1 Problem 14.9

Consider the following log-linear production function, where output (y) is related to a composite input index (x) through the equation

$$lny_t = \beta_1 + \beta_2 lnx_t + e_t$$

The error term e, consists of two components; one is a random completely unpredictable component, whereas the other is a measure of the managerial or technical efficiency of the tth firm. It is believed that the input level of a firm x_t could depend on, and hence be correlated with, the component of e_t that represents technical efficiency. To test this hypothesis the input price p_t facing the tth firm is used as an instrumental variable. Given that price is exogenously determined, it is assumed that price is uncorrelated with technical efficiency. However, x_t and p_t are likely to be correlated because the choice of the level of input will depend on its price.

1.a

Use the data on 40 firms given in Table 14.4 to find least squares and instrumental variable estimators for β_1, and β_2.

y	x	p	lny	lnx
31.18	91.19	50.81	3.439777	4.512945
32.08	112.7	38.56	3.468233	4.724729
30.46	120.67	43.42	3.416414	4.79306
28.28	93.65	59.89	3.342155	4.539564
30.34	116	46.24	3.412467	4.75359
32.29	104.97	59.33	3.474758	4.653675
31.54	104.17	51	3.451257	4.646024
34.35	92.6	49.14	3.536602	4.528289
36.26	98.87	45.66	3.590715	4.593806
31.4	93.16	47.73	3.446808	4.534318
27.82	74.37	56.01	3.325755	4.309053
30.24	109.03	46.68	3.409166	4.691623
32.92	122.5	43.17	3.49408	4.808111
28.33	92.55	54.82	3.343921	4.527749
27.54	88.13	51.02	3.315639	4.478813
32.72	106.68	52.28	3.487987	4.669834
27.43	80.47	54.34	3.311637	4.387884
27.86	97.08	50.21	3.327192	4.575535
28	84.85	49.33	3.332205	4.440885
33.63	86.26	55.39	3.515419	4.457366
33.85	103.04	48.64	3.521939	4.635117
26.25	78.95	56.25	3.267666	4.368815
35.33	114.84	46.58	3.564732	4.74354
36.54	100.38	53.78	3.598408	4.608963
32.77	118.93	44.04	3.489513	4.778535
33.95	107.54	54.44	3.524889	4.677863
38.59	128.48	45.49	3.652993	4.855773
26.14	93.95	54.31	3.263467	4.542763
31.09	112.76	47.17	3.436886	4.725262
33.51	102.94	48.01	3.511844	4.634146
33.81	105.27	28.74	3.520757	4.656528
31.74	97.7	49.84	3.457578	4.581902
31.66	88.2	55.74	3.455054	4.479607
29.97	101.91	44.47	3.400197	4.62409
30.41	82.86	42.69	3.414772	4.417152
28.51	76.39	59.5	3.350255	4.335852
28.39	79.86	59.99	3.346037	4.380275
30.69	92.26	48.56	3.423937	4.524611
25.57	78.4	58.62	3.24142	4.361824
31.04	115.37	61.56	3.435277	4.748144

$$
\begin{aligned}
b &= (Z'Z)^{-1}Z'y && (1)\\
&= \begin{bmatrix} 1.5 \\ 0.4218 \end{bmatrix} && (2)
\end{aligned}
$$

$$
\begin{aligned}
b_{IV} &= (X'Z)^{-1}X'y && (1)\\
&= \begin{bmatrix} -47.9703 \\ 17.2567 \end{bmatrix} && (2)
\end{aligned}
$$

1.b

Use Hausman's specification test on the estimates for β_2 to test whether the input and technical efficiency are correlated. Comment on the two sets of estimates.

H_0: $b = b_{IV}$
H_1: $b \neq b_{IV}$

$$
\begin{aligned}
\sigma^2_{ls} &= \frac{(y'y - b'Z'y)}{T-K}\\
&= 0.0060
\end{aligned}
$$

$$
\begin{aligned}
m_{ls} &= (b - b_{IV})'[\sigma^2_{ls}(Z'Z)^{-1} - \sigma^2_{ls}(X'Z)^{-1}X'X(X'Z)^{-1}]^{-1}(b - b_{IV})\\
&= 168949.867134107
\end{aligned}
$$

Since $m_{ls} > \chi^2_{L-K}$, we reject the null hypothesis and conclude the IV of p_t is invalid.

2

Consider the data used in SCHOOLING.xls. In the practice, we estimated the following model

$$lwage = \beta_1 + \beta_2 edu + \beta_3 exp + \beta_4 exp^2 + \beta_5 black + \beta_6 ma + \beta_7 south + e$$

using the least-squares, the IV, and the 2SLS estimators, and the potential endogeneity of the variable education. In this exercise, we will consider the potential endogeneity of exp (and therefore exp^2) in addition to edu.

2.a

Argue why experience and experience squared are endogenous.

Experience and experience squared is endogenous because greater experience could lead to a promotion, which results in greater income and this could potentially lead to more experience and skills gained, therefore there is a positive feedback loop between experience and wages.

2.b

Estimate the model using the least-squares method. Interpret the parameter estimates.

$$
\begin{aligned}
b &= (Z'Z)^{-1}Zy\\
&= \begin{bmatrix} 4.7314 \\ 0.0762 \\ 0.0852 \\ -0.0024 \\ -0.1770 \\ 0.1127 \\ -0.0960 \end{bmatrix}
\end{aligned}
$$

For every additional year of education, wage increases by 0.0762 percent, ceteris paribus. For experience, since there is exp and exp^2, our model predicts decreasing returns for experience, therefore we must calculate the following:

$$\begin{aligned}\frac{\partial lnwage}{\partial exp} &= \beta_3 + 2\beta_4 exp \\ &= 0.0852 + 2(-0.0024)exp \\ exp &= \frac{0.0852}{2(-0.0024)} \\ &= 18\end{aligned}$$

This means the effect of experience on wage is that it increases wage from ages 0-18, then decreases wage after the age of 18. If the individual is in a metropolitan area, we would expect their wage to increase by 0.1127 percent, ceteris paribus. As for race (black variable), if the individual is black, this results in a 0.177 percent decrease in wages, ceteris paribus. If the individual is in the south, then there is a 0.096 percent decrease in wages, ceteris paribus.

2.c

Use near college variable and age as instruments for edu, exp, and exp^2. Is there any issue?

Yes, this is an under-identified case. There are 2 instrumental variables and 3 parameters, therefore we cannot calculate b_{IV} in this case.

2.d

To solve the issue, you use age and age^2 as instruments for exp and exp^2, respectively.

$$\begin{aligned}b_{IV} &= (X'Z)^{-1}X'y \\ &= \begin{bmatrix} 3.6286 \\ 0.1690 \\ 0.0459 \\ -0.0002877 \\ -0.0539 \\ 0.0669 \\ -0.0438 \end{bmatrix}\end{aligned}$$

2.e

Compare the results from (b) and (d).

The signs of the coefficients are the same from (b) and (d). The magnitude of every coefficient from (d), however, are smaller than (b). However, b_{IV} is similar enough to b.

2.f

Are edu, exp, and exp^2 endogenous?

H_0: edu, exp, and exp^2 are endogenous
H_1: At least one is not endogenous

$$\begin{aligned}m_{ls} &= (b - b_{IV})'[\sigma^2_{ls}(Z'Z)^{-1} - \sigma^2_{ls}(X'Z)^{-1}(X'X)(X'Z)^{-1}]^{-1}(b - b_{IV}) \\ &= 11588\end{aligned}$$

The p-value of m is zero, which means it is significant. We reject our null and conclude at least one of $edu, exp,$ or exp^2 is not endogenous at $\alpha = 0.05$.

2.g

You also consider *momed* and *daded* as potential instruments for *edu*. Re-estimate the model.

$$b_{2SLS} = (X(X'X)^{-1}X'Z)^{-1}Z'X(X'X)^{-1}X'y$$

$$b_{2SLS} = \begin{bmatrix} 4.4117 \\ 0.0951 \\ 0.0929 \\ -0.0024 \\ -0.1599 \\ 0.1053 \\ -0.0885 \end{bmatrix}$$

2.h

Are the instruments valid?

H_0: *nearcollege*, *momed*, and *daded* are valid
H_1: At least one is not valid

We obtain $OIR = 6.9777$ and $\chi^2_c = 5.9915$. Since $OIR > \chi^2_c$ $(6.9777 > 5.9915)$, we reject the null hypothesis and conclude there is sufficient evidence at a 95% significance level that one of the instruments is invalid.

3

The following model tests whether there is a trade-off between participating in a 401(k) plan and having an individual retirement account (IRA)

$$pira = \beta_1 + \beta_2 p401k + \beta_3 inc + \beta_4 inc^2 + \beta_5 age + \beta_6 age^2 + e$$

where *pira* is the probability of participating in an IRA plan, *inc* is the income and *age* is the age of the respondent.

3.a

Use the 401ksubs.xls to estimate the model by OLS and discuss the estimate of $p401k$.

$$b = \begin{bmatrix} -0.1977 \\ 0.0537 \\ 0.0087 \\ -0.000023 \\ -0.0016 \\ 0.00012 \end{bmatrix}$$

3.b

What might be a problem with ordinary least squares (OLS)?

The variables may be correlated or even highly with each other. For example, older individuals would earn more income and with higher income, there could be a higher probability that the individual would participate in a 401k plan.

3.c

The variable $e401k$ is a binary (dummy) variable equal to one if a worker is eligible to participate in 401(k) plan. Explain what is required for $e401k$ to be a valid IV for $p401k$. Do these assumption seem reasonable?

e401k must be correlated with p401k but not correlated with the error term.

3.d

Estimate the first stage and verify that $e401k$ is statistically significant after controlling for other variables?

$$b_{\text{first stage}} \begin{bmatrix} 0.0591493112341638 \\ 0.688845445287413 \\ 0.00111166313624400 \\ 0.00000184100843902913 \\ -0.00472050956521741 \\ 0.0000520374125867843 \end{bmatrix}$$

If X is our model and z is our instrumental variable model, then:

$$b_{2SLS} = [X'z(z'z)z'X]^{-1}X'z(z'z)^{-1}z'y \tag{1}$$

$$= \begin{bmatrix} -0.207313611145421 \\ 0.0207011720541929 \\ 0.00899823830800461 \\ -0.0000241363983119473 \\ -0.00114664910239068 \\ 0.000112070160158153 \end{bmatrix} \tag{2}$$

3.e

Now estimate the model using IV method and compare the estimate of β_1 with the OLS estimate.

If X is our model and z is our instrumental variable model, then:

$$b_{IV} = (X'z)^{-1}X'y$$

$$= \begin{bmatrix} -0.207313611145421 \\ 0.0207011720541929 \\ 0.00899823830800461 \\ -0.0000241363983119473 \\ -0.00114664910239068 \\ 0.000112070160158153 \end{bmatrix}$$

$$b_{IV} = b_{2SLS}$$

$b_{1,IV/2SLS} < b_{1,OLS}$ $(-0.2073 < -0.1977)$. However, the numbers from our IV and OLS estimate are similar.

3.f

Test the null hypothesis that $p401k$ is in fact exogenous.

H_0: $p401k$ is exogenous
H_1: $p401k$ is not exogenous

We obtain $m = 11.7803$. Since $\chi^2_{ls} = 11.0705$ and $m > \chi^2_{ls}$, we fail to reject the null hypothesis.

4

Consider the statistical model $y = X\beta + e$, where $e \sim N(0, \sigma^2 I)$ and

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 0 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$y = \begin{bmatrix} 5 \\ 8 \\ 8 \\ -1 \\ 6 \\ 7 \end{bmatrix}$$

4.a

Use the above information to find the least squares estimates of β_1 and β_2.

$$\begin{aligned} b &= (X'X)^{-1}X'y \\ &= \begin{bmatrix} -0.5 \\ 3 \end{bmatrix} \end{aligned}$$

4.b

Is X statistically significant?
$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

$$\begin{aligned} cov(b) &= \sigma^2 (X'X)^{-1} && (1) \\ &= \frac{y'y - b'X'y}{T-K}(X'X)^{-1} && (2) \\ &= \begin{bmatrix} 0.7292 & -0.2917 \\ -0.2917 & 0.1458 \end{bmatrix} && (3) \end{aligned}$$

$$\begin{aligned} t &= \frac{b_2 - \beta_2}{\sqrt{var(b_2)}} \\ &= \frac{3-0}{\sqrt{0.1458}} \\ &= 7.857 \end{aligned}$$

$t_c = 4.7733$ with $\alpha = 0.05$ and 5 degrees of freedom. Since $t > t_c$ ($7.857 > 4.7733$), we reject our null hypothesis and have sufficient evidence conclude that X is statistically significant.

4.c

Is the regression statistically significant?

As we found in 4(b), since b_2 is the only parameter and it is statistically significant, then this regression is significant.

4.d

One issue with the previous model is that x may be endogenous, that is, $E[xe] \neq 0$. What is the implication of this on the estimates in $a)$?

With $E(xe) \neq 0$, our model could run into both biased and inconsistency problems. This implies we require the instrumental variable approach.

4.e

The use of an instrumental variable (IV) method provides a consistent estimator for β_1 and β_2. The following variable is suggested as an IV for x

$$z = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

What criteria should z have to be considered a valid IV?

$E(z'e) = 0$. z must be uncorrelated with e and must be highly correlated with X.

4.f

Estimate the parameters β_1 and β_2 using the IV method.

$$b_{2SLS} = [X'z(z'z)z'X]^{-1}X'z(z'z)^{-1}z'y$$
$$b_{2SLS} = \begin{bmatrix} 1 \\ 2.25 \end{bmatrix}$$

4.g

Using Hausman test for one parameter, test whether contemporaneous correlation between x and e exists.

$$\begin{aligned} m_{2SLS} &= (b - b_{2SLS})'[\sigma^2_{2SLS}(X' * X)^{-1} - \sigma^2_{2SLS}(z' * X)^{-1}(z' * z)(z' * X)^{-1}]^{-1}(b - b_{2SLS}) \\ &= 1.2857 \end{aligned}$$

4.h

Is z a valid instrument for x?

H_0: z is endogenous
H_1: z is not endogenous

$OIR = 4.0641$ We reject the null hypothesis and conclude there is sufficient evidence that z is not endogenous.