FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Advanced Econometrics

Take-home exam

Gerdie Everaert

2023-2024

GHENT UNIVERSITY

Main theme: dynamic panel data models

0. To warm up: OLS estimation
1. Fixed Effects (FE) estimation
   - Implement the FE estimator in R/Matlab
   - Simulate its properties in a dynamic setting
   - Use simulation methods for bias correction and inference
2. Generalized Method of Moments (GMM) estimation
   - Implement the GMM function in R/Matlab
   - Simulate its properties in a dynamic setting

To warm up: OLS estimation

- Sample R and Matlab **functions** implementing the basic OLS estimator are available on Ufora
  - inputs: y and X
  - output: OLS coefs, se, *t*-stats, *p*-values

- Replicate the OLS results (see next slide) for the following cigarette demand model (Baltagi, 2008)

$$\ln C_{it} = \alpha + \rho \ln C_{i,t-1} + \beta_1 \ln P_{it} + \beta_2 \ln Pn_{it} + \beta_3 \ln Y_{it} + \mu_{it}$$

with C_{it} the number of packs of cigarettes sold per person in state i in year t, P_{it} the price per pack, Y_{it} disposable income and Pn_{it} the minimum price found in neighboring states

  - data for 46 states over the period 1963-92 are available on Ufora ('Data_Baltagi.xlsx')

Table 8.1 Pooled Estimation Results.* Cigarette Demand Equation 1963–92

	$\ln C_{i,t-1}$	$\ln P_{it}$	$\ln Pn_{it}$	$\ln Y_{it}$
OLS	0.97	−0.090	0.024	−0.03
	(157.7)	(6.2)	(1.8)	(5.1)
Within	0.83	−0.299	0.034	0.10
	(66.3)	(12.7)	(1.2)	(4.2)
2SLS	0.85	−0.205	0.052	−0.02
	(25.3)	(5.8)	(3.1)	(2.2)
2SLS-KR	0.71	−0.311	0.071	−0.02
	(22.7)	(13.9)	(3.7)	(1.5)
Within-2SLS	0.60	−0.496	−0.016	0.19
	(17.0)	(13.0)	(0.5)	(6.4)
FD-2SLS	0.51	−0.348	0.112	0.10
	(9.5)	(12.3)	(3.5)	(2.9)
FD-2SLS-KR	0.49	−0.348	0.095	0.13
	(18.7)	(18.0)	(4.7)	(9.0)
GMM-two-step	0.70	−0.396	−0.105	0.13
	(10.2)	(6.0)	(1.3)	(3.5)
System GMM	0.70	−0.415	−0.003	0.09
	(8.8)	(4.3)	(0.1)	(3.4)

* Numbers in parentheses are t-statistics. All regressions except OLS and 2SLS include time dummies.
Source: Some of the results in this table are reported in Baltagi, Griffin and Xiong (2000).

Implementing the FE estimator

- Write a Matlab/R function that implements the FE estimator
  - inputs: y and X
  - output: FE coefs, se, t-stats (p-values)
- Note that it may be (computationally) convenient to
  - set up y as a $(T \times N)$ matrix and X as a $(T \times N \times K)$ matrix
  - loop over the N-dimension to compute

$$X' M_D X = \sum_{i=1}^{N} X_i' M_d X_i, \qquad X' M_D y = \sum_{i=1}^{N} X_i' M_d y_i$$

  with y_i and X_i the data for cross-sectional unit i and $M_d = I_T - T^{-1} \iota_T \iota_T'$
- Replicate the FE results reported by Baltagi (you can add time effects using dummies)

Simulating the properties of the FE estimator

Set up a Monte Carlo experiment to numerically illustrate the properties of the FE estimator in a dynamic setting

- Generate data from $y_{it} = \rho y_{i,t-1} + \varepsilon_{it}$ with $\varepsilon_{it} \sim N(0,1)$
  - use a burn-in of 25 periods (i.e., initialize $y_{i,-25} = 0$, generate y_{it} for $t = -24, \ldots, 0, 1, \ldots T$ and discard the first 25 draws)
  - by using $y_{i,0}, \ldots, y_{i,T}$ and after taking a lag, this implies that you have T effective observations
  - use 3 different values of $\rho = 0, 0.5, 0.9$
  - use all combinations of $T = 4, 10, 20, 50$ and $N = 10, 100$
- Evaluate the properties of the FE estimator $\widehat{\rho}_{FE}$ in the model $y_{it} = \alpha_i + \rho y_{i,t-1} + \varepsilon_{it}$
  - compare the mean of $\widehat{\rho}_{FE}$ to the population parameter ρ
  - compare the mean of the estimated standard error $\widehat{se}(\widehat{\rho}_{FE})$ to the true standard error $se(\widehat{\rho}_{FE})$

Use simulation methods for bias correction and inference

- Implement a half-panel jackknife bias-corrected FE estimator $\widehat{\rho}_{FEj}$
  - use the above Monte Carlo setup to compare the mean of $\widehat{\rho}_{FEj}$ to the population parameter ρ
- Implement a bootstrap to compute standard errors for the jackknife FE estimator
  - think about an appropriate resampling scheme
  - compare the mean of the bootstrapped standard error $\widehat{se}^{b}(\widehat{\rho}_{FEj})$ to the true standard error $se(\widehat{\rho}_{FEj})$