D0E28A Econometrics Test 2 (Tuesday) - Solutions

KU Leuven

Spring 2022

Question 1

1. $SER = 0.516$: the relative deviations of *wage* from its expected value, given *exper* and *black*, are on the order of 51.6%.

2. The marginal effect of *exper* on *wage* is

$$\frac{\partial \mathrm{E}(\log(wage)|black, exper)}{\partial exper} = \beta_2 + 2\beta_3 exper,$$

which depends on the value of *exper*. For example, when $exper = 5$, the estimated relative marginal effect is

$$\begin{aligned}\frac{\partial \widehat{\mathrm{E}(\log(wage)|black, exper)}}{\partial exper} &= \widehat{\beta}_2 + 2\widehat{\beta}_3 \times 5 \\ &= 0.0641.\end{aligned}$$

This means that for someone with 5 years of work experience, a unit level-change in *exper*, i.e., 1 year, is associated with (on average) a $(0.0641 \times 100)\% = 6.41\%$ increase in the wage.

3. Equation (1) is inadequate to answer the question if there is wage discrimination against blacks. The reason is that $\widehat{\beta}_1$ is likely to be biased due to omitted variables. For example, it is highly likely that $\log(wage)$ depends on the education level (an omitted variable) and that the education level is correlated with *black*, making *black* endogenous. In particular, if black workers on average have fewer years of education and education has a positive effect on wages, then the LS estimate $\widehat{\beta}_1$ is negatively biased. In sum, the negative estimate $\widehat{\beta}_1$ may be due to omitted variables and does not necessarily reflect racial discrimination.

4. The estimate of the effect of *black* when $exper = 5$ is -0.1305, which means the predicted probability of earning a high wage for black people with 5 years work experience is on average 13.05 percentage points lower than that for non-black people with the same work experience.

5. The effect of *black* can be calculated as

$$
\begin{aligned}
&\Pr\left[highwage = 1 \widehat{| black} = 1, exper = 5\right] \\
&\quad - \Pr\left[highwage = 1 \widehat{| black} = 0, exper = 5\right] \\
=&\Lambda\left(\widehat{\beta}_0 + \widehat{\beta}_1 \times 1 + \widehat{\beta}_2 \times 5 + \widehat{\beta}_3 \times 5^2\right) \\
&\quad - \Lambda\left(\widehat{\beta}_0 + \widehat{\beta}_1 \times 0 + \widehat{\beta}_2 \times 5 + \widehat{\beta}_3 \times 5^2\right) \\
=&\frac{1}{1 + \exp\left(-\widehat{\beta}_0 - \widehat{\beta}_1 \times 1 - \widehat{\beta}_2 \times 5 - \widehat{\beta}_3 \times 5^2\right)} \\
&\quad - \frac{1}{1 + \exp\left(-\widehat{\beta}_0 - \widehat{\beta}_1 \times 0 - \widehat{\beta}_2 \times 5 - \widehat{\beta}_3 \times 5^2\right)} \\
=&\frac{1}{1 + \exp\left(-(-1.695) - (-0.528) \times 1 - 0.506 \times 5 - (-0.0292) \times 5^2\right)} \\
&\quad - \frac{1}{1 + \exp\left(-(-1.695) - (-0.528) \times 0 - 0.506 \times 5 - (-0.0292) \times 5^2\right)} \\
=& -0.1305,
\end{aligned}
$$

which agrees with the result in the previous question.

6.

$$
\begin{aligned}
&\widehat{\log(wage_{it})} = \widehat{FE} + \underset{(0.0106)}{0.122}\, exper_{it} - \underset{(0.000688)}{0.00452}\, exper_{it}^2 \\
&R^2 = 0.173 \qquad SER = 0.352
\end{aligned}
$$

7. It is a balanced panel dataset because there are no missing data (for all i and t).

8. Because the regressor *black* is constant over time, it is eliminated by the within-transformation when estimating (3). Therefore, it is not possible to include *black* in (3): $black_i$ would be absorbed into α_i anyway.

Question 2

1. Yes. Apart from *educ* and *exper*, characteristics such as one's professional skills, work attitude, perseverance, etc., also tend to have effects on one's wage, which means they are contained in u in (4). Moreover, it is likely that these characteristics are positively correlated with education. Thus, *educ* is likely to be correlated with u and, hence, to be endogenous.

2. A set of instruments $Z_1, \ldots, Z_l$ must satisfy the following two conditions to be valid:

   (a) the set $Z_1, \ldots, Z_l$ is relevant: $\text{cov}\left(\{Z_1, \ldots, Z_l\}, \{educ, exper, exper^2\}\right)$ has rank 3;

   (b) all $Z_1, \ldots, Z_l$ are exogenous: $\text{corr}(Z_j, u) = 0$ for all j; this means that every Z_j must be uncorrelated with the causal determinants of $\log(wage)$, apart from *educ*, *exper*, and $exper^2$.

3. $motheduc$ and $fatheduc$ need to meet the two conditions in the previous question in order to be valid instruments for *educ*.

   (a) We expect that, regardless of the level of *exper*, $motheduc$ and $fatheduc$ will be positively correlated with *educ*: people tend to be more highly educated if their parents are highly educated. To check this condition, we can run a LS regression of *educ* on *motheduc*, *fatheduc* and the other exogenous regressors in (4):

$$\widehat{educ} = \underset{(0.424)}{9.103} + \underset{(0.0355)}{0.158}\, motheduc + \underset{(0.0324)}{0.190}\, fatheduc + \underset{(0.0419)}{0.0452} exper - \underset{(0.00132)}{0.00101} exper^2.$$

The coefficients of *motheduc* and *fatheduc* are both positive (as expected) and statistically significant at all conventional levels. Thus, there is clear statistical evidence that the relevancy condition is satisfied.

(b) The exogeneity condition that $\mathrm{corr}(motheduc, u) = 0$ and $\mathrm{corr}(fatheduc, u) = 0$ cannot be tested formally. However, it is reasonable to believe they are not exogenous. In particular, it is highly likely that the individual characteristics mentioned in Question 2.1 (professional skills, work attitude, perseverance, etc.) are influenced by *motheduc* and *fatheduc*. This causes correlation between *motheduc* and *fatheduc* on the one hand, and u on the other.

4.

$$\log(\widehat{wage}) = \underset{(0.430)}{0.0481} + \underset{(0.0333)}{0.0614} educ + \underset{(0.0155)}{0.0442} exper - \underset{(0.000430)}{0.000899} exper^2$$

$$R^2 = 0.136 \qquad SER = 0.675$$

5. $\widehat{\beta}_1 = 0.0614$: a unit level-change in *educ*, i.e., 1 year, is associated (on average) with a $\left(\widehat{\beta}_1 \times 100\right)\% =$ 6.14% increase in *wage*.

6. The R output confirms that

   (a) the 2SLS estimates and the IV estimates are the same: $\widehat{\beta}^{2SLS} = \widehat{\beta}^{IV}$;

   (b) the standard errors computed from the second stage of 2SLS are wrong: they are different from those of the IV estimates.

   The standard errors computed from the second stage of 2SLS are wrong because that computation uses the wrong residuals, namely the second-stage residuals

$$y_i - \widehat{x}_i'\widehat{\beta}^{2SLS} = y_i - \widehat{x}_i'\widehat{\beta}^{IV}$$

   instead of the correct residuals

$$\widehat{u}_i = y_i - x_i'\widehat{\beta}^{IV},$$

   which use x_i' instead of $\widehat{x}_i'$. If, instead, the correct residuals $\widehat{u}_i$ had been used, the second stage would also have produced correct standard errors.