

QE 2016

Start time :-

2020 - 05 - 20 1730

1.

- (a) [50%] Show that the residual e_i in the identity $Y_i = E[Y_i|X_i] + e_i$ is mean independent of X_i .

$$\text{Claim: } \mathbb{E}(e_i | X_i) = \mathbb{E}(e_i).$$

We have the identity $Y_i = \mathbb{E}[Y_i|X_i] + e_i$ with $\mathbb{E}(e_i) = 0$
 Taking conditional expectations on both sides and by construction,

by linearity, we have:

$$\mathbb{E}[Y_i|X_i] = \mathbb{E}[\mathbb{E}[Y_i|X_i]|X_i] + \mathbb{E}[e_i|X_i]$$

$$\text{By LIE, } \mathbb{E}[\mathbb{E}(\cdot)|X] = \mathbb{E}(\cdot), \text{ so}$$

$$= \mathbb{E}[Y_i|X_i] + \mathbb{E}[e_i|X_i]$$

$$\Rightarrow \mathbb{E}[e_i|X_i] = 0$$

We know by construction that $\mathbb{E}(e_i) = 0$ in the CEF decomposition identity, so

$$\mathbb{E}(e_i|X_i) = \mathbb{E}(e_i) = 0. \text{ Shown. } \square$$

- (b) [50%] Explain how measurement error causes attenuation bias in the linear regression model.

We have the linear regression model ^{pop.}

$$Y = \beta_0 + \beta_1 X + u. \quad \text{Let } X \text{ be measured}$$

with error, such that $X = X^* + e_x$, where X^*

is the measured value of X and $\mathbb{E}(e_x) = 0$,

$e_x \perp u, \beta_0, \beta_1, Y, X$.

Then The population LR of γ on X^* will recover

$$\begin{aligned}\beta_1 &= \frac{\text{Cov}(X^*, \gamma)}{\text{Var}(X^*)} \\ &= \frac{\text{Cov}(X^*, \beta_0 + \beta_1 X^* + \epsilon_x))}{\text{Var}(X^*)}\end{aligned}$$

$$= \beta_1 + \frac{\text{Cov}(X^*, \beta_1 \epsilon_x)}{\text{Var}(X^*)}$$

Recall that $X^* = X - \epsilon_x$, and substituting .

$$= \beta_1 + \frac{\text{Cov}(X - \epsilon_x, \beta_1 \epsilon_x)}{\text{Var}(X - \epsilon_x)}$$

$$= \beta_1 + - \frac{\text{Var}(\epsilon_x)}{\text{Var}(X - \epsilon_x)}$$

$$= \beta_1 - \frac{\text{Var}(\epsilon_x)}{\text{Var}(X) + \text{Var}(\epsilon_x)} \quad (\text{Cov}(X, \epsilon_x) = 0)$$

$$\Rightarrow \frac{\text{Var}(\epsilon_x)}{\text{Var}(X) + \text{Var}(\epsilon_x)} > 0 \quad \text{by definition of variance.}$$

Therefore, the estimated value of β_1 will be lower than the true value of β_1 . This shows that measurement error causes attenuation bias in the LR model of an independent variable is measured with error.

2. Consider the following AR(1) time-series model:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t$$

if both are I(1)

- (a) [50%] What econometric problems arise if $\beta = 1$? ?
- (b) [50%] Suppose we had some time-series data and estimated the above equation and obtained the following:

$$\begin{array}{l} y_t = 5.057 + 0.947 y_{t-1} \\ \quad (2.125) (0.022) \end{array}$$

where standard errors are reported in parentheses.

Can you reject that $\beta = 1$ at the 5% significance level?

- a) We get problems of inconsistent OLS estimates
and spurious regression -
- high R^2
- (elaborate) --

attenuation bias

- Because the OLS estimates may not be standard normal,
- b) We run a Dickey - Fuller test. At the 5% significance level the critical value is -2.86.

Rewrite the model as

$$y_t = \beta_0 + \gamma_0 y_{t-1} \quad \text{where } \gamma_0 = \beta_1 - 1 .$$

$$H_0 : \gamma_0 = 0$$

$$H_1 : \gamma_0 < 0 .$$

The t-statistic is

$$t = \frac{\hat{\gamma}_0 - \gamma_0}{\hat{se}(\hat{\gamma}_0)} \sim \text{Dickey-Fuller distribution}$$

Decision rule: Reject H_0 if

$$t^{\text{act}} < \underline{-2.86} . \quad (\text{one-tailed})$$

$$t^{\text{act}} = \frac{-0.053}{0.022} \approx -2.41 > \underline{-2.86}$$

\Rightarrow We cannot reject the null that $\beta = 1$ at the 5% significance level.

- (a) [20%] Test at the 5 percent significance level, the hypothesis that all other things being equal there is no difference in the average log hourly earnings of males and females.

Let's first set up the null and alternative hypotheses -

$$H_0: \beta_{\text{male}} = 0$$

$$H_1: \beta_{\text{male}} \neq 0.$$

By the CLT, we know that

$$t = \frac{\hat{\beta}_{\text{male}} - \beta_{\text{male}}}{\hat{se}(\beta_{\text{male}})} \sim N(0, 1)$$

Since we are testing at the 5% significance level,

The critical value is |1.96| and we have the decision rule:

Decision rule: reject H_0 iff $|t^{\text{act}}| > 1.96$.

Now, we calculate t^{act} .

$$t^{\text{act}} = \frac{0.0121 - 0}{0.0099} = 1.22 < 1.96. \quad H_0 \text{ not rejected.}$$

\Rightarrow We cannot reject the null hypothesis that there is no difference between men's and women's' avg log hourly wages.

- (b) [25%] Define fully the test statistic $F(21, 6937)$ reported above and interpret the result.

The F -test $F(21, 6937)$ reported above is the hypothesis test that in the OLS regression of

$$\log \text{wages} = \beta_0 + \beta_1 \text{GCSE} + \beta_2 \text{Age} + \beta_3 \text{Male} + \dots + \beta_{21},$$

$$H_0 : -\beta_1 = \dots = \beta_{21} = 0$$

$$H_1 : \exists l \text{ in } l \in \{1, 21\} \text{ s.t. } \beta_l \neq 0,$$

H_0 is rejected. That is to say, that the ^{coefficients} ~~values~~ estimated from the OLS regression are non zero.

Here the F -statistic is 55.9, so the p -value is ≈ 0 , so we can conclusively reject the null that the coefficients are 0 (no variables are relevant).

- (c) [25%] Interpret the coefficient on the variable *North*. Compute and interpret the p-value for the hypothesis that the parameter of *North* is zero.

The coefficient on the variable *North* has the following interpretation:

Residing in the North is associated with a 15% lower wage compared to residing in the West Midlands, holding all other variables in the regression constant.

Test if $\hat{\beta}_{\text{North}} = 0$:

$$H_0: \hat{\beta}_{\text{North}} = 0$$

$$H_1: \hat{\beta}_{\text{North}} \neq 0$$

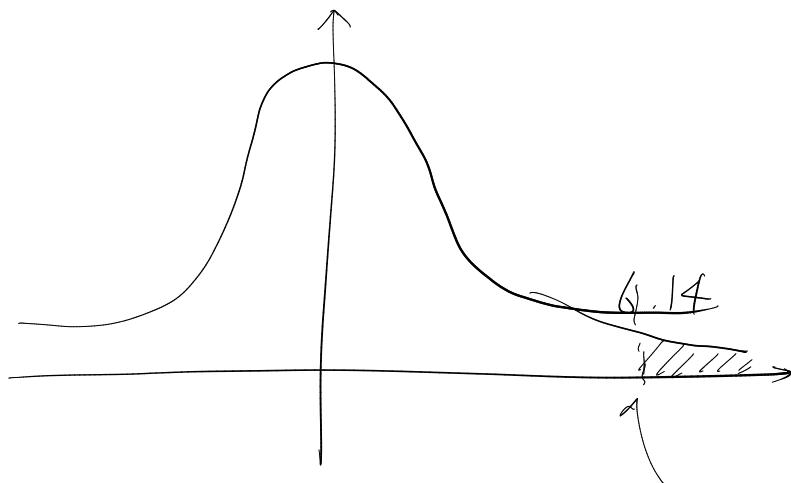
By the CLT,

$$t = \frac{\hat{\beta}_{\text{North}} - \beta_{\text{North}}}{\hat{s.e}(\hat{\beta}_{\text{North}})} \sim N(0, 1)$$

The p-value is the lowest level of significance under which the null hypothesis would be rejected, given the actual t-statistic that we observe.

We calculate t^{act} to find the p-value:

$$t^{\text{act}} = \frac{-0.1571 - 0}{0.0258} = -6.14$$



\Rightarrow the p-value $\approx 0 \rightarrow$ a first approximation .

- we can reject the null hypothesis that people living in
 (almost)

- the North have lower wages ceteris paribus at any significance level ,

than those living in West Midlands .

(d)

$$\hat{\beta}_{Inurr} = -0.084.$$

A 10% increase in the local area unemployment rate decreases hourly wages by 0.84%. The confidence

Interval is .

$$\begin{aligned} CI_{95\%} &= \left\{ \hat{\beta}_{Inurr} \pm 1.96 \hat{s.e}(\hat{\beta}_{Inurr}) \right\} \\ &= \left\{ -0.084 \pm 1.96 (0.0153) \right\} \\ &= \left\{ -0.114 , -0.054 \right\} \end{aligned}$$

95%

The confidence interval is the interval that contains the true effect of local area unemployment on an individual's hourly earnings in 95% of samples under the null hypothesis.

Normally I'd do Q4 but I'll skip it because I've done it.

5. According to a sample of 1000 men from the 2015 Labour Force Survey the mean earnings of men with university degrees 10 years after graduation was £31,500; whilst for men of the same age who did not have university degrees, the mean earnings was £29,000.

- (a) [20%] Explain why the difference in mean earnings between these groups may not reflect the causal effect of a university degree on earnings.

There may be omitted variable bias which would cause the OLS estimator to overstate the causal effect of university degree on earnings.

Suppose we had the causal model

$$\text{Earnings} = \beta_0 + \beta_1 \text{Degree} + \beta_2 \text{Ability} + u$$

and this satisfied OR $(\text{Cov}(u, \text{Degree}) = 0)$.

However, we only ran the "short" regression

$$\text{Earnings} = \gamma_0 + \gamma_1 \text{Degree} + u.$$

Then the OLS regression would consistently estimate

$$\begin{aligned}\gamma_1 &= \frac{\text{Cov}(\text{Degree}, \text{Earnings})}{\text{Var}(\text{Degree})} \\ &= \frac{\text{Cov}(\text{Degree}, \beta_0 + \beta_1 \text{Degree} + \beta_2 \text{Ability})}{\text{Var}(\text{Degree})} \\ &= \beta_1 + \beta_2 \frac{\text{Cov}(\text{Degree}, \text{Ability})}{\text{Var}(\text{Degree})}.\end{aligned}$$

If (as is highly likely) one's ability is very highly correlated with whether or not one gets a degree, then $\text{Cov}(\text{Degree}, \text{Ability}) > 0$ and you won't get the causal effect.

The Survey also records whether or not at least one of the respondent's parents went to university and graduated.

According to the survey the mean earnings for men who do not hold a degree and whose parents did not hold a degree either was £30,000; whilst for men who do not have a degree but whose parents did, earnings were £27,000 on average. Amongst men who have a degree those whose parents also had a degree earned £30,000 on average, and those whose parents did not graduate earned £33,000. There were 400 respondents who did not have a degree and whose parents did not have a degree either. The numbers in all of the other groups were 200.

- (b) [50%] Calculate the Local Average Treatment Effect (LATE) using whether or not at least one of the respondent's parents went to university and graduated as your instrumental variable.

The LATE is given by

$$\text{LATE} := \mathbb{E} \left[\beta_{\text{Degree}^i} \cdot \frac{\pi_{1i}}{\mathbb{E} \pi_{1i}} \right]$$

where π_{1i} is the coefficient on the

OLS regression

$$\frac{1}{2}$$

$$\text{Degree} = \pi_{0i} + \pi_{1i} \text{ Parents Degree}$$

$$\text{No. Degree} = 1, \text{ Parents} = 1 = 200$$

$$\text{No. Degree} = 1, \text{ Parents} = 0 = 200$$

$$" 0, " 1 = 200$$

$$" 0, " 0 = 400$$

We can calculate $\mathbb{E} \pi_{0i}$ and $\mathbb{E} \pi_{1i}$ this way:

$$\text{No. Degree} = 1 = 400$$

$$\text{No. Degree} = 0 = 600$$

$$\text{No. Degree} = 1 / \text{Parents} = 1 = 200$$

$$\text{Degree} = 1 / \text{Parents} = 0 = 200$$

$$\text{Degree} = 0.4$$

$\mathbb{E} \pi_{zi}$ must be the expected value of having a parent or getting a degree

$\frac{200}{400}$ for parents no degree $\frac{300}{400}$ for parents with degree

$$\frac{1}{3}$$

$$\frac{1}{2}$$

$$\frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

$$\mathbb{E} \pi_{zi} = \frac{1}{6}$$

$$\mathbb{E} \pi_{oi} = \text{Degree} - \mathbb{E} \pi_{zi}$$

$$= \frac{4}{10} - \frac{1}{6} \cancel{\text{less}}$$

$$= \frac{12}{30} - \frac{5}{30}$$

$$= \frac{7}{30}$$

$$\mathbb{E} \pi_{oi} = \frac{7}{30}, \quad \mathbb{E} \pi_{zi} = \frac{1}{6}.$$

Now Let's look at the numbers on the regression

$N = 200$, Degree = 1, Parents = 1, Earnings = 30000

$N = 200$, Degree = 1, Parents = 0, Earnings = 33,000

$N = 200$, " = 0, " = 1, " = 27,000

$N = 400$, " = 0, " = 0, " = 30,000

$$\text{Earnings} = \beta_0 + \beta_1 \text{Degree} + u$$

Average earnings ~~are~~

$$= \frac{30,000 + 33,000 + 27,000 + 30,000}{4}$$

5

$$= 31,000$$

Average earnings amongst those with degrees = $\frac{200 \cdot 30k + 200 + 33k}{400}$

$$= 31,500$$

Average " without degrees = $\frac{200 \cdot 27k + 400 + 30k}{600}$

$$\mathbb{E}\beta_0 = 29,000 \quad \mathbb{E}\beta_1 = 2,500 \quad = 29,000$$

$$\text{Earnings} = 29,000 + 2,500 \text{Degree}$$

$$\text{LATE} := \frac{\mathbb{E}[\beta_1 | \pi_{1i}]}{\mathbb{E}\pi_{1i}}$$

$$Y_i = \beta_{0i} + \beta_{1i} \text{Degree} + u$$

$$\begin{aligned} Y_i &= \underbrace{\beta_{0i} + \beta_{1i} (\pi_{0i} + \pi_{1i} \text{Parents Degree} + e_i)}_{\gamma_{0i} + \beta_{1i} \pi_{1i} \text{Parents Degree} + \beta_{1i} e_i} + u \\ &= \gamma_{0i} + \beta_{1i} \pi_{1i} \text{Parents Degree} + \beta_{1i} e_i + u \\ &= \gamma_{0i} + \beta_{1i} \pi_i Z_i + \varepsilon_i \end{aligned}$$

$$\mathbb{E}[Y_i | Z_i] = \mathbb{E}[\gamma_{0i} | Z_i] + \mathbb{E}[\beta_{1i} \pi_i] Z_i + \mathbb{E}[\varepsilon_i | Z_i]$$

We know that $\text{Cov}(u, Z_i) = 0$ by the fact that Z_i is an IV, and that $\text{Cov}(e_i, Z_i) = 0$ by construction. Thus,

$$= \mathbb{E}[\gamma_{0i}] + \mathbb{E}[\beta_{1i} \pi_i] Z_i + \mathbb{E}(\varepsilon_i)$$

Thus,

$$\begin{aligned} \mathbb{E}[Y_i | Z_i = 1] &= \mathbb{E}[Y_i | Z_i = 0] = \\ &\mathbb{E}[\gamma_{0i}] + \mathbb{E}[\beta_{1i} \pi_i] + \mathbb{E}(\varepsilon_i) - (\mathbb{E}[\gamma_{0i}] + 0 + \mathbb{E}(\varepsilon_i)) \end{aligned}$$

$$= \mathbb{E}[\beta_{1i} \pi_i]$$

$$\text{And therefore, the LATE} := \frac{\mathbb{E}[\beta_{1i} \pi_i]}{\mathbb{E} \pi_i} = \frac{(31k - 28.5k)}{16}$$

$$= -15k$$

- (c) [30%] Critically assess the use of this variable indicating the parents' educational attainment as a valid instrumental variable in this context.

Parents' educational attainment is a valid instrumental variable only if all the following are true:

$$\text{Relevance: } \text{Cov}(\text{Degree}, \text{Parents}) \neq 0$$

Here we have

$$\text{E} \pi_i = \frac{\text{Cov}(\text{Degree}, \text{Parents})}{\text{Var}(\text{Parents})} = \frac{1}{6} \neq 0$$

so the relevance condition is satisfied.

$$\text{Exogeneity: } \text{Cov}(\text{Parents}, u) \neq 0$$

This assumption is unlikely. We would have to assume that whether or not one's parents got a degree had no effect on one's earnings. But if for instance it was hereditary and affected both education, children's education, and earnings, then this assumption would not hold.

7.

After an empirical investigation of the relationship between economic growth, development aid and the quality of country institutions, the following set of OLS regression results obtained from a cross-section sample of 124 countries are reported:

Dependent variable: growth of GDP per capita 1990-1999		
Explanatory variables	Column 1	Column 2
Log per capita GDP 1990	-0.012 (2.37)	-0.002 (0.45)
Index of institutional quality	0.022** (3.67)	0.013** (3.06)
AID/GDP	-0.244 (1.83)	0.120 (0.71)
AID/GDP \times index of institutional quality		0.484* (2.01)
Constant	0.119 (2.62)	0.026 (0.67)
Observations	124	124
R-squared	0.15	0.39

Index of institutional quality: average of six governance indicators (ranges from -2 to 2; increasing with better quality institutions); AID/GDP is foreign aid as a proportion of GDP in 1990.

(a) [20%] The figures in parentheses are described as “t-values” and the coefficients marked with a ** are said to be “significant at the 1% level”. Explain in detail what is meant by each of these statements.

Bookwork

- (b) [20%] What do the results tell us about the relationship between aid, institutional quality and economic growth? Reviewers suggest that the researchers should re-estimate the model with institutional quality as a categorical variable. Explain how this would change the model specification and the interpretation of the relationship between the variables.

$$\begin{aligned}
 & \text{Growth} = \log \text{GDP} + \beta_1 \text{Institutional Quality} \\
 & + \beta_2 \frac{\text{AID}}{\text{GDP}} \quad \text{0.120} \\
 & + \beta_3 \frac{\text{AID}}{\text{GDP}} \cdot \text{Institutional Quality} \\
 & \quad \quad \quad \text{0.484}^{**}
 \end{aligned}$$

The results tell us the following:

1. Institutional quality has a positive correlation with GDP growth;
2. Aid has a positive (but not statistically significant) correlation with GDP growth;
3. Aid has a stronger correlation with GDP growth, the higher the institutional quality is.

Specific numerical predictions: a one unit increase in institutional quality is associated with a 0.013 point increase in growth rate.

One extra percentage point increase in $\frac{\text{Aid}}{\text{GDP}}$
is associated with a 0.484 point increase in
growth rate for each extra ^{unit} ~~point~~ in institutional
quality

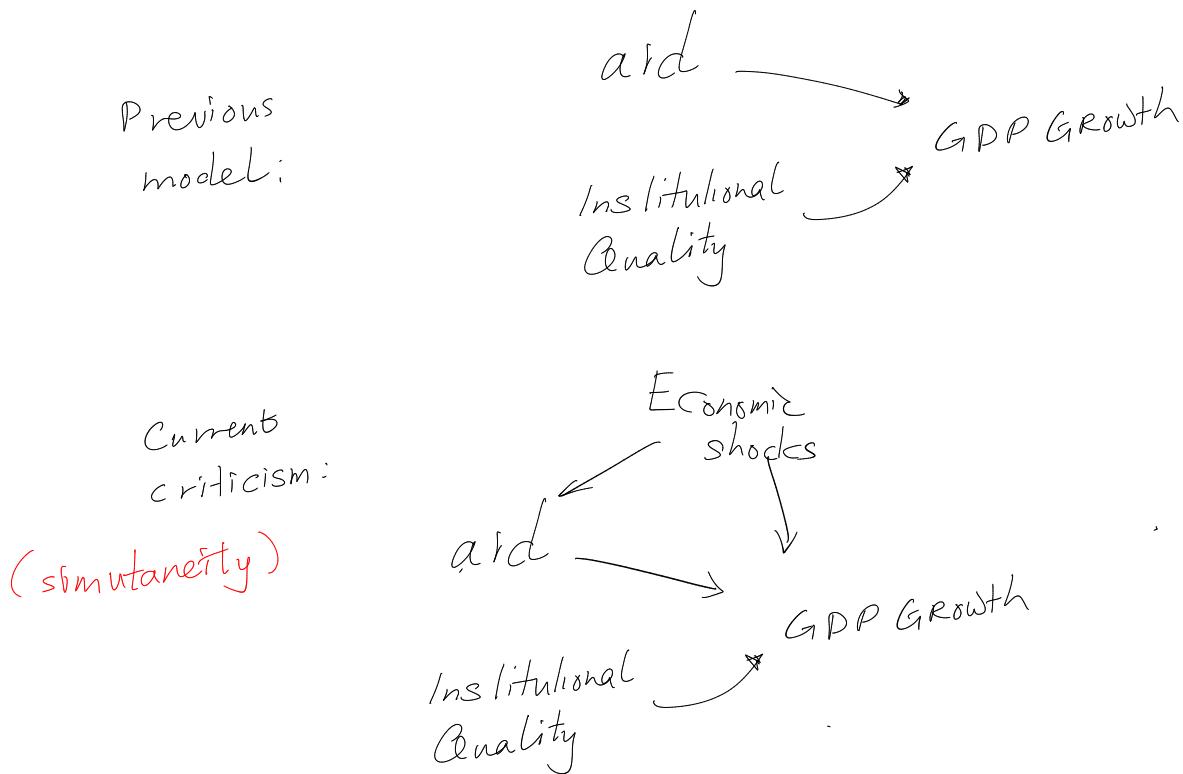
If institution quality were coded as a categorical
variable, this would mean that we would
code a dummy variable for each institutional
quality, taking care not to fall into the
dummy variable trap.

→ Most of the coeff's interpretations would remain
the same, but that of institutional quality
would be

"Being of a {high} inst. quality is associated
with an increased level of growth by X points"
compared to {low}
and the interaction term would be

"Being of a {high} inst. quality is associated
with an additional X increase in GDP growth
compared to {low} with every extra point of $\frac{\text{Aid}}{\text{GDP}}$ "

- (c) [30%] The study is criticised on the grounds that "donors respond to countries hit by unexpected negative economic shocks by increasing levels of aid and so the OLS estimates of the effects of aid are biased". Explain fully why such behaviour on the part of donors affects the validity of the OLS estimates. You can focus your discussion on the OLS estimates of column 1.



If indeed there is a hidden variable of economic shocks that affects both aid and growth, then OLR will no longer hold and the causal model won't coincide anymore with the pop LR.

To see why; consider the "short" causal model:

$$\text{Growth} = \beta_0 + \beta_1 \text{Aid} + \beta_2 \text{Quality} + u \quad (1)$$

If there is indeed a hidden variable EconShocks that affects both growth and quality, then $\text{Cov}(\text{Quality}, u) \neq 0$ because a change in the EconShocks variable (in u) affects both Growth and Quality.

- (d) [30%] Focusing on the specification used in column 1, explain how you would address this criticism, giving details of the alternative estimation procedure that you would use, its properties and any limitations.

You would simply add EconShocks as a control variable I think, that would restore the endogeneity.

If that isn't possible, use IV, some instrument that proxies for EconShocks