London School of Economics
Department of Economics

Vassilis Hajivassiliou
Autumn Term 2024

M.Sc. Econometrics (Ec402)
2024–2025
Answers for Problem Set #7

Question 1

Consider a linear regression model where the classic A1-A4 assumptions are believed to hold, with sample indexed by $s = 1, \cdots, S$. In particular, it is maintained that:

$A3 : E(\epsilon|X) = E(\epsilon)$ and $A4 : \epsilon_s \tilde{}\, i.i.d.$ An investigator wishes to carry out formal tests of two alternative hypotheses:

H$_1$ ϵ_i is i.i.d. *Gaussian*with marginal probability density function (p.d.f.) - mean 0 and scale v:

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

H$_2$ ϵ_i is i.i.d. *Logistic* with marginal probability density function (p.d.f.) - mean 0 and scale v:

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

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.

Explain how the investigator should *best* construct CIs for the following quantities: (a) β_7^{true}, (b) $\beta_3^{true} + \beta_4^{true}$, and (c) $\beta_5^{true} \cdot \beta_6^{true}$.

You must answer under the two alternative scenaria described here, namely:

$Scenario1 : A1 + A2linear + A3 + A4 + A5G$

$Scenario2 : A1 + A2linear + A3 + A4 + A5L$

The key to this question is the fact that the Gaussian and Logistic distribution do *not* belong to the same family of distributions — they are logically distinct possibilities, without any natural "nesting" being possible. For such situations, we will have to answer *separately* for each possible A1-A5 scenario.

***Summary of Scenario1: A1-A5Gaussian

Under these assumptions:

1. the GM theorem applies so the OLS estimator for β^{true} , namely $\hat{\beta}_{ols}$ will be BLUE

2. also $\hat{\beta}_{ols} \equiv \hat{\beta}_{mle.iid.Gaussian}$ so MLE Theorem 2 will apply, thus OLS will be the Best, Linear or Nonlinear CUAN estimator. Furthermore, since in finite samples it will be Unbiased, Theorem 1 of MLE will conclude that OLS will be BUE (best unbiased, linear or non linear estimator).

Consequently, the *best* estimator on which to base the Confidence Interval calculations is the OLS one — *best* in all the aforementioned senses, for which the finite sample distributional result of the NLRM tells us:

$$\hat{\beta}_{ols}|X \sim N(\beta^{true}, \sigma^2(X'X)^{-1})$$

For the two linear cases (a) and (b), we can use:

$$R\hat{\beta}_{ols}|X \sim N(R\beta^{true}, \sigma^2 R(X'X)^{-1}R')$$

where for (a) $R = (0,0,0,0,0,0,1,0,\cdots,0)$; and for (b) $R = (0,0,1,1,0,\cdots,0)$

Whereas for the nonlinear case (c), we need to use the asymptotic/approximate DELTA method, which says:

$$g(\hat{\beta}_{ols})|X \sim N(g(\beta^{true}), \sigma^2 \frac{\nabla g}{\nabla \beta}(X'X)^{-1}\frac{\nabla g}{\nabla \beta}')$$

where for this case:

$$\begin{aligned} g(\cdot) &= \beta_5 \cdot \beta_6 \\ \frac{\nabla g}{\nabla \beta} &= (0,0,0,0,\beta_6,\beta_5,0,\cdots,0) \end{aligned}$$

***Summary of Scenario2: A1-A5Logistic

Under these assumptions:

1. the GM theorem applies so the OLS estimator for β^{true} , namely $\hat{\beta}_{ols}$ will be BLUE — since the GM theorem does *NOT* require any A5.

2. But given A5.Logistic, $\hat{\beta}_{ols} \neq \hat{\beta}_{mle.iid.Logistic}$ so MLE Theorem 2 will apply for the $\hat{\beta}_{mle.iid.Logistic}$, which will thus be the Best, Linear or Nonlinear CUAN estimator. Note that this optimality result does *not* apply for the OLS — only for the MLE.iid.Logistic.

   Note also that the MLE.iid.Logistic will be a Nonlinear function of y, so in general the MLE will be *biased*. Hence Theorem 1 of ME will *not* apply for finite samples. But for infinite or very large samples, Theorem 2 of MLE *will* apply, as already explained.

   Thus in this Scenario2, the investigator has two options: either she says she will continue using as *BEST* the definition of BLUE — in which case she can continue with the CI constructions above. Or she proposes to adopt the Best, Linear or Nonlinear CUAN definition, and repeat all the constructions based on the ME.iid.Logistic estimates.