EC402 (2023/24)

RAGVIR'S SPEAKING NOTES FOR CLASS

Vassilis PS 6 (WK 7, AT)

QUESTION 1

2018

(Q1) $y = \overset{T\times k}{X}\beta + \varepsilon$; A1, A2, A3 Rmi

- Model I (A4GM) vs. Model II (A4. Ω) (Assume σ_ε^2 is known).

(a) The researcher has 2 estimators:
$\hat{\beta}_{OLS} := (X'X)^{-1}X'y$ and $\hat{\beta}_{GLS} := (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ where $\Omega := \begin{bmatrix} 1 & 0.45 & 0 & \cdots & 0 \\ 0.45 & 1 & 0.45 & & 0 \\ 0 & 0.45 & 1 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$

Conditional on X,

- $Var(\hat{\beta}_{OLS}) = E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right] = \sigma_\varepsilon^2(X'X)^{-1}$
- $Var(\hat{\beta}_{GLS}) = E\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right] = \sigma_\varepsilon^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$

These are the variances of the respective OLS and GLS estimators under Model I.

The G/M theorem tells us that $Var(\hat{\beta}_{GLS}) - Var(\hat{\beta}_{OLS})$ is a P.D. matrix under A1-A4GM, i.e. OLS is BLUE.

2018

Given $H_0: R\beta = q$
vs $H_1: R\beta \neq q$,

the researcher can use either:

$$V_{OLS} := (R\hat{\beta}_{OLS} - q)' \left[R\ \hat{\sigma}^2_\varepsilon (X'X)^{-1} R' \right]^{-1} (R\hat{\beta}_{OLS} - q) \qquad \text{OR}$$

$$V_{GLS} := (R\hat{\beta}_{GLS} - q)' \left[R\ \hat{\sigma}^2_\varepsilon (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}\Omega^{-1}X (X'X)^{-1} R' \right]^{-1} (R\hat{\beta}_{GLS} - q)$$

but if Model I is correct then the test based on V_{OLS} will be more powerful (i.e. for a given size α, the probability of making a Type II error will be lower with V_{OLS}).

(b) Conditional on X,

$$Var(\hat{\beta}_{OLS}) = c^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

$$Var(\hat{\beta}_{GLS}) = c^2 (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}\ \Omega\ \Omega^{-1}X (X'\Omega^{-1}X)^{-1} = c^2 (X'\Omega^{-1}X)^{-1}$$

with $c^2 := \sigma^2_\varepsilon = \sigma^2_v (1+\theta^2)$
assuming $\varepsilon_t = \theta v_{t-1} + v_t$
with $v_t \overset{iid}{\sim} (0, \sigma^2_v)$ for $t = 1, \dots, T$

These are the variances under Model II. The G/M theorem tells us that $Var(\hat{\beta}_{OLS}) - Var(\hat{\beta}_{GLS})$ is

2018

a P.D. matrix (under A1-A4 Ω), i.e. GLS is BLUE.

Given $H_0: R\beta = q$
vs $H_1: R\beta \neq q$,

the researcher can use either:

$$V_{OLS} := (R\hat{\beta}_{OLS} - q)' \left[R\sigma_\varepsilon^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} R' \right]^{-1} (R\hat{\beta}_{OLS} - q) \quad \text{OR}$$

$$V_{GLS} := (R\hat{\beta}_{GLS} - q)' \left[R\, \sigma_\varepsilon^2 (X'\Omega^{-1} X)^{-1} R' \right]^{-1} (R\hat{\beta}_{GLS} - q)$$

but if Model II is correct, then the test based on V_{GLS} will be more powerful (i.e. for a given size α, the probability of making a Type II error will be lower with V_{GLS}).

2018

Intuition: Say $X_i \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$ for $i = 1, \dots, N$ with σ^2 known.

Consider 2 LUE's for μ: X_1 and $\bar{X}$ (and let's assume $x_1 = \bar{x} = x^*$ in our sample)

Say $H_0: \mu = \mu_0$
$H_1: \mu = \mu_1 > \mu_0$

We can use either of 2 tests (of size $\alpha = 0.01$).

(i) Reject H_0 iff $x_1 = x^* > \mu_0 + 2.33\sigma$, FTR H_0 otherwise

(ii) Reject H_0 iff $\bar{x} = x^* > \mu_0 + 2.33\,\sigma/\sqrt{N}$, FTR H_0 otherwise

Even though both tests have the same size, clearly test (ii) has a higher power!

(The result is being driven by the lower variance of the second estimator relative to the first.)

"Ragvir, I don't understand your previous slide. What's the intuition for your intuition?!" 2021

Ans: Let's work out the power functions (from EC400).

$$\beta_{(i)}(\mu_1) = 1 - P\left(X_1 < \mu_0 + 2.33\,\sigma \;\middle|\; \mu = \mu_1\right)$$

$$= 1 - P\left(X_1 - \mu_1 < (\mu_0 - \mu_1) + 2.33\,\sigma \;\middle|\; \mu = \mu_1\right)$$

$$= 1 - P\left(\frac{X_1 - \mu_1}{\sigma} < \frac{(\mu_0 - \mu_1)}{\sigma} + 2.33 \;\middle|\; \mu = \mu_1\right)$$

$$= 1 - F_Z\left(\frac{\mu_0 - \mu_1}{\sigma} + 2.33\right) \text{ where } Z \sim N(0,1).$$

$$\beta_{(ii)}(\mu_1) = \ldots$$

$$= 1 - F_Z\left(\frac{\sqrt{N}(\mu_0 - \mu_1)}{\sigma} + 2.33\right) \text{ where } Z \sim N(0,1).$$

Clearly then, $\beta_{(ii)}(\mu_1) > \beta_{(i)}(\mu_1)$ for $N > 1$.

[Note: if this slide is also hard, please review Type I/II errors, size, power.]

QUESTION 2

(Q2) NOTE: See Vassilis' solution for this question. I present slightly modified examples of intuition for each case...

(a) Autocorrelation w. a lagged dependent variable:

Suppose $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t$ for $t = 1, \dots, T$

Recall: so long as $E[x_t \varepsilon_t] = 0$ AND $E[y_{t-1}\varepsilon_t] = 0$ for all t (i.e. A3R SM holds), $\hat{\beta}_{OLS}$ is a consistent estimator for β as $T \to \infty$.

However, suppose $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$ with $v_t \overset{IID}{\sim} N(0, \sigma_v^2)$ for $t = 1, \dots, T$

Now, we can rewrite the model as:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \underbrace{\rho \varepsilon_{t-1} + v_t}_{\varepsilon_t}$$

but it also ~~holds that~~

$$y_{t-1} = \beta_1 + \beta_2 x_{t-1} + \beta_3 y_{t-2} + \varepsilon_{t-1}$$

So CLEARLY $\operatorname{cov}(\varepsilon_t, y_{t-1}) \neq 0$! A3R SM is violated.

$\operatorname{plim}_{T\to\infty} \hat{\beta}_{OLS} \neq \beta$ (use IV?)

KEY MESSAGE

Our choice of A2, meaning a specification to include a lagged dependent variable, together with our choice of A4, meaning a specification to allow autocorrelation, forces a violation of even the weakest form of A3 that is needed to establish consistency of our estimators, meaning even A3Rfm.

(b) Measurement error in an explanatory variable:

let $y_t = \beta_1 + \beta_2 w_t + v_t$ where $v_t \overset{IID}{\sim} N(0, \sigma_v^2)$ for $t = 1, \dots, T$ and assume $E[v|w] = 0$

Suppose w_t cannot be measured accurately; instead we observe x_t:

$$x_t = w_t + u_t \quad ; \text{ ASSUME } u_t \overset{IID}{\sim} (0, \sigma_u^2) \text{ AND } u \perp\!\!\!\perp v \text{ AND } u \perp\!\!\!\perp w$$

The model becomes:

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 (x_t - u_t) + v_t \\
&= \beta_1 + \beta_2 \underline{x_t} + \underbrace{(v_t - \beta_2 u_t)} \\
&= \beta_1 + \beta_2 x_t + \varepsilon_t
\end{aligned}
$$

So CLEARLY $Cov(x_t, \varepsilon_t) \neq 0$! A3R srn is violated!

$\underset{T \to \infty}{\text{plim}} \hat{\beta} \neq \beta$ (use IV?)

QUESTION 3

Q1. GAUSSIAN CASE

RAGNIR:

$$y_i = x_i'\beta + \varepsilon_i \text{ for } i=1,\dots,N$$

where $\varepsilon_i | X \overset{iid}{\sim} N(0, v)$ with $0 < v < \infty$ known; $\dim\{x_i\} = K$;

$$X := \begin{bmatrix} x_1' \\ \vdots \\ x_N' \end{bmatrix}$$ is an $N \times K$ matrix s.t. $\text{rank}\{X\} = K$;

and $X \perp\!\!\!\perp \varepsilon$. let $y := (y_1, \dots, y_N)'$.

Find $\hat{\beta}_{G,MLE}$.

VASSILIS: A1, A2, A3Rfi, A4GM, A5N

(a) $$\mathcal{L}(\beta; y, X) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{1}{2v}(y_i - x_i'\beta)^2\right]$$

$$= (2\pi)^{-N/2} v^{-N/2} \exp\left[-\frac{1}{2v}\sum_{i=1}^{N}(y_i - x_i'\beta)^2\right].$$

$$l(\beta; y, X) = C - \frac{1}{2v}\sum_{i=1}^{N}(y_i - x_i'\beta)^2, \text{ for some constant } C.$$

$$s(\beta; y, X) = \frac{1}{v}\sum_{i=1}^{N} x_i(y_i - x_i'\beta)$$

Define $\hat{\beta}_{G,MLE}$ s.t. $s(\hat{\beta}_{G,MLE}; y, X) = 0$ —— ①

That is, $$\sum_{i=1}^{N} x_i y_i = \sum_{i=1}^{N} x_i x_i' \hat{\beta}_{G,MLE}$$

$$\therefore \hat{\beta}_{G,MLE} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i$$

APPENDIX A: SOCs

Typically, you'd have to also check the SoCs. Question doesn't require it apparently but I feel you should see it $\geq$ once and at least for the Gaussian case:

$$\frac{\partial \mathcal{L}}{\partial \beta'}(\beta; y, X) = -\frac{1}{v}\sum_{i=1}^{N} x_i x_i'$$

Given $0 < v < \infty$ and $\mathrm{rank}\{X\} = K$, $\lambda_1\left\{-\frac{1}{v}\sum_{i=1}^{N} x_i x_i'\right\} < 0$ for any β, where $\lambda_1\{\cdot\}$ represents the largest eigenvalue; this confirms that $\mathcal{L}(\hat{\beta}_{G,MLE}; y, X) = \max_{\beta} \mathcal{L}(\beta; y, X)$ for any β in the parameter space.

(b) Define $S(\beta; y, X) = \sum_{i=1}^{N} (y_i - x_i'\beta)^2$

Then, $\hat{\beta}_{OLS} := \arg\min_{\beta} S(\beta; y, X) = \ldots = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i$

$= \hat{\beta}_{G.MLE}$

Since we have (a) A1: $\text{rank}\{X\} = K$;

(b) A2: $y_i = x_i'\beta + \varepsilon_i$ with $E(\varepsilon_i) = 0$, $i = 1, \ldots, N$;

(c) A3Rfi: $X \perp\!\!\!\perp \varepsilon$;

(d) A4GM: $E[\varepsilon\varepsilon' | X] = \nu I_N$;

the G/M theorem tells us that $\hat{\beta}_{G,MLE} = \hat{\beta}_{OLS}$ is the BLUE for β.

Indeed, since (e) A5N: $\varepsilon | X \sim N(0, \nu I_N)$ holds, $\hat{\beta}_{G,MLE}$ is in fact the BUE for β.

Q1. LOGISTIC CASE · part (a)

Now assume instead that $\varepsilon_i | X \overset{iid}{\sim} \text{Logistic}(0, v)$. Find $\hat{\beta}_{L,MLE}$.

$$\mathcal{L}(\beta; y, X) = \prod_{i=1}^{N} \exp\left(-\frac{1}{v}(y_i - x_i'\beta)\right) / v\left(1 + \exp\left(-\frac{1}{v}(y_i - x_i'\beta)\right)\right)^2$$

$$\ell(\beta; y, X) = \sum_{i=1}^{N} \left[-\frac{1}{v}(y_i - x_i'\beta) - \log(v) - 2\log\left(1 + \exp\left(-\frac{1}{v}(y_i - x_i'\beta)\right)\right)\right]$$

$$s(\beta; y, X) = \sum_{i=1}^{N} \left[\frac{1}{v}x_i - \frac{2\left[\exp\left(-\frac{1}{v}(y_i - x_i'\beta)\right)\right]}{\left[1 + \exp\left(-\frac{1}{v}(y_i - x_i'\beta)\right)\right]} \cdot \frac{1}{v}x_i\right]$$

$$= \frac{1}{v}\sum_{i=1}^{N} x_i \left[1 - 2\,\frac{\exp\left[-\frac{1}{v}(y_i - x_i'\beta)\right]}{\left[1 + \exp\left[-\frac{1}{v}(y_i - x_i'\beta)\right]\right]}\right]$$

(SOCs not necessary; would have to use quotient rule, I guess.)

We define $\hat{\beta}_{L,MLE}$ s.t. $\ell(\hat{\beta}_{L,MLE}; y, X) = 0$. —— ②

Recall the systems of equations in ① and ②:

① is a system of K LINEAR equations in K unknowns.

② — " —— NON-LINEAR —— " ——

For this reason no closed-form solution for $\hat{\beta}_{L,MLE}$ exists. The solution must be obtained numerically.

This is what "iterative solution on a computer" is referring to in the official solution. For a concrete example, see Appendix B.

APPENDIX B: Newton/Raphson Procedure

For example, one could use the Newton-Raphson method as follows:

- let $K=1$ so that β is a scalar. Let $\tilde{\beta}$ denote our best guess of $\hat{\beta}_{L,MLE}$ (or just β).
- We approximate $s(\hat{\beta}_{L,MLE}; y, X)$ using a 1[st] order Taylor expansion:

Set $$s(\hat{\beta}_{L,MLE}; y, X) \simeq s(\tilde{\beta}; y, X) + s'(\tilde{\beta}; y, X)(\hat{\beta}_{L,MLE} - \tilde{\beta}) = 0$$

$$\Rightarrow \quad \hat{\beta}_{L,MLE} = \tilde{\beta} - \frac{s(\tilde{\beta}; y, X)}{s'(\tilde{\beta}; y, X)}.$$

So an iterative procedure may be devised as follows:

- Start with initial guess $\tilde{\beta}_0$.
- Obtain $\tilde{\beta}_1 = \tilde{\beta}_0 - s(\tilde{\beta}_0; y, X) / s'(\tilde{\beta}_0; y, X)$
- Repeat for $\tilde{\beta}_{J+1} = \tilde{\beta}_J - s(\tilde{\beta}_J; y, X) / s'(\tilde{\beta}_J; y, X)$

  for $J = 1, 2, \ldots$,

  until convergence.

(This is the simplest numerical procedure. There are many others but unfortunately I'm not an expert on these.)

part (b) See Vassilis' solution

Appendix C: Typical Use for Logistic CDFs

(i) Say $y_i^* = x_i'\beta + \varepsilon_i$ for $i = 1, \dots, N$ s.t.

• latent

• observed

$$y_i := \begin{cases} 1, & y_i^* \geq 0 \\ 0, & \text{o/w} \end{cases}, \quad \text{where } \varepsilon_i \overset{\text{IID}}{\sim} f_{\varepsilon_i}(\cdot) \ \forall i.$$

(ii) Clearly, $y_i = 1 \iff y_i^* \geq 0 \iff \varepsilon_i \geq -x_i'\beta$.

(iii) Assuming $f_{\varepsilon_i}(\cdot)$ is symmetric, $P(\varepsilon_i \geq -x_i'\beta) = P(\varepsilon_i \leq x_i'\beta) = F_{\varepsilon_i}(x_i'\beta)$.

(iv) Assume

$$f_{\varepsilon_i}(u) = \exp\left[-u/v\right] \Big/ v\left[1 + \exp\left[-u/v\right]\right]^2, \quad u \in \mathbb{R}.$$

$$\text{Fun Facts:} \left\{ \begin{array}{l} \text{Mean} = 0 \\ \text{Variance} = v^2\pi^2/3 \\ \text{Skew} = 0 \\ \text{Kurtosis} = [\text{fat tails} \dots \text{like } t_7] \end{array} \right\}$$

$$\Rightarrow F_{\xi_i}(u) = \left(1 + \exp[-u/v]\right)^{-1}, \quad u \in \mathbb{R}$$

That's pretty much it. Now we're in business because...

$$y_i \overset{\text{iid}}{\sim} \text{Bernoulli}\left(F_{\xi_i}(x_i'\beta)\right) \text{ for } i = 1, \ldots, N.$$

So...

$$\mathcal{L}(\beta; y, X) = \prod_{i=1}^{N} F_{\xi_i}(x_i'\beta)^{y_i} \left[1 - F_{\xi_i}(x_i'\beta)\right]^{1-y_i}$$

$$\vdots$$