ECON0019 Past Paper 2020

Question A.1

1. Pooled OLS regression (1) contains an individual-specific fixed effect a_i, which causes endogeneity issue, leading to biased and inconsistent estimator of β. By using the first-differencing (fixed effect) estimator, we get rid of the unobserved individual-specific effect and therefore address the heterogeneity issue, so that we will have consistent OLS estimate.

$$\begin{aligned}\Delta y_{it} &= y_{it} - y_{i,t-1} \\ &= \beta_0 + \beta_1 x_{it} + a_i + u_{it} - \left(\beta_0 + \beta_1 x_{i,t-1} + a_i - u_{i,t-1}\right) \\ &= \beta_1\left(x_{it} - x_{i,t-1}\right) + \left(u_{it} - u_{i,t-1}\right) \\ &= \beta_1 \Delta x_{it} + \Delta u_{it}\end{aligned}$$

2. The sum of squared residual is defined as

$$SSR = \sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta y_{it} - b_1 \Delta x_{it})^2$$

We minimize SSR by differentiating with respect to b_1 and solve for the first order condition

$$\frac{\partial SSR}{\partial b_1} = -2\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}(\Delta y_{it} - b_1 \Delta x_{it}) = 0$$

$$\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta y_{it} = \hat{\beta}_1 \sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta y_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}$$

3. Rearranging the formula, we have $e_{it} = u_{it} - u_{i,t-1} = \Delta u_{it}$, so

$$\mathbb{E}[\Delta u_{it} | X_1, \dots, X_n] = \mathbb{E}[e_{it} | X_1, \dots, X_n]$$

Given that we adopt random sampling technique, e_{it} shall only depend on X_{it} and $X_{it} = \Delta x_{i2}, \dots, \Delta x_{it}$ is a deterministic function of $(x_{i1}, \dots, x_{iT})$, we have

$$\begin{aligned}\mathbb{E}[\Delta u_{it} | X_1, \dots, X_n] &= \mathbb{E}[e_{it} | X_{it}] \\ &= \mathbb{E}[e_{it} | \Delta x_{i2}, .., \Delta x_{iT}]\end{aligned}$$

$$\mathbb{E}[\Delta u_{it}|X_1,\dots,X_n] = \mathbb{E}[e_{it}|x_{i1},\dots,x_{iT}] = 0$$

We rewrite the OLS estimator into actual value + sampling error

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}(\beta_1 \Delta x_{it} + \Delta u_{it})}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2} = \beta_1 + \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}$$

The OLS estimator is unbiased when its conditional expectation equals to its true value

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_1 \middle| X_1,\dots,X_n\right] &= \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}\middle| X_1,\dots,X_n\right] \\
&= \beta_1 + \frac{1}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}\,|X_1,\dots,X_n\right] \\
&= \beta_1 + \frac{1}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\,\underbrace{\mathbb{E}[\Delta u_{it}|X_1,\dots,X_n]}_{=0} \\
&= \beta_1
\end{aligned}$$

4. For $s = t$, recall that $e_{it} = u_{it} - u_{i,t-1} = \Delta u_{it}$. We use LIE in our derivation

$$\begin{aligned}
\text{Cov}(\Delta u_{is}, \Delta u_{is}|X_1,\dots,X_n) &= \mathbb{E}[e_{is}^2|X_1,\dots,X_n] - \mathbb{E}[e_{is}|X_1,\dots,X_n]^2 \\
&= \mathbb{E}[e_{is}^2|X_i] - \underbrace{\mathbb{E}[\mathbb{E}[e_{is}|x_{i1},\dots,x_{iT}]|X_i]^2}_{=0} \\
&= \mathbb{E}\left[\mathbb{E}[e_{is}^2|x_{i1},\dots,x_{iT}]\middle|X_i\right] \\
&= \sigma^2
\end{aligned}$$

For $s \neq t$

$$\begin{aligned}
\text{Cov}(\Delta u_{is}, \Delta u_{it}|X_1,\dots,X_n) &= \mathbb{E}[e_{is}e_{it}|X_1,\dots,X_n] - \mathbb{E}[e_{is}|X_1,\dots,X_n]\mathbb{E}[e_{it}|X_1,\dots,X_n] \\
&= \mathbb{E}[e_{is}e_{it}|X_i] - \mathbb{E}[e_{is}|X_i]\mathbb{E}[e_{it}|X_i] \\
&= \mathbb{E}\left[\underbrace{\mathbb{E}[e_{is}e_{it}|x_{i1},\dots,x_{it}]}_{=0}\middle|X_i\right] \\
&\quad - \mathbb{E}[e_{is}|X_i]\mathbb{E}\left[\underbrace{\mathbb{E}[e_{it}|x_{i1},\dots,x_{iT}]}_{=0}\middle|X_i\right] \\
&= 0
\end{aligned}$$

The conditional variance of the first-differencing estimator is

$$\text{Var}\left(\hat{\beta}_1 \middle| X_1,\dots,X_n\right) = \text{Var}\left(\beta_1 + \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}\middle| X_1,\dots,X_n\right)$$

$$= \text{Var}\left(\frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}\middle| X_i\right)$$

$$= \frac{1}{[\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2]^2}\text{Var}(\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}\,|X_i)$$

Apply the variance formula of weighted sum

$$\text{Var}(\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}\,|X_i) = \sum_{i,t}\sum_{j,s}\Delta x_{it}\Delta x_{js}\text{Cov}\big(\Delta u_{it}, \Delta u_{js}\big|X_1, \dots, X_n\big)$$

For every cross-individual term $i \neq j$, $\text{Cov}\big(\Delta u_{it}, \Delta u_{js}\big|X_1, \dots, X_n\big) = 0$ due to random sampling. For the same individual across time, the covariance structure we derived earlier implies

$$\text{Cov}(\Delta u_{is}, \Delta u_{it}|X_1, \dots, X_n) = \begin{cases}\sigma^2 \text{ if } s = t\\ 0 \text{ if } s \neq t\end{cases}$$

Therefore, only the diagonal elements of the weighted sum where $s = t$ are not zero

$$\sum_{i,t}\sum_{j,s}\Delta x_{it}\Delta x_{js}\text{Cov}\big(\Delta u_{it}, \Delta u_{js}\big|X_1, \dots, X_n\big) = \sum_{i,t}(\Delta x_{it})^2\sigma^2$$

Plug back to the conditional variance expression

$$\text{Var}\big(\hat{\beta}_1\big|X_1, \dots, X_n\big) = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}{[\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2]^2}\sigma^2$$

$$= \frac{1}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}\sigma^2 \blacksquare$$

5. Recall that the true parameter + sampling error formula was

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2}$$

Applying the law of large number (LLN) to both the denominator and numerator of the sampling error term

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta x_{it}\Delta u_{it} \xrightarrow{p} \sum_{t=2}^{T}\mathbb{E}[\Delta x_{it}\Delta u_{it}] = \sum_{t=2}^{T}\mathbb{E}\big[\mathbb{E}[\Delta x_{it}\Delta u_{it}|X_i]\big]$$

$$= \sum_{t=2}^{T}\mathbb{E}\left[\Delta x_{it}\underbrace{\mathbb{E}[\Delta u_{it}|x_{i1}, \dots, x_{it}]}_{=0}\right] = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta x_{it})^2 \xrightarrow{p} \sum_{t=2}^{T}\mathbb{E}[(\Delta x_{it})^2] > 0$$

Therefore, the first-differencing estimator is consistent

$$\hat{\beta}_1 \xrightarrow{p} \beta_1$$

Question A.2

1. The *i.i.d.* assumption allows us to conveniently calculate standard error of OLS parameters by making cross-term vanishes. Country-level data are likely to be independently and identically distributed, because the population of each country is considered to be a sample when the population is every country in the world, each country (sample) should be identically distributed and independent of each other if the sampling process is random. However, 104 countries are almost half of the population and are certainly not randomly selected because not all of then publishes reliable data, causing a violation of independent part of *i.i.d.* assumption.
   $\mathbb{E}[u|gpop, sk, educ] = 0$ ensures that OLS estimators are unbiased and consistent. This assumption is certainly not satisfied because there are numerous factors that correlate with *educ* and affect *relinc*, such as the average IQ and institutional features of the nation.

2. A 1 unit increase in the population growth rate will result in 5.869 less monetary unit of GDP per worker relative to the US. A one unit increase in average investment share of GDP will result in 0.738 higher monetary unit of GDP per work relative to the US. One more years of avg. education attainment will increase relative GDP per worker by 0.055. Intuitively, all coefficients have the expected sign.
   The robust *t*-statistics are
   $$t_{gpop} = -2.622, \qquad t_{sk} = 2.510, \qquad t_{educ} = 5.5$$
   All three coefficients are statistically significant at 5% level of significance, and average investment share is significant at 1% significance level.

3. For OECD countries:
   $$\widehat{relinc} = \beta_0 + \beta_1 gpop + \beta_2 sk + \beta_3 educ + \beta_4 oecd + \beta_5 (oecd \times gpop) + \beta_6 (oecd \times sk) + \beta_7 (oecd \times educ)$$
   $$\widehat{relinc} = (-0.068 + 0.381) + (-0.063 - 8.038) gpop + (0.719 - 0.43) sk + (0.044 + 0.003) educ$$

$$\widehat{relinc} = 0.313 - 8.101gpop + 0.289sk + 0.047educ$$

For non-OECD nations:

$$\widehat{relinc} = \beta_0 + \beta_1 gpop + \beta_2 sk + \beta_3 educ + \beta_4 oecd$$

$$\widehat{relinc} = -0.068 - 0.063gpop + 0.719sk + 0.044educ$$

Investment has greater impact on *relinc* of non-OECD nations while *gpop* is more negatively correlated with *relinc* for OECD nations. The effect of education of *relinc* is similar for both OECD and non-OECD nations.

4. We wish to test whether $\beta_4, \ldots, \beta_7$ are jointly equal to naught

$$\mathcal{H}_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

$$\mathcal{H}_1: \text{At least one coefficient is non} - \text{zero}$$

The 5% critical value of $F_{104,4}$ is approximately 2.46. Therefore, we reject the null hypothesis, meaning that the effect of explanatory variables on the outcome variables varies based on whether the country belongs to OECD.

5. Given that the t-statistics follow a standard normal distribution in large sample, the *p*-value for OECD is

$$p = 2(1 - 0.9992) = 0.0016$$

Therefore, we reject the null hypothesis that *oecd* term is equal to naught at all conventional level of significance, meaning that OECD countries should have greater *relinc* than non-OECD countries, ceteris paribus.

6. The null hypothesis is

$$\mathcal{H}_0: \beta_5 = \beta_6 = \beta_7 = 0$$

We failed to reject the null hypothesis at 5% level of significance, meaning that being an OECD country does not provide you with different effect of *gpop, sk* and *educ* on *relinc.* In conclusion, the property of being an OECD country itself correlates with greater prosperity while not changing any underlying partial effect.

Question B.1

1. Assuming U follows a logistic distribution with $\mu = 0$ and $s = 1$. For the binary dependent variable, the conditional probability of success is

$$\begin{aligned}\Pr(D_i = 1|R^p) &= \Pr(\beta_0 + \beta_1 R^p + U \geq 0) \\ &= \Pr(U \geq -\beta_0 - \beta_1 R^p) \\ &= 1 - \Pr(U < -\beta_0 - \beta_1 R^p) \\ &= 1 - \Lambda(-\beta_0 - \beta_1 R^p)\end{aligned}$$

The probability of failure is

$$\begin{aligned}\Pr(D_i = 1|R^p) &= \Pr(\beta_0 + \beta_1 R^p + U < 0) \\ &= \Pr(U < -\beta_0 - \beta_1 R^p) \\ &= \Lambda(-\beta_0 - \beta_1 R^p)\end{aligned}$$

The likelihood function is

$$\prod_{i=1}^{n} (1 - \Lambda(-\beta_0 - \beta_1 R^p))^{D_i} \times \Lambda(-\beta_0 - \beta_1 R^p)^{1-D_i}$$

Which corresponds to log-likelihood function

$$\sum_{i=1}^{n} D_i \ln[1 - \Lambda(-\beta_0 - \beta_1 R^p)] + (1 - D_i) \ln[\Lambda(-\beta_0 - \beta_1 R^p)]$$

Let $x = -\beta_0 - \beta_1 R^p$, we rearrange the log-likelihood function

$$\begin{aligned}\ell &= \sum_{i=1}^{n} D_i \ln\left[1 - \frac{e^x}{1 + e^x}\right] + (1 - D_i) \ln\left[\frac{e^x}{1 + e^x}\right] \\ &= \sum_{i=1}^{n} D_i \ln\left[\frac{1}{1 + e^x}\right] + (1 - D_i)[x - \ln(1 + e^x)] \\ &= \sum_{i=1}^{n} -D_i \ln(1 + e^x) + (1 - D_i)[x - \ln(1 + e^x)] \\ &= \sum_{i=1}^{n} (1 - D_i)\left[-\beta_0 - \beta_1 R^p - \ln\left(1 - e^{-\beta_0 - \beta_1 R^p}\right)\right] - D_i \ln\left(1 + e^{-\beta_0 - \beta_1 R^p}\right)\end{aligned}$$

We use the minimum likelihood estimation (MLE) to consistently estimate the coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\hat{\beta} = \arg\max_{\beta} \ell(\beta)$$

The difference in probability of success is

$$\Pr(D_i = 1|R^p = 10) - \Pr(D_i = 1|R^p = 0) = \left[1 - \Lambda\left(-\hat{\beta}_0 - \hat{\beta}_1 \cdot 10\right)\right] - \left[1 - \Lambda\left(-\widehat{\beta_0}\right)\right]$$

$$= \left[1 - \frac{e^{-\widehat{\beta}_0 - \widehat{\beta}_1 \cdot 10}}{1 + e^{-\widehat{\beta}_0 - \widehat{\beta}_1 \cdot 10}}\right] - \left[1 - \frac{e^{-\widehat{\beta}_0}}{1 + e^{-\widehat{\beta}_0}}\right]$$

$$= \frac{e^{-\widehat{\beta}_0}}{1 + e^{-\widehat{\beta}_0}} - \frac{e^{-\widehat{\beta}_0 - \widehat{\beta}_1 \cdot 10}}{1 + e^{-\widehat{\beta}_0 - \widehat{\beta}_1 \cdot 10}}$$

2. To address the endogeneity issue caused by reverse causality, we use 2SLS estimation protocol. We run the 2SLS first-stage regression by OLS

$$R_i^p = \pi_0 + \pi_1 Z_i + u_i$$

And save the first OLS estimates of the endogenous variable $\hat{R}_i^C$. The 2SLS second stage is

$$R_i^C = \alpha_0 + \alpha_1 \hat{R}_i^C + V_i$$

Which gives us a consistent estimate of $\hat{\alpha}$ if the exogeneity and relevance assumptions of instrument are satisfied. The validity assumption requires that $\mathrm{Cov}(Z_i, V_i) = 0$, meaning that any variable that affects children's risk attitude should remain uncorrelated with parent's religion. However, the validity assumption may not hold because there may exist community-level confounder that affects parent's religious view while simultaneously impact children's religious view and therefore their attitudes toward risk.

3. Using the rule of thumb (Stock & Goyo, 2005) that an instrument is significant when the first-stage *F*-statistics is greater than 10, both instruments are weak, and the relevance condition fails to hold. Therefore, the 2SLS estimator will be imprecise and biased.

4. For 2SLS protocol with one endogenous regressor and one instrument, the variance of 2SLS estimator under homoskedasticity is

$$\widehat{var}\left(\hat{\beta}_{2\mathrm{SLS}}\right) = \frac{\sigma^2}{SST_x R_{x,z}^2}$$

Since R-squared is defined over the range between zero and one, the variance of 2SLS estimator is strictly greater than that of OLS estimator unless $\mathrm{Cov}(x, z) = 0$, which corresponds the case of an irrelevant instrument.

5. This autoregressive model cannot be unbiased since it would never satisfy the strict exogeneity assumption because there is a lagged value of outcome variable as regressor, meaning that current outcome variable would depends on all past error terms.