London School of Economics
Department of Economics

Vassilis Hajivassiliou
Autumn Term 2024

# **MSc Econometrics** (Ec402)
## 2024-2025
## Problem Set #8

**Instructions:** To be discussed in week 10.

1. Consider the "Truncated" regression model with exogenous regressors, defined by:

$$y_s = \begin{cases} x_s'\beta + \epsilon_s & if \quad x_s'\beta + \epsilon_s > 0 \\ nothing\ observed & otherwise \text{ (not even the Xs)} \end{cases}$$

The sample is indexed by $s = 1, \cdots, S$ and the error term $\epsilon_s$ is distributed i.i.d.$N(0, \sigma^2)$ conditional on the regressors $X$. It can be shown that the following result holds about the distribution of $y|X$:

$$\text{Result1} : E(y_s|X, x_s'\beta + \epsilon_s > 0) = x_s'\beta + \sigma \frac{\phi\left(\frac{x_s'\beta}{\sigma}\right)}{\Phi\left(\frac{x_s'\beta}{\sigma}\right)}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution respectively.

(a) A colleague proposes to ignore Result1 and simply regress by OLS the observed $y_s$ on the observed vector of regressors $x_s'$. What is the thinking behind this proposal? How well do you expect this method to work in practice? Explain.

(b) A colleague proposes instead to use Result 1 and estimate $\beta$ and $\sigma$ by defining the Nonlinear Least Squares (NLLS) estimators:

$$\begin{pmatrix} \hat{\beta}_{nlls} \\ \hat{\sigma}_{nlls} \end{pmatrix} \equiv \arg\min_{\beta,\sigma} \sum_{s=1}^{S} \left( y_s - x_s'\beta - \sigma \frac{\phi\left(\frac{x_s'\beta}{\sigma}\right)}{\Phi\left(\frac{x_s'\beta}{\sigma}\right)} \right)^2$$

What is the intuition behind this proposal? When would you expect this method to perform well?

(c) Which proposal, (a) or (b), do you think is more likely to produce better estimates? Explain intuitively your answer. **NB: do not try to obtain analytical results that prove your conjectures about the NLLS method — only the intuition is required.**

2. Consider the regression model

$$y = X\beta + \epsilon, \qquad X \text{ of dimension } T \times k,$$

where the $\epsilon_t$ are iid $(0, \sigma_\epsilon^2)$ and the rows of the $X$ matrix $x_t'$, $t = 1, \cdots, T$, are random vectors of order $k$, iid $(\mu_x, \Sigma_x)$ and independent of the errors. Show that

$$s^2 = \frac{1}{T-k} \sum_{t=1}^{T} \hat{\epsilon}_t^2$$

is consistent for the true $\sigma_\epsilon^2$, where $\hat{\epsilon}_t$ are the least squares residuals. Does the same result hold in case $\epsilon_t$ and $x_t$ are merely uncorrelated?