Final Exam
Econometrics, PhD, Spring 2022
Sharif University of Technology, MH Rahmati

1. **(30 points)** Consider the following model:

$$y_{1i} = \beta x_{1i} + u_i$$
$$y_{2i} = \beta x_{2i} + u_i$$
$$U_{1i} = y_{1i} + \nu_{1i}$$
$$U_{2i} = y_{2i} + \nu_{2i}$$
$$u_i = \alpha_1 \nu_{1i} + \alpha_2 \nu_{2i} + \omega_i$$
$$i = 1,2,\cdots,n$$

define

$$y_i = \begin{cases} y_{1i} & if\ U_{1i} - U_{2i} > 0 \Leftrightarrow D_i = 1 \Leftrightarrow A_i \\ y_{2i} & if\ U_{1i} - U_{2i} < 0 \Leftrightarrow D_i = 0 \Leftrightarrow \overline{A_i} \end{cases}$$

Assume that (ν_{1i}, ν_{2i}) are i.i.d. $N(0,0,1,1,\rho)$, $\omega_i$are i.i.d $N(0,\sigma^2)$ independent of (ν_{1i}, ν_{2i}), and the statistician observes D_i, y_i, x_{1i}, x_{2i} for every i. The unknown parameters of the model are $\alpha_1, \alpha_2, \beta, \rho\ and\ \sigma^2$

   i. Derive the likelihood function
   ii. Define a two-step consistent estimator of β simpler to compute than the maximum likelihood estimator.

2. **(40 points)** Consider a linear panel data model with unobserved heterogeneity and a (potentially) endogenous explanatory variable, y_{it2}, where $y_{it2} > 0$:

$$y_{it1} = \alpha_1 y_{it2} + z_{it1}\delta_1 + c_{i1} + u_{it1}$$

$$E(u_{it1}|z_i, c_{i1}) = 0$$

where $z_{it} = (z_{it1}, z_{it2})$ and z_i is the entire history $\{z_{it} : t = 1,2,\ldots,T\}$.

   i. Assuming an appropriate rank condition – which you should provide – how would you estimate α_1 without further assumptions?
   ii. Let $\overline{z_i} = T^{-1} \sum_{r=1}^{T} z_{ir}$ be the vector of time averages of the exogenous variables and assume that

$$E(c_{i1}|z_i) = \psi_1 + \overline{z_i}\xi_1$$
$$E(y_{it2}|z_i) = \exp\left(\eta_2 + z_{it}\pi_2 + \overline{z_i}\xi_2\right)$$

   where $z_{it}\pi_2$ is assumed to depend on z_{it2}. Find $E(y_{it1}|z_i)$.

   iii. Consider the following two-step procedure.

   (a) Run a pooled nonlinear least squares estimation

$$\min_{\eta_2,\pi_2,\xi_2} \sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it2} - \exp(\eta_2 + z_{it}\pi_2 + \overline{z_i}\xi_2)]^2$$

   and obtain the fitted values, $\widehat{y_{it2}} = \exp\left(\widehat{\eta_2} + z_{it}\widehat{\pi_2} + \overline{z_i}\,\widehat{\xi_2}\right)$

   (b) Run the pooled OLS regression

y_{it1} on $\widehat{y_{it2}}, z_{it1}, 1, \overline{z_i}$.

Under the assumptions given in part (ii), is this two-step approach generally consistent for α_1 and δ_1? Explain two ways of performing inference and discuss the pros and cons of each. (No derivations are needed but careful explanations are.)

iv. Argue that if $E(y_{it2}|z_i) \neq \exp\left(\eta_2 + z_{it}\pi_2 + \overline{z_i}\xi_2\right)$ the approach from part (iii) is inconsistent. What is a better way of using the fitted values $\widehat{y_{it2}}$ generated in part (iii)?

v. If you extend the model to

$$y_{it1} = \alpha_1 y_{it2} + z_{it1}\delta_1 + y_{it2}z_{it1}\theta_1 + c_{i1} + u_{it1}$$

how would you estimate the parameters? Be very detailed in your description.

3. (30 points) Consider a structural equation:

$$y_t = \alpha d_t + \beta x_t + \varepsilon_t$$

Where $d_t$can be correlated with ε_t. Suppose you have access to an instrument variable z_t such that $E(\varepsilon_t|z_t) = 0$. Since d_t is a count variable that takes values $0,1,2,\cdots$ you specify a Poisson regression model, such that for $k = 0,1,2,\cdots$

$$P(d_t = k|x_t, z_t) = \frac{\lambda^k e^{-\lambda}}{k!}, \text{ where } \lambda = \exp(x_t'\alpha_1 + z_t'\alpha_2)$$

Suppose you estimate a Poisson regression model in the first stage and obtain the predicted value of $E[d_t|x_t, z_t]$, call it $\widehat{d_t}$

i. Will you be able to obtain consistent estimate for α and β if you regress y_t on $\widehat{d_t}$ and x_t in the second stage? State clearly any assumption you need to make to justify your answer.

ii. How will your answers change if you use $\widehat{d_t}$ as an instrument instead of regressor in the second stage regression?

iii. Given the information above what is the most efficient method you can come up with to estimate α and β? Derive your estimator!