Advanced Econometrics II
Homework Assignment No. 1

Deadline: 12.01.2015, 23:59

Submit to: `A.Borowska@tinbergen.nl`

Please submit your (typed) solution in a `pdf` file. For programming exercises (if there are any), the code (preferably a Matlab one) has to be put, together with the main solution file, in an archive file (e.g. `zip` or `rar`). Each code file shall contain your name, either as a comment or in its name.

Question 1

For this question we consider two cases in which correlation between the error terms and the regressors arises (endogeneity).

(a) *(Measurement errors)*
First, we consider measurement errors. For this case consider the "true" model

$$y^o = X^o\beta + u^o, \quad u^o \sim N(0, \Omega), \quad \Omega = \sigma_0^2 I_n, \quad \mathbb{E}(u^o|X^o) = 0,$$

where y^o is n-by-1 and X^o is n-by-k. However, instead of observing y^o and X^o directly, we observe

$$\begin{aligned} y &= y^o + v_1, & v_1 &\sim N(0, V_1) \\ X &= X^o + v_2, & \text{vec}(v_2) &\sim N(0, V_2), \end{aligned}$$

where v_1 and v_2 are independently distributed.

Determine whether the OLS estimator based on the observed sample is still unbiased and consistent. If this is not true then discuss where the problem arises.

(b) *(Misspecified dynamics)*
Second, we consider the case of an AR(1) model with the autocorrelated error terms. Hence, consider the model

$$\begin{aligned} y_t &= \alpha + \beta y_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \rho\varepsilon_{t-1} + u_t, \\ u_t &\sim N(0, \sigma^2). \end{aligned}$$

Determine whether here the OLS estimator is unbiased and/or consistent. If this is not true then discuss where the problem arises.

(c) *(Solution)*
Discuss how you would solve this problem for one of the two situations.

Question 2

Next, consider the model

$$y = X\beta_0 + u, \quad \mathbb{E}(u_t|X_t) \neq 0, \quad \mathbb{E}(u|W) = 0, \quad \mathbb{E}(uu^T|W) = \sigma_0^2 I_n,$$

where y is n-by-1 and X is n-by-k, and W is a n-by-l (where $l > k$, i.e. overidentification) matrix with valid and relevant instruments. Let the l-by-k matrix J be a weighting matrix.

(a) *(Asymptotic distribution)*
Derive the asymptotic distribution of the IV estimator based on the instruments WJ.

(b) *(Generalized IV estimator)*
Using $J = (W^T W)^{-1} W^T X$ we obtain the Generalized IV estimator. Show that this choice of J leads to valid and relevant instruments.

(c) *(Asymptotic efficiency)*
Show that the choice of the weighting matrix J as in (b) is asymptotically efficient (i.e. it minimises the asymptotic variance of the IV estimator).

(d) *(Confidence region)*
Show how a researcher with a large data set can construct a 95% asymptotic confidence region for β_0, explain all steps.

(e) *(Additional instruments)*
Suppose that we can obtain a set of m extra valid and relevant instruments which are available through the matrix W^*. How can we benefit from these additional instruments? Show that it is asymptotically efficient to use these extra instruments.

Question 3

In this simulation exercise we will investigate some finite sample properties of the IV estimator. For this purpose, consider the following DGP that simultaneously determines the vectors x and y which are both of dimension n-by-1:

$$y = x\beta_0 + \sigma_u u,$$
$$x = w\pi_0 + \sigma_v v.$$

Take the following parameter values $\sigma_u = \sigma_v = 1$, $\pi_0 = 1$, $\beta_0 = 1$, $\rho = 0.5$, where ρ stands for the correlation between u and v. For the exogenous instrument w use independent drawings from the standard normal distribution and then rescale w so that $w^T w = n$.

(a) *(Convergence to the asymptotic distribution*
Generate at least 1000 datasets for $n = 10$, $n = 100$ and $n = 1000$, for each dataset compute both the OLS estimator and the IV estimator. Then show in a graph for each value of n the empirical distribution function of the OLS and IV estimator and the CDF of the normal distribution with mean zero and variance $(\sigma_u^2/n)\pi_0$. Explain why this is an appropriate way to compare the finite-sample and asymptotic distributions of the estimator. Interpret your results.

(b) *(Including additional instruments)*

Generate additional instruments z in the same way as w was generated above, however keep the additional instruments z independent of everything else. Redo the analysis above with $n = 10$, but now replace the IV estimator based only on w by the Generalized IV estimator that also uses the additional instruments z in the optimal way. Afterwards redo this analysis using two and four additional instruments. What is the theoretical prediction on the effect of including more instruments on the finite sample distribution? Present your results and discuss the effects of including these additional (irrelevant) instruments.