EC475 Problem set 2

Linear panel data models

Antoine Goujard

13/11/09

- Antoine Goujard,a.j.goujard@lse.ac.uk ;
- Office hours: S684, Wednesday 12.30 – 13.30 ;
- Class webpage:
  http://personal.lse.ac.uk/goujard/
  Under Teaching/EC475.
  1. Introduction to Gauss (in progress) ;
  2. Examples of codes (in Gauss and Stata).
- Useful references for STATA programming (nothing is required):
  1. Baum, An Introduction to Stata Programming ;
  2. Cameron and Trivedi, Microeconometrics using STATA.
- Useful references for GAUSS programming:
  1. http://www.aae.wisc.edu/aae637/gausscode.htm ;
  2. Gauss user's guide (Aptech) ;

The GPNL data file is generated using Monte-Carlo simulations ($\forall(i,t)$):

$$y_{it} = 0.5.x_{1it} - 0.3.x_{2it} + 1 + 0.7.z_{1i} - 0.2.z_{2i} + \alpha_i + \nu_{it}$$

With: $\sigma_\alpha = 0.5$ and $\sigma_\nu = 0.25$ and $\mathrm{E}(\alpha|x_2) \neq 0$, $\mathrm{E}(\alpha|z_2) \neq 0$.
This is a balanced panel $T = 4$ and $N = 50$.

In this problem set we assume that the true specification is:

$$y_{it} = \beta_1.x_{1it} + \beta_2.x_{2it} + \gamma_0 + \gamma_1.z_{1i} + \gamma_2.z_{2i} + \epsilon_{it}$$

And we use different linear panel data estimators (OLS, Fixed effects, between, one-factor GLS and Hausman-Taylor), based on different assumptions about the data generating process.

- Plain OLS: is inconsistent because: $E((x_{1it}, x_{2it}, 1, z_{1i}, z_{2i}).\epsilon_{it}) \neq 0$ (even **A3Rsru** does not hold) (even if consistent OLS is inefficient as $\sigma_\alpha \neq 0$). As **A3Rmi** do not hold, we do not have unbiasedness.
- Fixed effects/Within: consistent for (β_1, β_2) not efficient. The transformed model is:

$$y_{it} - \overline{y}_{i.} = \beta_1.\underbrace{(x_{1it} - \overline{x}_{1i.})}_{\tilde{x}_{1i}} + \beta_2.\underbrace{(x_{2it} - \overline{x}_{2i.})}_{\tilde{x}_{2i}} + \nu_{it} - \overline{\nu}_{i.}$$

  Here $E(\nu_{it} - \overline{\nu}_{i.}|\tilde{x}_1, \tilde{x}_2) = 0$ so we have also unbiasedness.
- GLS is not feasible. We use FGLS estimating in a first step $\sigma_\alpha,\ \sigma_\nu$ to estimate: $\hat{\lambda}_i = 1 - \sqrt{\frac{\hat{\sigma}_\nu^2}{T_i.\hat{\sigma}_\alpha^2 + \hat{\sigma}_\nu^2}}$ and λ_i demean the variables at the individual level. The new error term is:

$$\zeta_{it} = \epsilon_{it} - \lambda_i.\overline{\epsilon}_{i.} = (1 - \lambda_i).\alpha_i + \nu_{it} - \lambda_i.\overline{\nu}_{i.}$$

- GLS (continued): ζ_{it} is homoskedastic and serially uncorrelated:

$$\text{var}(\zeta_{it}) = (1-\lambda_i)^2\sigma_\alpha^2 + \text{var}(\nu_{it} - \lambda_i.\overline{\nu}_{i.}) + 0$$

As α_i and ν_{it} are independently generated.

$$\text{var}(\zeta_{it}) = ... + \sigma_\nu^2 + \lambda_i^2.\sigma_\nu^2.1/T_i + 2\,\text{cov}(\nu_{it}, -\lambda_i.\nu_{it}.1/T_i)$$

Thus, $\text{var}(\zeta_{it}) = \sigma_\nu^2$.
Moreover, $\text{cov}(\zeta_{it}, \zeta_{it'}) = 0$, if $t \neq t'$.
So the transformed regression satisfies **A4**, and the GLS estimator should be consistent if $\text{E}(\alpha_i + \nu_{it}|x, z) = 0$ (not the case here) and efficient (if the one factor model is true).

- Hausman and Taylor (1981):
  $y_{it} = \beta_1.x_{1it} + \beta_2.x_{2it} + \underbrace{\gamma_0 + \gamma_1.z_{1i} + \gamma_2.z_{2i}}_{d_i} + \epsilon_{it}$
  We assume x_2 and z_2 may be correlated with α_i and apply the 2 step- estimator:
  1. Run **FE**, get $\hat{\beta}_{fe}$ (consistent) and compute $\hat{d}_i = \bar{y}_{i.} - \bar{x}_{i.}\hat{\beta}_{fe}$
  2. Use **2SLS** on : $\hat{d}_i = \gamma_0 + \gamma_1.z_{1i} + \gamma_2.z_{2i} + v_i$
     where v_i is a new error term and where we instrument z_{2i} by $\bar{x}_{1i.}$.
- **Rk1:** By def., the new error term is:

$$v_i = \bar{\epsilon}_{i.} - \bar{x}_{i.}.(\underbrace{\hat{\beta}_{fe} - \beta}_{=o_p(1)})$$

- **Rk2:** Need more variables x_1 than z_2.
- **Rk3:** STATA xthtaylor implements a one step version of this estimator (more efficient), using IVs: $x_{1it}, x_{2it} - \bar{x}_{2i.}, z_{1i}, \bar{x}_{1i.}$

VAR	OLS	GLS	BET	FE	HTAY	TRUE
ONE	1.1053	1.0998	1.1087	.	1.0576	1
X1	0.5206	0.5321	0.4963	0.5340	0.5340	0.5
X2	-0.3628	-0.3364	-0.3693	-0.3288	-0.3288	-0.3
Z1	0.5157	0.5172	0.5120	.	1.0414	0.7
Z2	-0.0321	-0.0429	-0.0251	.	-0.3774	-0.2

Hausman specification tests:

**** QUAD FORM GLS/FE: 1.1267458

**** PVALUE ($> \chi^2_2$): 0.56928568

**** QUAD FORM GLS/BETW: 1.1557304

**** PVALUE ($> \chi^2_2$): 0.56109492

**** QUAD FORM FE/BETW: 1.1566248

**** PVALUE ($> \chi^2_2$): 0.56084405

**** QUAD FORM OLS/GLS: 8.2343601

**** PVALUE ($> \chi^2_5$): 0.14378380

- $H_0 : \; E(\alpha|x, z) = 0$ vs $H_1 : \; E(\alpha|x, z) \neq 0$.
- The 1st three tests should be $=$ (Baltagi, 4.3 or Hausman-Taylor, 1981). An "intuition" for that is that under H_0 all estimators are consistent and thus we expect the quadratic form to be close to 0. Under $H_1 = H_a$, only *FE* is consistent, so the intuition for the test GLS/FE is clear. But it is possible to write the GLS estimator as a (matrix) weighted average of the FE estimator and the between estimator. So the *plim* of $GLS - BET$ and the *plim* of $FE - BET$ will be $\neq 0$ under H_1.
  This fact gives us some power to test H_0.
- Here, none of our 3 tests allows to reject H_0. They have **Low power**.

- Not clear how to interpret OLS vs GLS. If $H_0: \; E(\alpha|x,z) = 0$ is true both OLS and GLS are consistent and GLS is more efficient if $\sigma_\alpha^2 \neq 0$. Under $H_1 = H_a$ both estimators are inconsistent. Thus the test results are difficult to interpret because it can still be the case that under H_1: $(\hat{\beta}_{GLS}, \hat{\gamma}_{GLS}) \simeq (\hat{\beta}_{OLS}, \hat{\gamma}_{OLS})$ or that:
  (*) $plim(\hat{\beta}_{GLS}, \hat{\gamma}_{GLS}) = plim(\hat{\beta}_{OLS}, \hat{\gamma}_{OLS})$
  Hence under H_0, the difference between the 2 estimators should be close to 0 but this may also be the case under H_1. So the power of the test is impossible to evaluate and could be $= 0$ if (*) is true under H_1 (as it is already true under H_0).
- Computing a Hausman test for HT vs GLS would require to correct the var-covar of the HT estimator.