EC402 (2023/24)

RAGVIR'S SPEAKING NOTES FOR CLASS

---

Vassilis PS 4 (WK 5, AT)

# PLAN

~~(Q1)~~

~~(Q2)~~

~~(Q3)~~

( MAYBE: FWL THEOREM IF WE HAVE TIME. OTHERWISE SEE MY NOTES ON WEBPAGE )

PS6 2018

Q1) (a) $h_i = \beta_0 + \sum_{j=1}^{4} \beta_j d_{ji} + \underset{(1\times k)(k\times 1)}{x_i' \gamma} + \varepsilon_i$ for $i = 1, \dots, N$

Define: $\underset{N\times(K+5)}{Q} := [\, i \; d_1 \; d_2 \; d_3 \; d_4 \; X \,]$. Clearly, $i = \sum_{j=1}^{4} d_j$ since every individual belongs in 1 and only 1 category (by design).

$\therefore$ $\exists$ at least 1 perfect linear relationship among the regressors $\Rightarrow$ A1 is violated!

Define: $\underset{(K+5)\times 1}{B} := [\beta_0, \dots, \beta_4, \gamma']'$. Then, $\hat{B}_{OLS} = (Q'Q)^{-1} Q' h$ cannot be estimated since $\text{rank}(Q'Q) < (K+5)$ and $(Q'Q)^{-1}$ will not exist.

This is a classic instance of the "dummy variable trap".

(b) 2 possible ways to avoid the aforesaid "trap":

(i) $h_i = \beta_0 + \sum_{j=2}^{4} \beta_j d_{ji} + x_i' \gamma + \varepsilon_i$ for $i = 1, \dots, N$ → drop "d1"

Here, "individuals with guaranteed income level 1" becomes the reference or benchmark category.

$E[h|Q] = \beta_0 + X\gamma$ for this reference category, then for $j = 2, \dots, 4$,

$\beta_j$ is the marginal/additional effect (note: not necessarily positive!) on the conditional mean of $h$ for individuals belonging in the $j^{th}$ guaranteed income level group; that is, over and above that for individuals in the reference category.

Note: $\beta_0$ is the intercept for the reference category

$\beta_0 + \beta_2$ — " — group 2

$\beta_0 + \beta_3$ — " — group 3

$\beta_0 + \beta_4$ — " — group 4

(ii) $h_i = \sum_{j=1}^{4} \theta_j d_{ji} + x_i'\delta + \varepsilon_i$ for $i = 1, \dots, N$ → drop the constant

Here, we no longer have a reference category

i.e. The intercepts by category are simply $\theta_j$ for $j = 1, \dots, 4$ respectively.

[Tip: In practice, specification (i) is more commonly used. The choice of reference category is arbitrary and usually based on the question of interest to the researcher.]

(iii) How are the coefficients related?

$$\theta_1 = \beta_0 \; ; \; \theta_2 = \beta_0 + \beta_2 \; ; \; \theta_3 = \beta_0 + \beta_3 \; ; \; \theta_4 = \beta_0 + \beta_4 \; ; \; \delta = \gamma$$

The OLS residuals should be identical since $\hat{h} = Q\hat{\beta}_{OLS}$ will be identical in both regressions.

2018

Q2) $y = S\beta_1 + X_2\beta_2 + \varepsilon$ ; Note: $X_2$ cannot contain a constant!

$M_S := [I_T - S(S'S)^{-1}S']$ where $\underset{T\times 4}{S} = [d_1\ d_2\ d_3\ d_4]$

Consider $z = \delta_1 d_1 + \delta_2 d_2 + \delta_3 d_3 + \delta_4 d_4 + \eta$ ; • $[S'S]_{ij} = d_i'd_j$ for $i,j = 1,\dots,4$

• note then that $d_i'd_j = \begin{cases} T_i & \text{if } i=j \\ 0 & \text{o/w} \end{cases}$ so that $(S'S) = \mathrm{diag}\{T_1,\dots,T_4\}$

and $(S'S)^{-1} = \mathrm{diag}\left\{\frac{1}{T_1},\dots,\frac{1}{T_4}\right\}$

NOTATION: "$\mathbb{1}(\cdot)$" is an indicator function.
{Q1 denotes a "Quarter1" identifier ;
{Q2, Q3, Q4 defined analogously.

• Further, $S'z = \begin{bmatrix} d_1'z \\ d_2'z \\ d_3'z \\ d_4'z \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{T} z_i \mathbb{1}(Q1)_i \\ \sum_{i=1}^{T} z_i \mathbb{1}(Q2)_i \\ \sum_{i=1}^{T} z_i \mathbb{1}(Q3)_i \\ \sum_{i=1}^{T} z_i \mathbb{1}(Q4)_i \end{bmatrix} = \begin{pmatrix} \sum_{i=1}^{T_1} z1_i \\ \sum_{i=1}^{T_2} z2_i \\ \sum_{i=1}^{T_3} z3_i \\ \sum_{i=1}^{T_4} z4_i \end{pmatrix}$

• Thus, $\hat{\delta} = (S'S)^{-1}S'z = [\bar{z1},\dots,\bar{z4}]'$

• So, $S\hat{\delta}$ yields a $T\times 1$ matrix s.t. for any $t = 1,\dots,T$, the $t^{th}$ element of $S(S'S)^{-1}S'z$ is $\begin{cases} \bar{z1} & \text{if } t \text{ is a Q1 obs}, \\ \bar{z2} & \text{" Q2 "} \\ \bar{z3} & \text{" Q3 "} \\ \bar{z4} & \text{" Q4 "}. \end{cases}$

• The claimed structure of $M_S z = z - S\hat{\delta}$ follows.

(b) Consider the deseasonalised variables regression :

$$y^* = X^* \alpha_2 + \varepsilon^* \quad , \qquad \text{where } * \text{ denotes pre-multiplication by } M_S .$$

Then, $\hat{\alpha}_{2_{OLS}} = (X^{*\prime} X^*)^{-1} X^{*\prime} y^* = ((M_S X)' M_S X)^{-1} (M_S X)' M_S y = (X' M_S X)^{-1} X' M_S y = \hat{\beta}_{2_{OLS}}$ by the FWL theorem.

# APPENDIX A

2018

Dear all,

A student in one class asked me a question about the DV trap. The question was roughly as follows:

"Consider $y = \alpha + \beta_M M + \beta_F F + \varepsilon$ where M and F are male and female dummies. Obviously, this specification suffers from the DV trap since $i = M + F$ so either we drop $i$ or M or F. The student asked that if we drop $i$, we still have a linear relationship in the sense that $M + F = 1$. So why is this not a problem?"

MATH:

① • obviously, $X := [i \ M]$ is not a problem since $i$ and M are not collinear. The rank of X is full and $(X'X)^{-1}$ can be computed.

② • But this is identical to a regression of $y$ on $X := [M \ F]$. To see this:

consider
$$
\begin{aligned}
y &= \delta_M M + \delta_F F + \varepsilon \\
&= \delta_M M + \delta_F (1 - M) + \varepsilon \\
&= \delta_M M + \delta_F - \delta_F M + \varepsilon \\
&= \underline{\delta_F} + \underline{(\delta_M - \delta_F)} M + \varepsilon \\
&= \theta_0 + \theta_1 M + \varepsilon
\end{aligned}
$$

∴ since the two specifications are effectively the same, if you are happy with ① you must also be happy with ②.

2018

INTUITION:

I think the best way to approach this problem is to think directly in terms of rank$(X'X)$ since that is the precise condition to check for AI. When $X := [i\ M\ F]$, rank of $X'X$ will be 2 but also when $X := [M\ F]$ and when $X := [i\ M]$ and when $X := [i\ F]$.

Hope that helps,

Ragvir

Q3. · $z \in \mathbb{R}^n$, $z \sim (\mu_z, V_z)$

· constant $q \in \mathbb{R}^n$

(a) Define $w := q'z$. Find $E(w)$ & $Var(w)$.

$E(w) = q'E(z) = q'\mu_z$.

$Var(w) = q' Var(z) q = q' V_z q$.

(b) Suppose that for a particular choice of $q \neq 0$, $Var(w) = 0$. What does this say about the random vector $z$? What does it say about the *rank* of the variance-covariance matrix $V_z$?

(i) (ii)

(i) Since there exists a $q \neq 0$ s.t. $Var(q'z) = 0$, there exists a perfect linear relationship between the elements of $z$.

You don't need to read this (sketch of a) proof. Last year students wanted me to give extra intuition for the second paragraph of Vassilis' solution so I came up with the reasoning below. That's all this is!

(ii) let's assume WLOG that there exists just one vector $q \neq 0$ s.t. $Var(q'z) = 0$.

Then, recognise that

1. $Var(q'z) = q' V_z q = 0$ iff $V_z q = 0$ for $q \neq 0$;
2. $Null(V_z) := \{ q \in \mathbb{R}^n : V_z q = 0 \}$ and so
$$Nullity(V_z) := \dim\{Null(V_z)\} = 1;$$
3. By the rank-nullity theorem,
$$rank(V_z) + Nullity(V_z) = n.$$

$\therefore$ By 1, 2, 3, $rank(V_z) = n-1$.

# APPENDIX B: Intuition

## ATTEMPT 1

Let's see if I can think of an example for you for Q3:

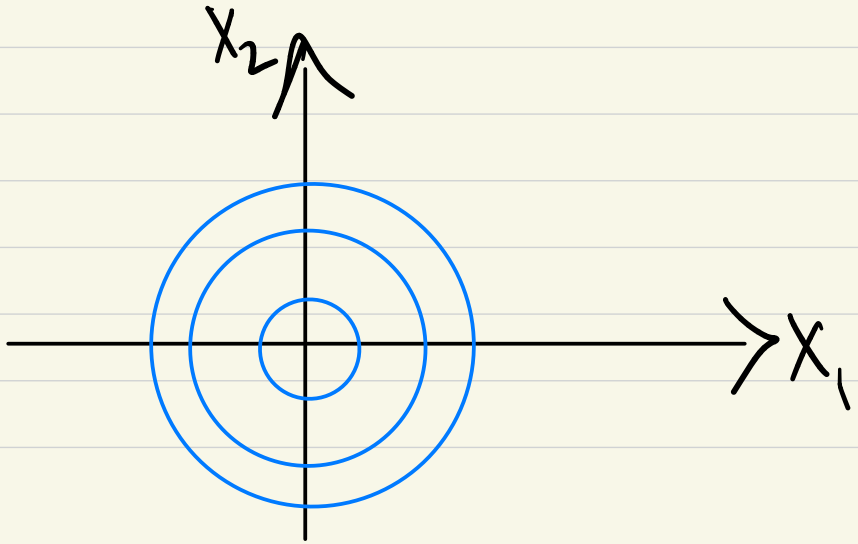
Consider $X \sim N(0, \sigma^2)$ and the vector $(X, 2X)'$.

Then $\begin{pmatrix} X \\ 2X \end{pmatrix}$ has variance $\begin{pmatrix} \sigma^2 & 2\sigma^2 \\ 2\sigma^2 & 4\sigma^2 \end{pmatrix}$, right?

Now ask yourself what is the rank of the matrix above?!

## ATTEMPT 2

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$$




$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$$ where $\sigma_{12} > 0$