London School of Economics
Department of Economics

Vassilis Hajivassiliou
Autumn Term 2024

MSc Econometrics (Ec402)
2024-2025
Problem Set #5

Instructions: Prepare for week 7.

1. Consider a linear regression model with data sample (y, X) of size S, that satisfies:

$$A1 : rank(X) = k$$
$$A2 : y = X\beta^{true} + \epsilon^{true}, \qquad E\epsilon^{true} = 0$$
$$A3Rmi : E(\epsilon^{true}|X) = E\epsilon^{true}$$

An investigator wishes to estimate the unknown parameter vector β^{true} and use those estimates to test, in the best possible way, a set of r linear restrictions on this parameter vector specified as: $R\beta^{true} = q$ where R is $r \times k$ and q is $r \times 1$. She is unsure, however, as to which $A4$ assumption characterizes the variance-covariance matrix of the true error vector conditional on the X variables, i.e., what to assume exactly about:

$$varcov(\epsilon^{true}|X) = E(\epsilon^{true}\epsilon^{true\prime}|X)$$

The investigator suspects one of two possible models for $varcov(\epsilon^{true}|X)$ specified as:

Model I	$A4.GM:$	$E(\epsilon^{true}\epsilon^{true\prime}|X) = \sigma_\epsilon^2 I_S$	GM error vcov
Model II	$A4.\Omega_{ma1}:$	$E(\epsilon^{true}\epsilon^{true\prime}|X) = \sigma_\epsilon^2 \Omega_{ma1}$	MA1 with $corr(\epsilon_s^{true}, \epsilon_{s-1}^{true}) = \lambda = 0.45$

$$\text{where:} \quad \Omega_{ma1} = \begin{pmatrix} 1 & \lambda & 0 & 0 & \cdots & 0 & 0 \\ & 1 & \lambda & 0 & \ddots & \ddots & 0 \\ & & 1 & \lambda & \ddots & \ddots & \vdots \\ & & & \ddots & \ddots & 0 & 0 \\ & & & & 1 & \lambda & 0 \\ & & & & & 1 & \lambda \\ & & & & & & 1 \end{pmatrix}$$

Consequently, she considers two alternative estimators for the β^{true} coefficient vector:

$$\hat{\beta}_{ols} = (X'X)^{-1}X'y \qquad \hat{\beta}_{igls.ma1} = (X'\Omega_{ma1}^{-1}X)^{-1}X'\Omega_{ma1}^{-1}y$$

Outline the properties of the two estimators she considers and explain what your recommendation would be.

2. Consider the linear regression model with k regressors:

$$y_i = x_i'\beta^{true} + \epsilon_i$$

estimated from a (large) cross-section indexed by $i = 1, \ldots, N$. The $N \times k$ regressor matrix X has full column rank k. Every true error disturbance ϵ_i (for all i) is fully statistically independent of the matrix X (i.e., for all regressor variables and all data points).

A researcher wants to estimate the true coefficient vector β^{true} and considers Maximum Likelihood (ML) estimation.

(a) The researcher derives two ML estimators for β^{true} by assuming in turn the following *two alternative error processes*:

[G] i.i.d. $Gaussian(0, v)$ with marginal probability density function (p.d.f.)

$$f_G(\epsilon_i) = \frac{1}{\sqrt{2\pi v}} \exp(-\frac{\epsilon_i^2}{2v^2})$$

[L] i.i.d. $Logistic(0, v)$ with marginal probability density function (p.d.f.)

$$f_L(\epsilon_i) = \frac{\exp\left(-\frac{\epsilon_i}{v}\right)}{v\left(1 + \exp\left(-\frac{\epsilon_i}{v}\right)\right)^2}$$

Reminder: The parameter v is the *scale* parameter of each distribution — it is not necessarily equal to the variance of ϵ_s in all cases.

Denote the two ML estimators she will obtain by $\hat{\beta}_{ml.G}$ and $\hat{\beta}_{ml.L}$ respectively. Set up the log likelihood optimization problem for each of these two estimators. Which, if any, of these formulae have closed-form solutions?

(b) Are there conditions that would make some or all of these estimators *Best Linear Unbiased* ($BLUE$)? BUE?

3. In the two-variable model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \qquad i = 1, ..., 11,$$

suppose that $x_1'x_1 = 2$, $x_2'x_2 = 2$, $x_1'x_2 = 1$, $x_1'y = 1$, $x_2'y = 1$, $y'y = 4/3$ where x_1, x_2, and y are the column vectors with typical elements x_{1i}, x_{2i} and y_i respectively.

Assume $\epsilon_i \sim$ i.i.d.$N(0, \sigma_\epsilon^2)$. Suppose you would like to make out-of-sample predictions about the left-hand-side (dependent) variable for two hypothetical observations with the following characteristics:

Obs.	$\mathbf{x}_1$	$\mathbf{x}_2$
12	5	-2
13	3	-7

(a) Construct 80% prediction intervals for the dependent variable y for observations 12 and 13.

(b) Construct 80% prediction intervals for the *expected value* of y_{12} and y_{13}.

(c) Do the answers to (i) and (ii) differ? Why?

4. Consider the linear regression model:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \epsilon_t, \quad t = 1, \cdots, T,$$

with $\rho(X'X) = 4$ and $\epsilon | X \sim N(0, \sigma_\epsilon^2 I_T)$. We are interested in the following hypotheses:

(a) H_0: $\beta_2 - 3\beta_3 = 4$, $\beta_1 = 2\beta_4$

(b) H_0': $\beta_1 = 1$, $\beta_2 = 3\beta_4 - 1$, $\beta_3 = 0$

(c) H_0'': $\beta_2 = \beta_3 = 2\beta_4$

i. Explain how you would use a statistic of the form $W = (R\hat{\beta} - q)'(R\hat{V}(\hat{\beta})R'^{-1}(R\hat{\beta} - q)/r$ to carry out tests of the 3 null hypotheses above. Be explicit about the quantities R, q, and r, and the degrees of freedom involved.