# Advanced Econometrics II
# Homework 1

Deadline: 2011-01-17 09:30

You are to hand in your homework via email to ta@zamojski.net by 9:30 on Tuesday. I would appreciate if your answers were TeXed in full, if you insist you can scan your handwritten answers at TI and send them to me as well.

You can work in groups of two.

If there is a computational exercise included in the homework, the quality of your coding will be judged (commenting, efficiency, etc.). Your code should be easy to read. Things that make it easier: lots of comments (e.g. explaining what loops are meant to do), camel case variable names (e.g. mErrorsUniformlyDistributed), consistency. You are to include your code in the body of the report, e.g. if you are TeXing your answers then with the listings package. You can use Ox, Python, C++, or Matlab. If I am not able to run your code after extracting your answers (assuming they are zipped) to a separate folder you will lose points. Looking ahead, it is in your best interest to combine Ox (computations, sometimes graphics) and Python (database management, multiprocessing management, graphics) as it will cut the time needed for simulations considerably. In the empirical exercises you are expected to provide comments for your results and methods (e.g tests) used.

## 1 Exercise 1

Consider a very simple consumption function, of the form

$$c_i = \beta_1 + \beta_2 y_i^* + u_i^* \tag{1}$$

$$u_i^* \sim \text{IID}\left(0, \sigma^2\right) \tag{2}$$

where $c_i$ is the logarithm of consumption by household $i$, and $y_i^*$ is the permanent income of household $i$, which is not observed. Instead, we observe current income $y_i$, which is equal to $y_i^* + v_i$, where $v_i \sim \text{IID}\left(0, \omega^2\right)$ is assumed to be uncorrelated with $y_i^*$ and $u_i^*$. Therefore, we run the regression

$$c_i = \beta_1 + \beta_2 y_i + u_i \tag{3}$$

Under the plausible assumption that the true value $\beta_{20}$ is positive, show that $y_i$ is negatively correlated with $u_i$. Using this result, evaluate the *plim* of the OLS estimator $\hat{\beta}_2$, and show that this *plim* is less than $\beta_{20}$.

## 2 Exercise 2

Show that the GIV estimator $\hat{\beta}_{IV}$ is consistent by explicitly computing the probability limit of the estimator for a DGP such that

$$y = X\beta_0 + u \tag{4}$$

What assumptions do you need to make? Given the answers to this exercise and Exercise 1, which estimator do you prefer?

# 3 Exercise 3

Suppose that $W_1$ and $W_2$ are, respectively, $n \times l_1$ and $n \times l_2$ matrices of instruments, and that $W_2$ consists of $W_1$ plus $l_2 - l_1 > 0$ additional columns. Prove that the generalized IV estimator using $W_2$ is asymptotically more efficient than the generalized IV estimator using $W_1$ . To do this, you need to show that the matrix $\left(\mathbf{X}^{\mathrm{T}}\mathbf{P}_{\mathbf{W_1}}\mathbf{X}\right)^{-1} - \left(\mathbf{X}^{\mathrm{T}}\mathbf{P}_{\mathbf{W_2}}\mathbf{X}\right)^{-1}$ is positive semidefinite, why? You can use the result of exercise 3.8 without deriving it.

# 4 Exercise 4

## a)

Use the DGP (8.40) to generate at least 1000 (a well written code should handle many more, e.g. $10^7$, in a reasonable amount of time) sets of simulated data for $x$ and $y$ with sample size $n = \{10, 100, 1000\}$, using normally distributed error terms and parameter values $\sigma_u = \sigma_v = 1$, $\pi_0 = 1$, $\beta_0 = 0$, and $\rho = 0.5$. For the exogenous instrument $w$, use independent drawings from the standard normal distribution, and then rescale $w$ so that $\mathbf{w}^{\mathrm{T}}\mathbf{w}$ is equal to $n$, rather than 1 as in Section 8.4.

For each simulated data set, compute the IV estimator (8.41). Then draw the empirical distribution of the realizations of the estimator on the same plot as the CDF of the normal distribution with mean zero and variance $\frac{\sigma_u}{n\pi_0}$. Explain why this is an appropriate way to compare the finite-sample and asymptotic distributions of the estimator. How quickly does the finite-sample distribution of the IV estimator converge to the asymptotic distribution?

In addition, for each simulated data set, compute the OLS estimator, and plot the EDF of the realizations of this estimator on the same axes as the EDF of the realizations of the IV estimator.

You may also want to plot histograms of thus created data to help with the conclusions.

## b)

What do you expect to happen if you changed the values of $\pi_0$, and $\rho$? Why? Redo this exercise for other values of these parameters (e.g. $\pi_0 = 0.4$). Comment on your results. Given your results which estimator, IV or OLS do you prefer?

## c)

Redo the simulations in a) generating the exogenous instrument $w$ as follows. For the first experiment, use independent drawings from the uniform distribution on $[-1, 1]$. For the second, use drawings from the $AR(1)$ process

$$w_t = \alpha w_{t-1} + \varepsilon_t \tag{5}$$

where $w_0 = 0$, $\alpha = 0.8$, and the $\varepsilon_t \sim \mathrm{N}(0, 1)$. In all cases, rescale $w$ so that $\mathbf{w}^{\mathrm{T}}\mathbf{w} = n$. To what extent does the empirical distribution of $\beta_{IV}$ appear to depend on the properties of $w$? What theoretical explanation can you think of for your results?

## d)

Include one more instrument in the simulations of a). Continue to use the same DGP for $y$ and $x$, but replace the simple IV estimator by the generalized one, based on two instruments $w$ and $z$, where $z$ is generated independently of everything else in the simulation. Repeat the simulations for three additional instruments. What happens to the distribution of the estimator as the number of instruments increases (you may want to plot EDFs and histograms)? Relate your answer to Exercise 3. Can you increase the number of instruments indefinitely?