

1. Lecture note 7: Question 1-(2)

$Y_i = \exp(\beta_0)L_i^{\beta_L}K_i^{\beta_K}\exp(u_i)$ 와 $\ln Y_i = \beta_0 + \beta_L \ln L_i + \beta_K \ln K_i + u_i$ 는 L,K가 양수라면 수학적으로 같은 식입니다. 강의에서는 첫 번째 식은 non-linear form이고 두 번째 식은 linear form이기 때문에 둘이 차이가 있고 그것을 그림을 통해 보여줬는데요.

그러나 첫 번째 식은 Y-L 축 or Y-K 축에서는 non-linear이고 두 번째 식은 lnY-lnL 축 혹은 lnY-lnK 축에서 linear한건데 두 식을 같은 평면에 놓고 비교하여 차이가 있다고 하는 게 이해가 되지 않습니다. 마치 $y = x^2$ 을 $\ln y = 2\ln x$ 로 바꾸고 두 그래프를 하나의 종이 위에 그려놓고 차이가 있다고 하는 것처럼 들립니다. 축이 다른, 그러나 실제로는 같은 두 그래프를 하나의 평면 위에 그려놓는게 의미가 있는지 모르겠습니다.

그리고 두 번째 linear form의 식이 점마다 달라지는 기울기를 잡아낼 수 없다고 하는데, 2 번째 식에서 β_L, β_K 를 하나씩만 구하기는 하지만 그걸 다시 exponential을 씌워서 1번 식처럼 바꾸면 각 점에서 기울기는 당연히 달라지지 않나요? 제가 어디를 잘못 이해한건지 모르겠습니다.

2. Lecture note 7: Question 1-(6)

회사마다 평균 임금이 크게 변하지 않는다 => 회사마다 β_L 의 추정값들이 비슷하게 나왔으니 그에 따른 임금도 비슷하다 => β_L 에 measurement error가 없다고 볼 수 있으니 IV 안써도 되고 L에 measurement error가 있다는 접근을 해야한다(?)

β_L 의 measurement error와 L의 measurement error를 어떤 방향으로 이해해야하는지 헛갈리고, β_L 은 L값을 통해 계산이 되는건데 L에 measurement error가 있어도 IV는 써야하지 않을까하는 생각이 들었습니다.

3. Lecture note 7: Question 2-(4)

여기서 more data를 regressor의 수가 늘어난다고 해석해야하는지 관측치의 개수가 늘어난다고 해석해야하는지 혼동이 오는데요. 강의에서는 more data면 R_squared가 늘어날 여지가 있다고 했는데 그럼 regressor의 수가 증가한다는 의미일까요?

4. Lecture note 8: Question 1- (3), (4)

강의에서는 $Y_i = \beta_0 + \beta_D D_i + u_i$ 에서 $\beta_D = E[Y_{1i} - Y_{0i}]$ 라고 나와있는데 공감이 되지 않습니다. 저 논리가 성립하려면 $P_i = 0$ 인 경우까지 포함해서 regression을 돌려야할 것 같은데요.

($\beta_D = E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 0] = E[Y_{1i} - Y_{0i}]$ 성립)

즉, for whom $P_i = 1$ 라는 조건이 없어야 대표님 논리가 맞다고 생각했습니다.

그런데 문제에서는 이미 $P_i = 1$ 인 경우만 고려한다고 나와있습니다. 그래서

$\beta_D = E[Y_i | D_i = 1, P_i = 1] - E[Y_i | D_i = 0, P_i = 1] = E[Y_{1i} | P_{1i} = 1] - E[Y_{0i} | P_{0i} = 1]$ 가 돼야하지 않을까 생각했습니다. 그리고 문제에서는 $E[Y_{1i} - Y_{0i} | P_{1i} = 1]$ 와 β_D 와의 관계를 물어봤으니

$E[Y_{1i} - Y_{0i} | P_{1i} = 1] = \beta_D + E[Y_{0i} | P_{0i} = 1] - E[Y_{0i} | P_{1i} = 1]$ 가 되고,

뒤의 $E[Y_{0i} | P_{0i} = 1] - E[Y_{0i} | P_{1i} = 1]$ 은 admission을 받지 못하고 의사가 된 사람의 2110년 연봉과 / admission을 받고 의사가 된 사람이 있는데 만약에 이 사람이 admission을 받지 못했다면 2110년에 받았을 가상의 연봉의 차이를 의미하게 됩니다. 즉, 의사가 된 사람들 중

에서 admission을 받은 사람들과 못 받은 사람들에 sampling bias가 있는지 없는지를 나타내는 항이라고 생각합니다. 그리고 (4)에서 2062년 이후의 출생자들로 regression을 하면 $P_{0i} = 1$ 인 부분이 사라지니($D_i = 0$ 이면 무조건 $P_i = 0$ 이 됨),

$\beta_D = E[Y_{1i}|P_{1i} = 1] - E[Y_{0i}|P_{0i} = 1] = \underline{E[Y_{1i}|P_{1i} = 1]}$ 이 된다고 생각했습니다.

이렇게 답안을 작성했을 시에 어떤 논리적 결함이 있는지 궁금합니다.

5. Lecture note 8: Question 3-(6)

강의 시간에 대표님께 질문드린 내용인데 논리가 잘 이해가 안되어 다시 의논해보고 싶어서 질문드립니다.

X_2^* 가 이미 Y 의 설명변수로 들어가있는데 IV로 사용해도 되는지에 대한 질문이었습니다.

대표님은 X_2^* 가 v 랑 exogenous하다고 가정하면 IV로 사용할 수 있다고 답을 해주셨는데요.

X_2^* 를 IV로 쓴다면 $Y_i = \alpha_0 + \alpha_1 X_{1i} + e_i$ 꼴로 IV regression을 돌린다는 뜻인데,

e_i 에 이미 X_2^* (일종의 omitted variable)와 v 가 포함되어 있기 때문에 $Cov(X_{2i}^*, e_i) \neq 0$ 가 되어서 orthogonality가 깨진다는게 제 생각입니다.