ECON0019 Past Paper 2022

Question A1

1. The true model is

$$score = \beta_0 + \beta_1 str + \beta_2 abil + \beta_3(str \times abil) + u$$

The mis-specified model due to inability to collect data is

$$\widehat{score} = \tilde{\beta}_0 + \tilde{\beta}_1 str + u$$

Let $y = score$ and $x = str$ to simplify notation, the OLS estimator of $\tilde{\beta}_1$ is

$$\tilde{\beta}_1 = \frac{\sum(x_i - \bar{x})y}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})(\beta_0 + \beta_1 x + \beta_2 abil + \beta_3(x \times abil) + u_i)}{\sum(x_i - \bar{x})^2}$$

$$= \beta_0 \frac{\underbrace{\sum(x_i - \bar{x})}_{=0}}{\sum(x_i - \bar{x})^2} + \beta_1 \frac{\underbrace{\sum(x_i - \bar{x})\,x}_{=\sum(x_i - \bar{x})^2}}{\sum(x_i - \bar{x})^2} + \beta_2 \frac{\sum(x_i - \bar{x})abil}{\sum(x_i - \bar{x})^2} + \beta_3 \frac{\sum(x_i - \bar{x})\,(x \times abil)}{\sum(x_i - \bar{x})^2} + \frac{\sum(x_i - \bar{x})\,u}{\sum(x_i - \bar{x})^2}$$

$$= \beta_1 + \beta_2 \frac{\sum(x_i - \bar{x})abil}{\sum(x_i - \bar{x})^2} + \beta_3 \frac{\sum(x_i - \bar{x})\,(x \times abil)}{\sum(x_i - \bar{x})^2} + \frac{\sum(x_i - \bar{x})\,u_i}{\sum(x_i - \bar{x})^2}$$

Apply LLN to each component of $\tilde{\beta}_1$, we have

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \xrightarrow{p} \mathbb{E}[(x_i - \mathbb{E}[x])^2] = \mathrm{Var}(x) > 0 \text{ by MLR. 3}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})abil \xrightarrow{p} \mathbb{E}[(X - \mathbb{E}[x])abil] = \mathbb{E}[x \cdot abil] - \mathbb{E}[x]\mathbb{E}[abil] = \mathrm{Cov}(x, abil)$$

$$\frac{1}{n}\sum(x_i - \bar{x})\,(x \times abil) \xrightarrow{p} \mathbb{E}[x \cdot x \times abil] - \mathbb{E}[x]\mathbb{E}[x\, abil] = \mathrm{Cov}(x, x \times abil)$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i \xrightarrow{p} \mathbb{E}[(x_i - \mathbb{E}[x])u_i] = \mathbb{E}[xu] - \mathbb{E}[x]\mathbb{E}[u] = 0$$

Therefore,

$$\tilde{\beta}_1 \xrightarrow{p} \beta_1 + \beta_2 \frac{\mathrm{Cov}(str, abil)}{\mathrm{Var}(str)} + \beta_3 \frac{\mathrm{Cov}(str, str \times abil)}{\mathrm{Var}(str)}$$

2. We rewrite the model

$$score = \beta_0 + \beta_1 str + (\beta_2 + \beta_3 str) ability + u$$

The reward for higher ability is

$$\frac{\partial score}{\partial abil} = \beta_2 + \beta_3 str$$

Intuitively, it's reasonable to presume that students with greater ability tends to perform better on tests, namely $\beta_2 > 0$. Also, we may also speculate that the reward for higher ability will be higher under smaller class size, which means $\beta_3 < 0$. Therefore, these sign restrictions make sense intuitively. Using the asymptotic bias formula, we may derive the asymptotic bias of $\tilde{\beta}_1$

$$\text{Bias} = \tilde{\beta}_1 - \beta_1 \xrightarrow{p} \beta_2 \underbrace{\frac{\text{Cov}(str, abil)}{\text{Var}(str)}}_{<0} + \beta_3 \underbrace{\frac{\text{Cov}(str, str \times abil)}{\text{Var}(str)}}_{<0} < 0$$

Therefore, $\tilde{\beta}_1$ will suffer from a negative asymptotic bias.

3. The first step in equation (4) suggests that *ability* and *str* are mean-independent, namely

$$\mathbb{E}[abil|str] = \mathbb{E}[abil]$$

Which means that their covariance

$$\text{Cov}(str, abil) = \mathbb{E}[str \times abil] - \mathbb{E}[str]\mathbb{E}[abil]$$

By the Law of Iterated Expectation,

$$\mathbb{E}[str \times abil] = \mathbb{E}\big[\mathbb{E}[str \times abil|str]\big] = \mathbb{E}\big[str \times \mathbb{E}[abil|str]\big] = \mathbb{E}[abil|str]\mathbb{E}[str]$$
$$= \mathbb{E}[abil]\mathbb{E}[str]$$

Back to the covariance:

$$\text{Cov}(str, abil) = \mathbb{E}[abil]\mathbb{E}[str] - \mathbb{E}[str]\mathbb{E}[abil] = 0$$

This is a strong assumption which is unlikely to hold empirically. What we observe in the real world is that students with stronger capability tends to be assigned to "preferred" classes that learn more challenging contents. These classes tend to be equipped with more teachers in addition to smaller class size compared to ordinary class. Hence, $\text{Cov}(str, abil) < 0$ is a more plausible property empirically.

4. We have derived that under (4), $\text{Cov}(str, abil) = 0$. For the covariance between str and its interaction term:

$$\text{Cov}(str, str \times abil) = \mathbb{E}[str \times str \times abil] - \mathbb{E}[str]\mathbb{E}[str \times abil]$$

Using the law of iterated expectation again

$$\mathbb{E}[str^2 \times abil] = \mathbb{E}\big[\mathbb{E}[str^2 \times abil|str]\big] = \mathbb{E}\big[str^2\mathbb{E}[abil|str]\big] = \mathbb{E}[str^2]\mathbb{E}[abil]$$

Hence, the covariance term becomes

$$\begin{aligned}\text{Cov}(str, str \times abil) &= \mathbb{E}[str^2]\mathbb{E}[abil] - \mathbb{E}[str]\mathbb{E}[str]\mathbb{E}[abil] \\ &= \mathbb{E}[abil](\mathbb{E}[str^2] - \mathbb{E}[str]^2) \\ &= \mathbb{E}[abil]\text{Var}(str)\end{aligned}$$

The probability limit of $\tilde{\beta}_1$ was

$$\begin{aligned}\tilde{\beta}_1 \xrightarrow{p} \beta_1 + \beta_2 \frac{\text{Cov}(str, abil)}{\text{Var}(str)} + \beta_3 \frac{\text{Cov}(str, str \times abil)}{\text{Var}(str)} \\ = \beta_1 + \beta_3 \cdot \frac{\mathbb{E}[abil]\text{Var}(str)}{\text{Var}(str)} \\ = \beta_1 + \beta_3\mathbb{E}[abil] \blacksquare\end{aligned}$$

This probability limit measures the average asymptotic effect of larger class size on grade over the population distribution of ability. This limit is useful when policy maker only cares about the effect of larger class size on grade on average. But it is not useful when making separate policy for low and high ability students.

5. Note that both conditions are actually requirement for *mark* to be a valid proxy for the unobservable *ability*. Condition (3) says the proxy is properly omitted from the original equation if we could include the unobservable. Condition (4) says that "*mark*" is a good proxy in a way that *str* would not help to predict the unobservable *ability* when *mark* is included as a proxy.
   They seem reasonable, because it allows correlation between *str* and *ability*.

6. Taking conditional expectation of the original model

$$
\begin{aligned}
&\mathbb{E}[score|str, mark] \\
&\qquad = \hat{\beta}_0 + \hat{\beta}_1 \mathbb{E}[str|str, mark] + \hat{\beta}_2 \mathbb{E}[abil|str, mark] \\
&\qquad + \hat{\beta}_3 \mathbb{E}[str \times abil|str, mark] + \underbrace{\mathbb{E}[u|str, mark]}_{=0}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[score|str, mark] &= \hat{\beta}_0 + \hat{\beta}_1 str + \hat{\beta}_2 \theta \cdot mark + \hat{\beta}_3 str \mathbb{E}[abil|str, mark] \\
&= \hat{\beta}_0 + \hat{\beta}_1 str + \hat{\beta}_2 \theta \cdot mark + \hat{\beta}_3 str \mathbb{E}[abil|str, mark] \\
&= \hat{\beta}_0 + \hat{\beta}_1 str + \hat{\beta}_2 \theta \cdot mark + \hat{\beta}_3 \theta \cdot str \cdot mark
\end{aligned}
$$

Recall that for the proxied equation

$$
\widehat{score} = \hat{\beta}_0 + \hat{\beta}_1 str + \hat{\beta}_2 mark + \hat{\beta}_3 (str \times mark) + u
$$

We observe that the conditional expectation of the original model aligns with the modified model, so OLS will consistently estimate coefficients.

Question A2

1. Let $\Delta\widehat{rd}$ be the change in expenditure on R&D, assuming MLR.1 – MLR.4, we have

$$
\begin{aligned}
\Delta\widehat{rd} &= 0.21 \ln \Delta Sales \\
&= 0.21 \ln 1.1 \cdot Sales - 0.21 \ln Sales \\
&= 0.21 \ln 1.1 \\
&= 0.02
\end{aligned}
$$

   Therefore, *rd* increased by 2% for a 10% in sales. This effect is quite large.

2. The null and alternative hypotheses are

$$
\mathcal{H}_0 \colon \beta_1 = 0
$$
$$
\mathcal{H}_1 \colon \beta_1 > 0
$$

   The *t*-statistics of this test is

$$
t = \frac{0.21 - 0}{0.116} = 1.81
$$

   Since $n > 30$, we may use the CDF table of a standard normal distribution instead of a *t*-distribution. The one-sided critical value for 5% and 10% level of significance are 1.645 and 1.282 respectively. Therefore, we reject the null hypothesis at both 5% and 10% level of significance.

$$
P(z \leq 1.81) = 0.9649
$$

Then, the p-value is

$$p = \Pr(z > 1.81) = 1 - \Pr(z < 1.81) = 0.0351$$

Therefore, it's reasonable to infer that rd does change with sales.

3. Given that the 5% critical value of $F(2{,}42)$ distribution is 3.22. We reject the null hypothesis at 5% level because our test statistics is greater than the critical value.

4. The heteroskedasticity robust standard error we used to calculate *t*-statistics in (b) are derived under large sample and may not be consistent under finite sample. Furthermore, the critical values we used for both tests are large-sample approximation, which may be understated than the true critical value given that our sample size is small.

   Despite we may use exact finite sample distribution along with normality assumption for inference, yet normality assumption is awfully violated because our dependent variable is strictly positive. Therefore, we should not trust the test results.

5. The marginal effect of *sales* on *rd* is

$$\frac{\partial \widehat{rd}}{\partial sales} = 0.030 - 0.000014 sales$$

   The effect changes sign when *sales* is evaluated at the point where the partial derivative equals to naught.

$$0.03 = 0.000014 sales$$

$$sales = 2142.86$$

   Therefore, expenditure on R&D began decreasing when annual sales exceeds 2143 million GBP.

6. We construct the following composite model

$$\widehat{rd} = \beta_0 + \beta_1 sales - \beta_2 sales^2 + \beta_3 profit + \beta_4 \log(sales) + u$$

   In order to determine which model is preferred, we test two null hypotheses

$$\mathcal{H}_0\colon \beta_1 = \beta_2 = 0$$

$$\mathcal{H}_1\colon \beta_4 = 0$$

using F-test and t-test respectively. We are able to determine which model is preferred if one of the null hypotheses is rejected while the other is failed to be rejected. However, this method will not work if we end up with reject-reject or accept-accept.

7. Model must be selected in an ad hoc manner. First-differencing estimator enjoys smaller bias when companies suffer from heterogeneity issue in the form of individual-specific fixed effects, which will be removed by first-differencing. However, first-differencing estimator requires strict exogeneity assumption to be consistent. Nevertheless, pooled regression may yield inference of higher power since we can utilize all 90 observations rather than a half of that using fixed effect estimator.
   We may use Breusch-Pagan LM Test to test for heteroskedasticity, we avoid using pooled OLS if we reject the null hypothesis of homoskedasticity.

Question B.1

1. Under classical linear regression assumptions, OLS estimate of β is inconsistent whenever the exogeneity assumption fails. One potential reason contributing to the failure of exogeneity could be omitted variable bias, where a variable that affects the number of riots while correlates with income per capita is omitted from the model. For example, there could be district-specific effects not accounted by the district dummy. On the other hand, a reverse causality may hold which violates the exogeneity assumption. For example, a district with greater number of riots tends to suffer from more destruction to infrastructure and productivity, therefore rending as lower income per capita.

2. We will use two-stage least square (2SLS) method to estimate β. First, we regress the endogenous regressor Y_i on all instruments R_{1i}, R_{2i} and all other exogenous variables $\sum_{i=1}^{28} \mathbf{1}\{S_i = s\}$, obtain the OLS fitted value $\hat{Y}_i$. Then, the second-stage regression is

$$C_i = \beta Y_i + \sum_{i=1}^{28} \gamma_s \mathbf{1}\{S_i = s\} + \epsilon_i$$

The 2SLS estimate $\hat{\beta}_1$ is consistent when the relevance and exogeneity assumption of instruments are satisfied:

$$\text{Cov}(R_{1i}, Y_i) \neq 0, \text{Cov}(R_{2i}, Y_1) \neq 0$$

$$\text{Cov}(R_{1i}, \epsilon_i) = 0, \text{Cov}(R_{2i}, \epsilon_i) = 0$$

Since we have two instruments for a single endogenous regressor, we may perform the overidentification test to test for the exogeneity of instruments. We first estimate coefficients by 2SLS and calculate residual $\widehat{u_i}$ using actual observed data rather than estimate from the 2SLS second-stage regression. Then, regress $\hat{u}_i$ on all exogenous variables and record R^2. Under the null hypothesis that IVs are exogenous, the test statistics follow a chi-squared distribution

$$nR^2 \sim \chi_1^2$$

The IVs are valid when R^2 is small and we failed to reject $\mathcal{H}_0$.

We can also test for the relevance condition by conducting F-test on the 2SLS first-stage coefficients.

3. The typical level of rainfall in a district could be the result of persistent district-specific variables that are not included from the model. Therefore, exogeneity assumption of IV is violated and de-meaning ensures only the random, with-in district shocks serve as instruments, the 2SLS estimators are more likely to be consistent as a result.

4. The first stage 2SLS regression is

$$Y_i = \pi_0 + \pi_{1i} R_{1i} + V_i$$

Note that π_{1i} now varies with regions, indicating heterogenous effect. The second-stage 2SLS regression coefficient now identifies Local Average Treatment Effect (LATE), which is the effect of income on conflict averaged by how much Y_i is affected by R_{1i}. The LATE estimator is

$$\beta_{LATE} = \frac{\mathbb{E}[\beta_{1i}\pi_{1i}]}{\mathbb{E}[\pi_{1i}]} = \mathbb{E}\left[\beta_{1i} \frac{\pi_{1i}}{\mathbb{E}[\pi_{1i}]}\right]$$

LATE estimator requires monotonicity assumption, $\pi_{1i} > 0$ or $\pi_{1i} < 0$ for all i, meaning that the direction of impact of rainfall on income must be the same across all districts. Intuitively, it's reasonable to assume that $\pi_{1i} > 0$ as more rainfalls mean better crop yield and higher income. Besides, LATE also requires independence assumption to hold true, meaning that R_{1i} must be randomly assigned or as-if randomly assigned. This assumption seem plausible because the demeaned rainfall is likely to be a random shock that uncorrelated with the regressor or error term.

5. Sarsons' finding states that the rainfall is not a relevant instrument to income in regions with a dam, which makes sense intuitively because the rainfall instrument dictates a causality chain, namely Rainfall → Crop yield → Income. However, a dam tears down this chain because crop yield becomes independent of rainfall. Hence, the rainfall in regions with a dam is significantly less correlated with income than regions without a dam and $\text{Cov}(R_{i1}, Y_i) \to 0$ for regions with a dam, making this instrument irrelevant.

Question B.2

1. We will use a Tobit model to model this censoring problem. The latent process is

$$y^\star = \ln y_i = \beta_0 + x_i'\beta_x + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Define the outcome variable of interest as

$$y_i = \begin{cases} \exp(y^\star) \text{ if } y_i^\star \leq \ln(c_i) \\ c_i \text{ if } y_i^\star > \ln(c_i) \end{cases}$$

The probability of being censored is

$$\begin{aligned}
\Pr(y_i^\star > \ln c_i) &= \Pr(\beta_0 + x_i'\beta_x + \epsilon_i > \ln c_i) \\
&= \Pr\left(\frac{\epsilon_i}{\sigma} > \frac{\ln c_i - \beta_0 + x_i'\beta_x}{\sigma}\right) \\
&= 1 - \Pr\left(\frac{\epsilon_i}{\sigma} < \frac{\ln c_i - \beta_0 + x_i'\beta_x}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{\ln c_i - \beta_0 + x_i'\beta_x}{\sigma}\right)
\end{aligned}$$

The density for non-censored data is

$$f(y_i^\star | x_i, y_i > c_i) = \phi\left(\frac{\ln c_i - \beta_0 + x_i'\beta_x}{\sigma}\right)$$

The log-likelihood function is

$$\ell(\beta_0, \beta, \sigma) = \sum_{i=1}^{n} \mathbf{1}\{c_i \geq \boldsymbol{y_i}\} \ln \phi\left(\frac{\ln c_i - \beta_0 + x_i' \beta_x}{\sigma}\right)$$
$$+ \mathbf{1}\{c_i < y_i\} \ln\left(1 - \Phi\left(\frac{\ln c_i - \beta_0 + x_i' \beta_x}{\sigma}\right)\right)$$

We estimate MLE by finding the set of parameters that maximize the log-likelihood function.

2. We would run Heckman correction for this selection problem. For both scenarios, OLS will remain consistent because when the selection depends on γ_d only through a constant or through variables that are independent of d_i, the Inverse Mills ratio induces only a shift in intercept.

3. We also run Heckman correction to restore consistency due to selection problem. First, we use MLE to estimate the selection equation as a probit model, then we evaluate the inverse Mills' ratio $\hat{\lambda}(\hat{\gamma}_0 + \hat{\gamma}_d d_i + \hat{\gamma}_x' \mathrm{x}_i)$ for observations with $h_i = 1$ using MLE estimates. Second, we run the equation (11) using the selected sample only with $\hat{\lambda}$ as an additional regressor. This method estimates β consistently.

4. We perform a test of autocorrelation. We first regress q_t on all week dummies for all $t = 1, \dots, n$, then obtain the OLS residual $\hat{u}_1, \dots, \hat{u}_n$. Regress $\hat{u}_t$ on $\hat{u}_{t-1}$ for $t = 2, \dots, n$ and obtain an estimate of the coefficient of the lagged residual $\hat{\rho}$. We perform a *t*-test to the null hypothesis of no autocorrelation $\mathcal{H}_0: \hat{\rho} = 0$ against an one-sided alternative hypothesis as the direction of autocorrelation must have been intuitively determined in priori. This test must be conducted under the assumption that all week dummies are strictly exogenous.

5. OLS would be consistent when a set of classical time series assumptions are satisfied, namely no perfect collinearity, exogeneity (strict exogeneity is not required), weak dependence and second-order stationarity. Autocorrelation does not compromise consistency if the above assumptions are satisfied.