計量經濟學導論暨實習期末考參考解答

2021.6.23.

1 (c)

2 (d)

3 (b)

4 (a)

5 (b)

6 (d)

7 (b)

8 (d)

9 (d)

10 (b)

11 (a)

12 (b)

13 (b)

14 (b)

15 (b)

二、非選擇題:

1a β_2 is not identified since there are no instruments available for equation (1) (i.e. there are no omitted exogenous regressors in (1) which are present in (2).

γ_2 is identified because we can use Z as an instrument in equation (2), because it is independent of v (exogenous) and is correlated with Y (relevant) as long as $\beta_3 \neq 0$.

1b Solving the system yields

$$Y = \frac{1}{1-\beta_2\gamma_2}(\beta_1 + \beta_2\gamma_1 + \beta_3 Z + u + \beta_2 v)$$

and thus

$$\text{Cov}(Y, v) = \frac{\beta_2}{1-\beta_2\gamma_2}\sigma_v^2$$

since u, v, and Z are uncorrelated.

1c We have

$$\hat{\gamma}_2^{OLS} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})X_i}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

and substituting (2) yields

$$\begin{aligned} \hat{\gamma}_2^{OLS} &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})(\gamma_1 + \gamma_2 Y_i + v_i)}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= \gamma_2 + \frac{\sum_{i=1}^n (Y_i - \bar{Y})v_i}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \end{aligned}$$

so that

$$\begin{aligned} \hat{\gamma}_2^{OLS} &\xrightarrow{p} \gamma_2 + \frac{\text{Cov}(Y, v)}{\text{Var}(Y)} \\ &= \gamma_2 + \frac{\beta_2}{1 - \beta_2\gamma_2}\frac{\sigma_v^2}{\sigma_Y^2} \neq \gamma_2 \end{aligned}$$

$\hat{\gamma}_2^{OLS}$ is in general not consistent. However, if $\beta_2 = 0$, then Y is not influenced by X, there is no simultaneity and $\hat{\gamma}_2^{OLS}$ will be consistent.

1d γ_2 is identified because the exogenous variable Z is omitted in equation (2), so it can be used as instrument as it is independent of v. Since it is present in equation (1) it is correlated with Y (relevant) if $\beta_3 \neq 0$.

The IV estimator is

$$\hat{\gamma}_2^{IV} = \frac{\sum_{i=1}^n (Z_i - \bar{Z})(X_i - \bar{X})}{\sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})} = \frac{\sum_{i=1}^n (Z_i - \bar{Z})X_i}{\sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})},$$

and substituting equation (2) yields

$$\begin{aligned} \hat{\gamma}_2^{IV} &= \frac{\sum_{i=1}^n (Z_i - \bar{Z})(\gamma_1 + \gamma_2 Y_i + v_i)}{\sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})} \\ &= \gamma_2 + \frac{\sum_{i=1}^n (Z_i - \bar{Z})v_i}{\sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})} \\ &\xrightarrow{p} \gamma_2 + \frac{\text{Cov}(Z, v)}{\text{Cov}(Z, Y)} = \gamma_2 \end{aligned}$$

because $\text{Cov}(Z, v) = 0$ and $\text{Cov}(Z, Y) \neq 0$ if $\beta_3 \neq 0$ since

$$\text{Cov}(Z, Y) = \frac{\beta_3}{1 - \beta_2\gamma_2}\text{Var}(Z),$$

and thus the IV estimator $\hat{\gamma}_2^{IV}$ is consistent.

1e The 2SLS estimator is

$$\hat{\gamma}_2^{2SLS} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(X_i - \bar{X})}{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2} \equiv \frac{s_{\hat{Y}X}}{s_{\hat{Y}}^2},$$

where $s_{\hat{Y}X}$ is the sample covariance between $\hat{Y}$ and X in the second stage. Because $\hat{Y}_i = \hat{\delta}_1 + \hat{\delta}_2 Z_i$,

$$\begin{aligned} \hat{\gamma}_2^{2SLS} &= \frac{s_{\hat{Y}X}}{s_{\hat{Y}}^2} \\ &= \frac{\hat{\delta}_2 s_{ZX}}{\hat{\delta}_2^2 s_Z^2} = \frac{s_{ZX}}{\hat{\delta}_2 s_Z^2} \\ &= \frac{s_{XZ}}{s_{YZ}} \quad \text{since } \hat{\delta}_2 = \frac{s_{YZ}}{s_Z^2} \\ &= \hat{\gamma}_2^{IV} \end{aligned}$$

Hence, $\hat{\gamma}_2^{2SLS}$ is equivalent to $\hat{\gamma}_2^{IV}$, and therefore consistent.

1f If $\gamma_1 = 0$, then we have

$$\frac{\bar{X}}{\bar{Y}} = \frac{\gamma_2 \bar{Y} + \bar{v}}{\bar{Y}} = \gamma_2 + \frac{\bar{v}}{\bar{Y}} \xrightarrow{p} \gamma_2$$

because

$$\begin{aligned} \bar{v} &\xrightarrow{p} \mathrm{E}(v) = 0 \\ \bar{Y} &\xrightarrow{p} \mathrm{E}(Y) = \frac{1}{1-\beta_2\gamma_2}\mathrm{E}(\beta_1 + \gamma_1\beta_2 + \beta_3 Z + u + \beta_2 v) \neq 0 \end{aligned}$$

Therefore, $\frac{\bar{X}}{\bar{Y}}$ is a consistent estimator of γ_2.

2a

$$\begin{aligned} \mathrm{E}(Y_t) &= \mathrm{E}\left(\sum_{i=1}^{t} u_i\right) = \sum_{i=1}^{t} \mathrm{E}(u_i) = 0. \\ \mathrm{Var}(Y_t) &= \mathrm{Var}\left(\sum_{i=1}^{t} u_i\right) = \sum_{i=1}^{t} \mathrm{Var}(u_i) = t\sigma_u^2, \end{aligned}$$

since $\mathrm{Cov}(u_i, u_j) = 0$ for $i \neq j$.

2b $Y_i = \sum_{i=1}^{t} u_i$ and $Y_{t-k} = \sum_{i=1}^{t-k} u_i$, so that $\mathrm{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma_u^2$.

2c From (a) the variance of Y_t depends on t, from (b) the covariance between Y_t and Y_{t-k} also depends on t, so Y_t is nonstationary.

3a For the differences estimator, the regression model is

$$Y_{i2} = \beta_0 + \beta_1 X_{i2} + (\alpha_i + u_{i2}),$$

the variance of the error term is

$$\mathrm{Var}(\alpha_i + u_{i2}) = \mathrm{Var}(\alpha_i) + \mathrm{Var}(u_{i2}) + 2\mathrm{Cov}(\alpha_i, u_{i2}) = \sigma_\alpha^2 + \sigma_u^2.$$

Therefore, the variance of the differences estimator is

$$\mathrm{Var}(\hat{\beta}_1^{differences}) = \frac{\mathrm{Var}(\alpha_i + u_{i2})}{n\mathrm{Var}(X_{i2})} = \frac{\sigma_\alpha^2 + \sigma_u^2}{n\mathrm{Var}(X_{i2})}.$$

3b For the differences-in-differences model, the regression model is

$$\begin{aligned} \Delta Y_i = Y_{i2} - Y_{i1} = \beta_0 + \beta_1 \Delta X_i + \Delta u_i &= \beta_0 + \beta_1(X_{i2} - X_{i1}) + (u_{i2} - u_{i1}) \\ &= \beta_0 + \beta_1 X_{i2} + (u_{i2} - u_{i1}), \text{ since } X_{i1} = 0. \end{aligned}$$

Thus, the variance of the error term is

$$\mathrm{Var}(u_{i2} - u_{i1}) = \mathrm{Var}(u_{i2}) + \mathrm{Var}(u_{i1}) - 2\mathrm{Cov}(u_{i2}, u_{i1}) = 2\sigma_u^2$$

Therefore, the variance of the differences-in-differences estimator is

$$\mathrm{Var}(\hat{\beta}_1^{diffs-in-diffs}) = \frac{\mathrm{Var}(u_{i2} - u_{i1})}{n\mathrm{Var}(X_{i2})} = \frac{2\sigma_u^2}{n\mathrm{Var}(X_{i2})}$$

3c From the results in (a) and (b), when $\sigma_\alpha^2 > \sigma_u^2$, $\mathrm{Var}(\hat{\beta}_1^{differences}) > \mathrm{Var}(\hat{\beta}_1^{diffs-in-diffs})$, then differences-in-differences estimator is more efficient then the differences estimator. Thus, if there is considerable large variance in the individual-specific effects, α_i, it is better to use the differences-in-differences estimator.

4a From model (1) in the STATA log file, the call-back rate for whites is the estimated intercept, 0.097, and the call-back for blacks is 0.097 − 0.032 = 0.065. The difference is -0.032, which is statistically significant at the 1% level (t-statistic = -4.11). This number implies that 9.7% of resumes with white-sounding names generated a call back. Only 6.5% of resumes with black-sounding names generated a call back. Since the average call-back rate as shown in the summary statistics is 8.05%, the difference of 3.2% is large in a real-world sense.

4b From model (2) in the STATA log file, the call-back rate for male blacks is 0.097 − 0.038 = 0.059, and for female blacks is 0.097 − 0.038 + 0.008 = 0.067. The difference is 0.008, which is the coefficient of the interactive term between $black$ and $female$, and is not significant at the 5% level (t-statistic = 0.69).

4c From model (3), the call-back rate for low-quality resumes is 0.073 and the call-back rate for high-quality resumes is 0.073 + 0.014 = 0.087. The difference is 0.014, which is not significant at the 5% level, but is at the 10% level (t-statistic = 1.80).

4d From model (4), the (high-quality)/(low-quality) difference for whites is 0.023 and for blacks is 0.023-0.018=0.005; the black-white difference is 0.018 which is not statistically significant at the 5% level (t-statistic = -1.14).