,t}) = 0, \mathbb{E}(u_{i,t} u_{i,s}) = 0 \; \forall t \neq s$
- $\mathbb{E}(y_{i,1} u_{i,t}) = 0$: assumption on initial values $y_{i,0}$
- Lagged levels ($y_{i,s}$) can be weak instruments
- Arellano and Bover (1995) add moments that use lagged differences of $y_{i,t}$ as instruments

$$\mathbb{E}(\nu_{i,t} \Delta y_{i,t-1}) = 0, \quad t = 3, \ldots, T$$

- Blundell and Bond (1998): $\Delta y_{i,1}$ is observed, additional moment

$$\mathbb{E}(\nu_{i,2} \Delta y_{i,1}) = 0$$

Needs restrictions on the initial conditions that generates $y_{i,0}$

Blundell–Bond: Initial Conditions

- Specify $y_{i,0} = \frac{\alpha_i}{1-\rho} + \nu_{i,0}$
- Then $\mathbb{E}(\nu_{i,2}\Delta y_{i,1}) = 0$ is equivalent to

$$\mathbb{E}\big[(\alpha_i + u_{i,2})(u_{i,1} + (\rho - 1)\nu_{i,0})\big] = 0$$

- Necessary conditions: $\mathbb{E}(\nu_{i,0}\alpha_i) = \mathbb{E}(\nu_{i,0}u_{i,2}) = 0$, $i = 1, \ldots, N$

In sum, Blundell–Bond use the following moment conditions:

$$\mathbb{E}\big[(\Delta y_{i,t} - \rho\Delta y_{i,t-1})y_{i,s}\big] = 0, \quad t = 3, \ldots, T;\, s = 1, \ldots, t-2$$
$$\mathbb{E}(\nu_{i,t}\Delta y_{i,t-1}) = 0, \quad t = 2, \ldots, T$$

Remarks:

- Blundell and Bond propose a conditional GLS estimator that allow for more moment conditions to be included
- Assume homoskedasticity across both individual and time:

$$\mathbb{E}(\alpha_i)^2 = \sigma_\alpha^2, \quad \mathbb{E}(u_{i,t})^2 = \sigma_u^2, \quad i = 1, \ldots, N,\ t = 1, \ldots, T$$

▸ GMM Details

Causal Dynamic Linear Panel (Acemoglu et al., 2019)

$$Y_{i,t} = \alpha_i + b_t + D'_{i,t}\theta + W'_{i,t}\beta + \varepsilon_{i,t}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T$$

- $Y_{i,t}$: outcome (log GDP) for a country i in year t
- $D_{i,t}$: vector of treatments of causal interest
- $W_{i,t}$: controls including a constant and **lags of** $Y_{i,t}$
- α_i and b_t: unobserved unit and time effects
- $\varepsilon_{i,t}$: error term normalized to have zero mean for each unit that satisfies the **weak sequential exogeneity** condition

$$\varepsilon_{i,t} \perp \mathcal{I}_{i,t}, \quad \mathcal{I}_{it} := \left\{(D_{i,s}, W_{i,s}, b_s)_{s=1}^{t}, a_i\right\}$$

- $Z_i := \left\{(Y_{i,t}, D'_{i,t}, W'_{i,t})'\right\}_{t=1}^{T}$ i.i.d. across i

Fixed Effects Approach

- Treat unit and time effects as parameters to be estimated
- Applies OLS in the model:

$$Y_{i,t} = D_{i,t}'\theta + X_{i,t}'\gamma + \varepsilon_{i,t}, \quad X_{i,t} := (W_{i,t}', Q_i', Q_t')'$$

- Q_i: N-dimensional vector of unit fixed effects
- Q_t: T-dimensional vector of time fixed effects
- FE can be seen as an **exactly identified** GMM estimator with the score function

$$g(Z_i, \theta, \gamma) = \left\{(Y_{i,t} - D_{i,t}'\theta - X_{i,t}'\gamma)M_{i,t}\right\}_{t=1}^{T},$$

where $M_{i,t} := (D_{i,t}', X_{i,t}')'$ and $Z_i := \left\{(Y_{i,t}, D_{i,t}', W_{i,t}')'\right\}_{t=1}^{T}$

Problems with FE: High-Dimensionality

- Biased with order $N/NT = 1/T$
- Why? Estimate N nuisance parameters with NT observations
- Need $T \to \infty$ for $\widehat{\theta}$ to approach θ_0: drive bias of $\widehat{\theta}$ to zero under the weak exogeneity condition
- But this alone does not suffice for inference, since bias b is large compared to the stochastic error $O(1/\sqrt{NT})$ of $\widehat{\theta}$
- Problems with FE with small T are well documented in econometrics and machine learning (e.g., Neyman and Scott, 1948; Nickell, 1981; Hahn and Kuersteiner, 2011)
- Need **bias correction** for validity of inference

AB Approach: A Review

- Eliminate α_i by taking differences across time

$$\Delta Y_{i,t} = \Delta D'_{i,t}\theta + \Delta X'_{i,t}\gamma + \Delta\varepsilon_{i,t}, \quad X_{i,t} = (W'_{i,t}, Q'_t)'$$

- Moment conditions for the variables in differences

$$\Delta\varepsilon_{i,t} \perp M_{i,t}, \quad M_{i,t} = [(D'_{i,s}, W'_{i,s})_{s=1}^{t-1}, Q'_t], \qquad t = 2, \ldots, T$$

- Lead to overidentified GMM with score function

$$g(Z_i, \theta, \gamma) = \{(\Delta Y_{i,t} - \Delta D'_{i,t}\theta - \Delta X'_{i,t}\gamma)M_{i,t}\}_{t=2}^{T}$$

Problems with AB: High Dimensionality

- Consistent under short-panel asymptotics, but can be biased when T is large due to the many instrument problem[4]
- The number of instruments or moment conditions is

$$m = \dim(g(Z_i, \theta, \gamma)) = O(T^2)$$

- Bias of order $m/NT = O(T/N)$ and may not be small compared to the sampling error $O_p(1/\sqrt{NT})$ of the estimator
- Leads to invalid inference, need **bias correction**

[4] Recall the first TA session

High Dimensional Asymptotic Approximation

- Denote $n = NT$
- In the FE approach, the dimension of θ is low, but the dimension of nuisance parameters γ might be high. Think of this as:

$$p = \dim(\gamma) \to \infty, \quad n \to \infty, \quad \dim(\theta) = \text{const}.$$

- In the AB approach, the number of moment conditions

$$m = \dim(g(Z_i, \theta, \gamma)),$$

can be high, while the dimension of θ is low. Think of this as:

$$m \to \infty, \quad n \to \infty, \quad \dim(\theta) = \text{const}.$$

GMM in High Dimensions

- The approximate normality and consistency results of GMM estimator $\widehat{\theta}$ continue to hold[5]

$$\sqrt{n}(\widehat{\theta} - \theta) \sim \mathcal{N}(0, V_{11})$$

if p^2 and m^2 are small compared to n, formally **rate condition**

$$(p \vee m)^2/n := \max(p^2, m^2)/n \to 0 \text{ as } n \to \infty$$

Can interpret as the small bias condition

- This fails in FE

$$p^2 = O(N^2 + T^2) \text{ is not small compared to } n = NT$$

- In the AB approach, if T is large

$$m^2 = O(T^4) \text{ is not small compared to } n = NT$$

[5] V_{11} is the $d_\theta \times d_\theta$ upper–left block of GMM variance estimator matrix corresponding to $\widehat{\theta}$

GMM in High Dimensions

- To understand the rate condition, focus on the exactly identified case where $p = m$
- An asymptotic second order expansion of $\widehat{\theta}$ around θ gives

$$\widehat{\theta} - \theta = Z_n/\sqrt{n} + b/n + r_n,$$

- $Z_n \sim \mathcal{N}(0, V_{11})$, $b = O(p)$ is a first order bias term, and r_n is the higher order remainder such as

$$r_n = O_p((p/n)^{3/2} + p^{1/2}/n)$$

- $\sqrt{n}(\widehat{\theta} - \theta) \sim \mathcal{N}(0, V_{11})$ holds if

$$\sqrt{n}b/n \to 0, \quad \text{i.e. } p^2/n \to 0,$$

AND

$$\sqrt{n}r_n \to_P 0, \quad \text{i.e. } p^{3/2}/n \to 0$$

Bias Corrections

- The bias is the bottleneck in the expansion, so can do two things:
  - A) *Analytical bias correction*: estimate b/n using analytical expressions for the bias and set
    $$\check{\theta} = \widehat{\theta} - \widehat{b}/n.$$
  - B) *Split-sample bias correction*: split the sample into two parts, compute the estimator on the two parts $\widehat{\theta}_{(1)}$ and $\widehat{\theta}_{(2)}$, and then set
    $$\check{\theta} = 2\widehat{\theta} - \overline{\theta}, \quad \overline{\theta} = (\widehat{\theta}_{(1)} + \widehat{\theta}_{(2)})/2.$$

    In some cases we can average over many splits to reduce variability
- With bias correction, a weaker rate requirement for GMM:
  $$(p \vee m)^{3/2}/n \to 0 \text{ as } n \to \infty$$
  which leads to
  $$\sqrt{n}(\check{\theta} - \theta) \sim \mathcal{N}(0, V_{11})$$

Why does the Sample-Splitting Method Work?

- The key is to split the sample such that the number of nuisance parameters and moment conditions are the same in all the parts
- Then, assuming that the parts are homogenous, the first order biases of $\widehat{\theta}$, $\widehat{\theta}_{(1)}$, and $\widehat{\theta}_{(2)}$ are

$$\frac{b}{n}, \quad \frac{b}{n/2}, \quad \frac{b}{n/2}$$

- The first order bias of $\check{\theta}$ is

$$2\frac{b}{n} - \left(\frac{1}{2}\left[\frac{b}{n/2}\right] + \frac{1}{2}\left[\frac{b}{n/2}\right]\right) = 0$$

Implementation of Split-Sample Bias Correction

- Need to determine the right partition of the data
- In FE approach: halve the panel along the time series dimension (Dhaene and Jochmans, 2015)

  Preserves the time series structure and delivers two panels with the same number of unit fixed effects and half the number of observations
- In AB approach: halve the panel along the cross section dimension

  Delivers two panels where the number of observations relative to the number of instruments is half of the original panel

  Can average across multiple splits to reduce variability because the cross-sectional ordering of the observations is arbitrary

Analytical Bias Correction for FE Approach

- First order bias b/n satisfies:

$$Hb = -\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\mathbb{E}[D_{i,s}\varepsilon_{i,t}], \quad H = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\widetilde{D}_{i,t}\widetilde{D}'_{i,t}],$$

  where $\widetilde{D}_{i,t}$ is the residual of the linear projection of $D_{i,t}$ on $X_{i,t}$
- $b = O(N)$ because the source of the bias is the estimation of the N unit fixed effects
- There is no bias from time fixed effects because the model is linear
- An estimator of the bias (Hahn and Kuersteiner, 2011)

$$\hat{H}\hat{b} = -\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{(t+M)\wedge T}\frac{D_{i,s}\widehat{\varepsilon}_{i,t}}{T-s+t}, \quad \hat{H} = \frac{1}{NT}\sum_{k=1}^{N}\sum_{t=1}^{T}\widetilde{D}_{i,t}\widetilde{D}'_{i,t},$$

  where $\widehat{\varepsilon}_{i,t}$ is the fixed effect residual and M is a trimming parameter such that $M/T \to 0$ and $M \to \infty$ as $T \to \infty$

MLE Inefficiency: An Example

Let $\widehat{\theta}_T = \overline{X}_T$ be the sample mean of realizations from a $\mathcal{N}(\theta, 1)$ population. Let $|\gamma| < 1$ and $\delta \in (0, 1/2)$ and define

$$\widetilde{\theta}_T = \begin{cases} \widehat{\theta}_T & |\widehat{\theta}_T| > T^{-\delta} \\ \gamma\widehat{\theta}_T & |\widehat{\theta}_T| \leq T^{-\delta} \end{cases}$$

- $\widehat{\theta}_T$ is the MLE of θ: $\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$
- Claim: $\widetilde{\theta}_T \to \widehat{\theta}_T$ w.p 1 when $\theta \neq 0$, $\widetilde{\theta}_T \to \gamma\widehat{\theta}_T$ w.p 1 when $\theta = 0$
  1. $Pr\left(|\widehat{\theta}_T| \leq T^{-\delta}\right) = Pr\left(\sqrt{T}(-T^{-\delta} - \theta) \leq Z_T \leq \sqrt{T}(T^{-\delta} - \theta)\right)$, where $Z_T = \sqrt{T}(\widehat{\theta}_T - \theta)$. If $\theta > 0$, $T^{-\delta} - \theta < 0$ for large T; if $\theta < 0$, $-T^{-\delta} - \delta > 0$ for large T. Either case entails $Pr\left(|\widehat{\theta}_T| \leq T^{-\delta}\right) \to 0$
  2. Similarly can show that $Pr\left(|\widehat{\theta}_T| \leq T^{-\delta}\right) \to 1$ when $\theta = 0$. Hence by definition of $\widetilde{\theta}_T$, the claim holds
- When $\theta = 0$, $\sqrt{T}(\widetilde{\theta}_T - \theta) \xrightarrow{d} \mathcal{N}(0, \gamma^2)$. **More efficient than MLE** because $\gamma^2 < 1$. Can further show that $\sqrt{T}(\widetilde{\theta}_T - \theta_T) \xrightarrow{d} L(h)$, where $\theta_T = h/\sqrt{T}$ and $\{X_t\}_{t=1}^{T}$ are from $\mathcal{N}(\theta_T, 1)$. $\widetilde{\theta}$ is not a regular estimator Back

MLE Inconsistency: An Example

Consider independent $\{X_{i,t}\}$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$ where

$$X_{i,t} \sim \mathcal{N}(\mu_i, \sigma^2)$$

Assume short panel: T is fixed and $N \to \infty$

Goal: estimate σ^2 consistenty

- MLE estimators:

$$\widehat{\mu}_i = \overline{X}_i, \quad \widehat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{i,t} - \overline{X}_i)^2$$

- Note $\widehat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sigma^2 W_i$, where W_i are independent χ^2_{T-1}
- By LLN, $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2 \frac{T-1}{T}$, **inconsistent**
- Why? $\widehat{\mu}_i$ is nuisance, and its dimension grows with N (incidental parameters). With T fixed, the estimate is noisy and affects $\widehat{\sigma}^2$
- Can adjust $\widehat{\sigma}^2$ by multipling $\frac{T}{T-1}$, in general hard ◂ Back

Semiparametric Nonlinear Panel

- Data: $Y = (Y_1, \ldots, Y_T)'$ and $X = [X_1, \ldots, X_T]'$
- Goal: identify effects of X_t on Y_t controlling for unobserved variables correleated with X_t
- Distribution of outcome Y_t given observables X_t

$$Y_t | Y^{t-1}, X_t, \alpha \sim f_{Y_t}(\cdot | Y^{t-1}, X_t, \alpha; \theta), \quad t = 1, \ldots, T$$

$Y^{t-1} = (Y_0, \ldots, Y_t)$ and $X^t = (X_1, \ldots, X_t)$

- α: unobserved fixed effects (individual)
- f: known conditional density of θ (parametric part)
- Doesn't specify distribution of fixed effects nor relationships with covariates (nonparametric part)
- X_t predetermined w.r.t Y_t and Y_0 is initial condition

Back

Two Examples

- Normal Linear Model with Strictly Exogenous Regressors

$$Y_t | Y^{t-1}, X_t, \alpha \ \sim \ Y_t | X, \alpha \ \sim \ \mathcal{N}(X_t'\theta + \alpha, \sigma^2)$$

- Dynamic Binary Choice Model, e.g., Labor Force Participation

$$Y_t | Y^{t-1}, X_t, \alpha \ \sim \ \left[F_\varepsilon(\theta_y Y_{t-1} + X_t'\theta_x + \alpha)\right]^{Y_t} \left[1 - F_\varepsilon(\theta_y Y_{t-1} + X_t'\theta_x + \alpha)\right]^{1-Y_t}$$

where $Y_{i,t} = 1(X_{i,t}'\theta + \alpha_i \geq \varepsilon_{i,t}), \quad \varepsilon_{i,t} | X_i^t, \alpha \ \sim \ F_\varepsilon$

◂ Back

MLE of Nonlinear Panel

- FE is maximizing the conditional log likelihood

$$\mathcal{L}(\theta, \alpha_1, \ldots, \alpha_N) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f_{Y_t}(Y_{i,t} | y_i^{t-1}, X_i^t, \alpha_i; \theta)$$

- Concentrate out α_i, sub in and maximize over θ:

$$\widehat{\alpha}_i(\theta) = \arg\max_{\alpha} \sum_{t=1}^{T} \log f_{Y_t}(Y_{i,t} | Y_i^{t-1}, X_i^t, \alpha; \theta)$$

$$\widehat{\theta} = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f_{Y_t}(Y_{i,t} | Y_i^{t-1}, X_i^t, \widehat{\alpha}_i(\theta); \theta)$$

- $\widehat{\alpha}$ inconsistent: N grows and T fixed, dimension of grows with sample size, incidental parameter problem
- Why? Only T observations to estimate each α, so that as N grows the estimate of α remains random
- In linear panel, no such issue since α is differenced out

MLE of Nonlinear Panel: Inconsistency

- For fixed T, $plim_{N\to\infty}\widehat{\theta}$ is

$$\theta_T = \arg\max_{\theta} \sum_{t=1}^{T} \mathbb{E}[\log f_{Y_t}(Y_t|Y^{t-1}, X^t, \widehat{\alpha}(\theta); \theta)]$$

- Inconsistency of $\widehat{\alpha}$ leads to inconsistency of $\widehat{\theta}$
- Would be consistent if we replace $\widehat{\alpha}$ with

$$\overline{\alpha}(\theta) = \arg\max_{\alpha} \sum_{t=1}^{T} \mathbb{E}[\log f_{Y_t}(Y_{i,t}|Y_i^{t-1}, X_i^t, \alpha; \theta)]$$

- Interpretation: when T is large, we have a better estiamte of nuisance parameters. See Fernández-Val and Weidner (2018) for theoretical details.

Back

GMM Details of Arellano–Bond

- Moment conditions $\mathbb{E}\left[(\Delta y_{i,t} - \rho \Delta y_{i,t-1}) y_{i,s}\right] = 0$ can be writen as $\mathbb{E}(\boldsymbol{Z_i}'\overline{\boldsymbol{u}}_{\boldsymbol{i}}) = 0$, where $\boldsymbol{Z_i}$ is $(T-2) \times \frac{(T-1)(T-2)}{2}$ as follows

$$\boldsymbol{Z_i} = \begin{pmatrix} y_{i,1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i,1} & y_{i,2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i,1} & \cdots & y_{i,T-2} \end{pmatrix}$$

  and $\overline{\boldsymbol{u}}_{\boldsymbol{i}} = (\Delta u_{i,3}, \Delta u_{i,4}, \ldots, \Delta u_{i,T})'$
- GMM minimizes $\overline{\boldsymbol{u}}' \boldsymbol{Z W_N Z'} \overline{\boldsymbol{u}}$, where $\boldsymbol{Z}' = (\boldsymbol{Z_1}', \ldots, \boldsymbol{Z_N}')$ and $\overline{\boldsymbol{u}}' = (\overline{\boldsymbol{u}}_{\boldsymbol{1}}', \ldots, \overline{\boldsymbol{u}}_{\boldsymbol{N}}')$
- Estimator: $\widehat{\rho} = (\overline{\boldsymbol{y}}_{\boldsymbol{-1}}' \boldsymbol{Z W_N Z'} \overline{\boldsymbol{y}}_{\boldsymbol{-1}})^{-1} \overline{\boldsymbol{y}}_{\boldsymbol{-1}}' \boldsymbol{Z W_N Z'} \overline{\boldsymbol{y}}$, where $\overline{\boldsymbol{y}}_{\boldsymbol{i}}' = (\Delta y_{i,3}, \ldots, \Delta y_{i,T})$, $\overline{\boldsymbol{y}}_{\boldsymbol{i,-1}}' = (\Delta y_{i,2}, \ldots, \Delta y_{i,T-1})$; $\overline{\boldsymbol{y}}$ and $\overline{\boldsymbol{y}}_{\boldsymbol{-1}}$ are stacked up across i's as $\overline{\boldsymbol{u}}_{\boldsymbol{i}}$
- Optimal weighting matrix: $\boldsymbol{W_N} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{Z_i'} \widehat{\overline{\boldsymbol{u}}}_{\boldsymbol{i}} \widehat{\overline{\boldsymbol{u}}}_{\boldsymbol{i}}' \boldsymbol{Z_i}\right)^{-1}$, where $\widehat{\overline{\boldsymbol{u}}}_{\boldsymbol{i}}$ are residuals from first–step estimation

◂ Back

Weak Instrument in Arellano–Bond

- When $T = 3$, model is just–identified. IV regression

$$\Delta y_{i,2} = \pi y_{i,1} + r_i, \quad i = 1, \ldots, N$$

- Rewrite $y_{i,2} = \rho y_{i,1} + \alpha_i + u_{i,2}$ to be

$$\Delta y_{i,2} = (\rho - 1) y_{i,1} + \alpha_i + u_{i,2}$$

- Assume stationarity and denote $\sigma_\alpha^2 = Var(\alpha_i)$, $\sigma_u^2 = Var(u_{i,t})$,

$$plim\, \widehat{\pi} = (\rho - 1) \frac{k}{(\sigma_\alpha^2/\sigma_u^2) + k}, \quad k = \frac{(1-\rho)^2}{1-\rho^2}$$

- $plim\, \widehat{\pi} \to 0$ if $\rho \to 1$ or $\sigma_\alpha^2/\sigma_u^2 \to \infty$
- For sufficiently high ρ or variance of α_i, we have weak instrument
- Least square $\widehat{\rho} - 1$ biased upwards in general

◂ Back

GMM Details of Blundell–Bond

The matrix form is

$$\boldsymbol{Z_i}^+ = \begin{pmatrix} \boldsymbol{Z_i} & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i,1} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \Delta y_{i,T-1} \end{pmatrix}$$

where $\boldsymbol{Z_i}$ is defined for in GMM details for Arellano–Bond. Set up the moments as in Areallno–Bond and conduct two step estimation. ◂ Back