Homework 2

ECON 7023: Econometrics II
Maghfira Ramadhani
February 9, 2022

Spring 2023

Chapter 4

Problem 4.11

a. In Example 4.3, use KWW and IQ simultaneously as proxies for ability in equation (4.29). Compare the estimated return to education without a proxy for ability and with IQ as the only proxy for ability.
Answer:
Table 1 show the estimates for return to education with and without proxies for ability.

Table 1: Regression result for Problem 4.11.a.

	(1)	(2)	(3)
years of work experience	0.013***	0.014***	0.014***
	(0.003)	(0.003)	(0.003)
years with current employer	0.011***	0.011***	0.012***
	(0.002)	(0.002)	(0.002)
=1 if married	0.192***	0.200***	0.199***
	(0.039)	(0.039)	(0.039)
=1 if live in south	-0.082***	-0.080***	-0.091***
	(0.026)	(0.026)	(0.026)
=1 if live in SMSA	0.176***	0.182***	0.184***
	(0.027)	(0.027)	(0.027)
=1 if black	-0.130***	-0.143***	-0.188***
	(0.040)	(0.039)	(0.038)
years of education	0.050***	0.054***	0.065***
	(0.007)	(0.007)	(0.006)
IQ score	0.003***	0.004***	
	(0.001)	(0.001)	
knowledge of world work score	0.004**		
	(0.002)		
Constant	5.176***	5.176***	5.395***
	(0.128)	(0.128)	(0.113)
Observations	935	935	935

Standard errors in parentheses
Data: NLS80.DTA
Wooldridge (2011)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The following code is used to run the regression.

```
reg lwage exper tenure married south urban black educ iq kww
reg lwage exper tenure married south urban black educ iq
reg lwage exper tenure married south urban black educ
```

The regression estimates by using KWW and IQ as proxies is shown in model (1) with the estimated return to education of about 5%. From model (2), When only IQ is used as proxy the estimated return is approximately 5.4%. The case when no proxy is used the return to education estimates is about 6.5% as shown in model (3). OVerall, we have lower estimates when more proxy is used and the inference is still statistically significant.

b. Test KWW and IQ for joint significance in the estimated equation from part a.
Answer:
We can use the F-test for joint significance of both variable with the following hypothesis,

$$\begin{aligned} &\mathrm{H}_0: \ \beta_{KWW} = \beta_{IQ} = 0 \\ &\mathrm{H}_1: \ \text{at least one of } \beta_{KWW} \text{ or } \beta_{IQ} \text{ is not zero} \end{aligned}$$

Below are the output for testing joint significance of KWW and IQ, the P-value is smaller than either $0.1, 0.05$ or 0.01. Thus we can reject the null hypothesis. And we can conclude that there is enough certainty to say that KWW and IQ are jointly significant.

```
. quietly reg lwage exper tenure married south urban black educ iq kww
. test kww iq
 ( 1)  kww = 0
 ( 2)  iq = 0
       F(  2,   925) =    8.59
            Prob > F =    0.0002
```

c. When KWW and IQ are used as proxies for *abil*, does the wage differential between nonblacks and blacks dissapear? What is the estimated differential?
Answer:
From estimates (1) in part a, the wage differential between nonblacks and blacks is still significant. Blacks are estimated to have about 13% lower wage than non blacks holding other factors constant.

d. Add the interaction $educ(IQ - 100)$ and $educ(KWW - \overline{KWW})$ to the regression from part a, where $\overline{KWW}$ is the average score in the sample. Are these terms jointly significant using a standard F test? Does adding them affect any important conclusions?
Answer:
We can use the F-test for joint significance of the interaction variables with the following hypothesis,

$$\begin{aligned} &\mathrm{H}_0: \ \beta_{educ(KWW-\overline{KWW})} = \beta_{educ(IQ-100)} = 0 \\ &\mathrm{H}_1: \ \text{at least one of } \beta_{educ(KWW-\overline{KWW})} \text{ or } \beta_{educ(IQ-100)} \text{ is not zero} \end{aligned}$$

From the F-test, the P-value is 0.015, which indicates a significant level of 1.5%. Thus we can reject the null hypothesis, and conclude with some certainty that $educ(IQ-100)$ and $educ(KWW-\overline{KWW})$ are jointly significant.

```
. gen educiq=educ*(iq-100)
. egen meankww=mean(kww)
. gen educkww=educ*(kww-meankww)
. label variable educkww "educ*(kww-mean(kww))"
. label variable educiq "educ*(iq-100)"
. quietly reg lwage exper tenure married south urban black educ iq kww educiq educkww
. test educiq educkww
 ( 1)  educiq = 0
 ( 2)  educkww = 0
       F(  2,   923) =    4.19
            Prob > F =    0.0154
```

Model (1) in Table 2 show the estimates for return to education by adding interaction terms interaction $educ(IQ-100)$ and $educ(KWW-\overline{KWW})$. The return to education is even smaller know compared to estimates in Table 1 from part a, about 4.5%. Regarding the effect of KWW, as it increases above its mean, the return of education will be larger. If the KWW is one point above the mean, the return to education will be 4.5%+0.02% or about 4.52%. This evidence show that the effect of knowledge of world work score to the return of education is positive, consistent with common intuition or at least my intuition.

Table 2: Regression result for Problem 4.11.d.

	(1)
years of work experience	0.012***
	(0.003)
years with current employer	0.011***
	(0.002)
=1 if married	0.198***
	(0.039)
=1 if live in south	-0.081***
	(0.026)
=1 if live in SMSA	0.178***
	(0.027)
=1 if black	-0.138***
	(0.040)
years of education	0.045***
	(0.008)
IQ score	0.005
	(0.006)
knowledge of world work score	-0.025**
	(0.011)
educ*(iq-100)	-0.000
	(0.000)
educ*(kww-mean(kww))	0.002***
	(0.001)
Constant	6.080***
	(0.561)
Observations	935

Standard errors in parentheses
Data: NLS80.DTA
Wooldridge (2011)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Problem 4.13

Use the data in CORNWELL.RAW (from Cornwell and Trumball, 1994) to estimate a model of county-level crime rates, using the year 1987 only.

a. Using logarithms of all variables, estimate a model relating the crime rate to the deterrent variables $prbarr, prbconv, prbpris$, and $avgsen$.

Answer:
Table 3 shows the estimates of the constant elasticity model (log-log model) of crime rate.

Table 3: Regression result for Problem 4.13.a and 4.13.b.

	(1)	(2)
log(prbarr)	-0.724***	-0.185***
	(0.115)	(0.063)
log(prbconv)	-0.473***	-0.039
	(0.083)	(0.047)
log(prbpris)	0.160	-0.127
	(0.206)	(0.099)
log(avgsen)	0.076	-0.152*
	(0.163)	(0.078)
log(crmrte)[t-1]		0.780***
		(0.045)
Constant	-4.868***	-0.767**
	(0.432)	(0.313)
Observations	90	90

Standard errors in parentheses
Data: CORNWELL.DTA
Cornwell and Trumball (1994)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We might expect that crime rate will decrease as the probability of arrest and probability of conviction increase. As shown in model (1), the elasticities of crime or coefficient on probability of arrest and probability of conviction have negative sign as expected and are statistically significant. The elasticities of crime on probability of serving prison term and average sentence length showing positive sign but statistically not significant.

b. Add log(*crmrte*) for 1986 as an additional explanatory variable, and comment on how the estimated elasticities differ from part a.
Answer:
As shown in estimate of model (2) in Table 3, by adding log(*crmrte*) for 1986 as an additional explanatory variable there are some change in the elasticities. First, the elasticity of crime rate on probability of arrest decrease significantly and still statistically significant. Second, the elasticity of crime rate with respect to probability of conviction also decrease significantly and become statistically insignificant. The elasticities on probability of serving prison term and on average sentence length remain insignificant and both change sign, with the latter become statistically significant at 10% level. The elasticities on the lagged crime rate is comparably large and statistically significant.

c. Compute the F statistic for joint significance of all of the wage variables (again in logs), using the restricted model from part b.
Answer:
We can use the F-test for joint significance of all of the wage variables with the following hypothesis, in this case we use assume homoskedasticity.

$$\begin{aligned} &\mathrm{H}_0: \ \beta_{lwcon} = \ldots = \beta_{lwloc} = 0 \\ &\mathrm{H}_1: \ \text{at least one of } \beta_{lwcon} \text{ or } \ldots \text{ or } \beta_{lwloc} \text{ is not zero} \end{aligned}$$

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lcrmrte1 lwcon-lwloc if d87
```

```
      Source |       SS           df       MS      Number of obs   =        90
-------------+----------------------------------   F(14, 75)       =     43.81
       Model |  23.8798774        14  1.70570553   Prob > F        =    0.0000
    Residual |  2.91982063        75  .038930942   R-squared       =    0.8911
-------------+----------------------------------   Adj R-squared   =    0.8707
       Total |   26.799698        89  .301120202   Root MSE        =    .19731

------------------------------------------------------------------------------
     lcrmrte | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.1725122   .0659533    -2.62   0.011    -.3038978   -.0411265
    lprbconv |  -.0683639    .049728    -1.37   0.173    -.1674273    .0306994
    lprbpris |  -.2155553   .1024014    -2.11   0.039    -.4195493   -.0115614
     lavgsen |  -.1960546   .0844647    -2.32   0.023     -.364317   -.0277923
    lcrmrte1 |   .7453414   .0530331    14.05   0.000     .6396942    .8509887
       lwcon |  -.2850008   .1775178    -1.61   0.113    -.6386344    .0686327
       lwtuc |   .0641312    .134327     0.48   0.634    -.2034619    .3317244
       lwtrd |    .253707   .2317449     1.09   0.277    -.2079525    .7153665
       lwfir |  -.0835258   .1964974    -0.43   0.672    -.4749687    .3079171
       lwser |   .1127542   .0847427     1.33   0.187    -.0560619    .2815703
       lwmfg |   .0987371   .1186099     0.83   0.408    -.1375459    .3350201
       lwfed |   .3361278   .2453134     1.37   0.175    -.1525615    .8248172
       lwsta |   .0395089   .2072112     0.19   0.849    -.3732769    .4522947
       lwloc |  -.0369855   .3291546    -0.11   0.911    -.6926951    .6187241
       _cons |  -3.792525   1.957472    -1.94   0.056    -7.692009    .1069593
------------------------------------------------------------------------------

. testparm lwcon-lwloc
 ( 1)  lwcon = 0
 ( 2)  lwtuc = 0
 ( 3)  lwtrd = 0
 ( 4)  lwfir = 0
 ( 5)  lwser = 0
 ( 6)  lwmfg = 0
 ( 7)  lwfed = 0
 ( 8)  lwsta = 0
 ( 9)  lwloc = 0
       F(  9,    75) =    1.50
            Prob > F =    0.1643
```

From the F-test we have that all of the wage variables are jointly insignificant since the P-value is more than 0.1. We might also interested in looking at individual elasticity and find out that the sign of the elasticity are not consistent.

d. Redo part c, but make the test robust to heteroskedasticity of unknown form.
Answer:
Analogous to c, we use the F-test for joint significance of all of the wage variables with the following hypothesis, now we use assume heteroskedasticity and use robust standard error.

$$\begin{aligned} \mathrm{H}_0 &: \ \beta_{lwcon} = \ldots = \beta_{lwloc} = 0 \\ \mathrm{H}_1 &: \ \text{at least one of } \beta_{lwcon} \text{ or } \ldots \text{ or } \beta_{lwloc} \text{ is not zero} \end{aligned}$$

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lcrmrte1 lwcon-lwloc if d87, robust
Linear regression                               Number of obs     =         90
                                                F(14, 75)         =     110.75
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8911
                                                Root MSE          =     .19731

------------------------------------------------------------------------------
             |               Robust
     lcrmrte | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.1725122   .0831236    -2.08   0.041    -.3381028   -.0069215
    lprbconv |  -.0683639   .0874696    -0.78   0.437    -.2426123    .1058844
    lprbpris |  -.2155553   .0895319    -2.41   0.019    -.3939121   -.0371986
     lavgsen |  -.1960546   .0976231    -2.01   0.048    -.3905298   -.0015794
    lcrmrte1 |   .7453414   .1594535     4.67   0.000     .4276937    1.062989
       lwcon |  -.2850008   .1276141    -2.23   0.029    -.5392212   -.0307805
```

```
       lwtuc |   .0641312   .1108165     0.58   0.565    -.1566265     .284889
       lwtrd |    .253707   .1712913     1.48   0.143    -.0875227    .5949368
       lwfir |  -.0835258   .1477461    -0.57   0.574     -.377851    .2107995
       lwser |   .1127542   .0715635     1.58   0.119    -.0298077     .255316
       lwmfg |   .0987371   .1083497     0.91   0.365    -.1171065    .3145807
       lwfed |   .3361278   .4416827     0.76   0.449    -.5437491    1.216005
       lwsta |   .0395089   .1829791     0.22   0.830    -.3250042     .404022
       lwloc |  -.0369855   .2825442    -0.13   0.896    -.5998425    .5258714
       _cons |  -3.792525   3.383901    -1.12   0.266     -10.5336    2.948552
------------------------------------------------------------------------------

. testparm lwcon-lwloc

 ( 1)  lwcon = 0
 ( 2)  lwtuc = 0
 ( 3)  lwtrd = 0
 ( 4)  lwfir = 0
 ( 5)  lwser = 0
 ( 6)  lwmfg = 0
 ( 7)  lwfed = 0
 ( 8)  lwsta = 0
 ( 9)  lwloc = 0

       F(  9,    75) =    2.19
            Prob > F =    0.0319
```

Since the P-value is less than 0.1, we will reject the null hypothesis. Now the F-test show that we have all of the wage variables jointly significant. Looking at individual significance, the only statistically significant elasticity is on variable *lwcon* with negative sign.

Problem 4.14

Use the data in ATTEND.RAW to answer this question

a. To determine the effects of attending lecture on final exam performance, estimate a model relating *stndfnl* (the standardized final exam score) to *atndrte* (the percent of lectures attended). Include the binary variables $frosh$ and *soph* as explanatory variables. Interpret the coefficient on *atndrte*, and discuss its significance.
Answer:
Model (1) in Table 4 shows estimate of a model relating standardized final exam score to the percent of lectures attended and include the binary variables $frosh$ and *soph*. The coefficient of the percent of lectures attended is significant at 5% level. Note that final exam score is standardized so, the interpretation is if the attendance rate increase by 10% the standardized test score will increase by around 10×0.008 or 0.08 standard deviation.

b. How confident are you that the OLS estimates from part a are estimating the causal effect of attendance? Explain.
Answer:
It looks like the data is not from a random experiment, which indicate that there is likely to be a selection bias. It may also be possible that the student with higher exam grade or better students is the one who attend the class regularly, not because they attend the class regularly so that they get a good grade. There is also other variable that may be omitted but affect attendance rate for example the distance of the class room from the place where student lives on campus. Thus, controlling for only year in college by adding $frosh$ and *soph* dummy may not solve the endogeneity problem.

c. As proxy variables for student ability, add to the regression *priGPA* (prior cumulative GPA) and *ACT* (achievement test score). Now what is the effect of *atndrte*? Discuss how the effect differs from that in part a.
Answer:
Model (2) in Table 4 shows the estimate. Now the effect of attendance rate decrease compared to the result from part a. The effect is still statistically significant but at 5% level which is weaker than the 1% level in part a. The effect of *priGPA* (prior cumulative GPA) and *ACT* (achievement test score) is positive on the standardized exam score and both very significant.

Table 4: Regression result for Problem 4.14.a., 4.14.c, 4.14.e, and 4.14.f

	(1)	(2)	(3)	(4)
percent classes attended	0.008***	0.005**	0.006***	0.006
	(0.002)	(0.002)	(0.002)	(0.011)
=1 if freshman	-0.290**	-0.049	-0.105	-0.105
	(0.116)	(0.108)	(0.107)	(0.107)
=1 if sophomore	-0.118	-0.160*	-0.181**	-0.181**
	(0.099)	(0.090)	(0.089)	(0.089)
cumulative GPA prior to term		0.427***	-1.526***	-1.525***
		(0.082)	(0.474)	(0.476)
ACT score		0.084***	-0.112	-0.112
		(0.011)	(0.098)	(0.098)
priGPA^2			0.368***	0.368***
			(0.089)	(0.089)
ACT^2			0.004*	0.004*
			(0.002)	(0.002)
atndrte^2				0.000
				(0.000)
Constant	-0.502**	-3.297***	1.385	1.394
	(0.196)	(0.309)	(1.239)	(1.267)
Observations	680	680	680	680

Standard errors in parentheses

Data: ATTEND.DTA

Wooldridge (2011)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

d. What happens to the significance of the dummy variables in part c as compared with part a? Explain.
Answer:
The significance of the dummy variable changes after controlling for $priGPA$ (prior cumulative GPA) and ACT (achievement test score). The coefficient of *soph*, which indicate the effect of being in sophomore relative to being in year three or above, still have same sign with decreased magnitude, and become statistically significant at 10% level. The coefficient of $frosh$, which indicate the effect of being in freshmen relative to being in year three or above, still have same sign with significantly lower magnitude, and become statistically insignificant. We can say that adding the two new dummy eliminate the effect of being a freshman.

e. Add the squares of $priGPA$ and ACT to the equation. What happens to the coefficient *atndrte*? Are the quadratics jointly significant?
Answer:
Model (3) in Table 4 shows the estimate after adding the squares of $priGPA$ and ACT. The coefficient on the square of $priGPA$ is very significant, while the coefficient on the square of ACT is significant only at 10% level. The effect on the attendance rate after adding the two variables now increases compared to model (2) and become more statistically significant. However, note that the effect on ACT now become insignificant and changes sign compared to model (2).
We can also test using F-test by using the following hypothesis.

$$\begin{aligned} \mathrm{H}_0 &: \ \beta_{priGPAsq} = \beta_{ACT} = 0 \\ \mathrm{H}_1 &: \ \text{at least one of } \beta_{priGPAsq} \text{ or } \beta_{ACT} \text{ is not zero} \end{aligned}$$

```
. quietly reg stndfnl atndrte frosh soph priGPA ACT priGPAsq ACTsq
. test priGPAsq ACTsq
 ( 1)  priGPAsq = 0
 ( 2)  ACTsq = 0
       F(  2,   672) =   11.30
            Prob > F =    0.0000
```

The F-test yield a near-zero P-value, so we will reject the null hypothesis. The result indicates that both variable are jointly significant.

f. To test for a nonlinear effect of *atndrte*, add its square to the equation from part e. What do you conclude?
Answer:
Model (4) in Table 4 shows that the coefficient on the square of the attendance rate is very insignificant. Also, note that the standard error of the coefficient of the attendance rate increase by over five fold become statistically not significant. This evidence suggests the attendance rate and its square might be highly collinear. Table 5 shows that the two variable is highly correlated.

Table 5: Correlation coefficient for Problem 4.14.f

	(1) percent classes attended	atndrte^2
percent classes attended	1.000	
atndrte^2	0.983***	1.000

Data: ATTEND.DTA
Wooldridge (2011)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Problem 4.15

Assume that y and x_j have finite second moments, and write the linear projection of y on $(1, x_1, \ldots, x_K)$ as

$$\begin{aligned} &y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u = \beta_0 + \mathbf{x}\boldsymbol{\beta} + u, \\ &E(u) = 0, \qquad E(x_j u) = 0, \qquad j = 1, 2, \ldots, K. \end{aligned}$$

a. Show that $\sigma_y^2 = \mathrm{Var}(\mathbf{x}\boldsymbol{\beta}) + \sigma_u^2$.

Answer:

We know that y and x_j have finite second moments, let $\mathrm{Var}(y) = \sigma_y^2$ and $\mathrm{Var}(x_j) = \sigma_{x_j}^2$. Also we can take β_0 as constant and $\boldsymbol{\beta}$ as a vector of constant since this is a linear projection model. Recall that $y = \beta_0 + \mathbf{x}\boldsymbol{\beta} + u$, so we have

$$\begin{aligned} &\mathrm{Var}(y) = \mathrm{Var}(\beta_0 + \mathbf{x}\boldsymbol{\beta} + u) \\ &\Leftrightarrow \mathrm{Var}(y) = \mathrm{Var}(\mathbf{x}\boldsymbol{\beta} + u) && [\text{ because } \beta_0 \text{ is constant}] \\ &\Leftrightarrow \sigma_y^2 = \mathrm{Var}(\mathbf{x}\boldsymbol{\beta}) + \mathrm{Var}(u) + \sum_{i=1}^{K} \mathrm{Cov}(x_i\beta_i, u) \\ &\Leftrightarrow \sigma_y^2 = \mathrm{Var}(\mathbf{x}\boldsymbol{\beta}) + \sigma_u^2 + \sum_{i=1}^{K} \beta_i \mathrm{Cov}(x_i, u) \\ &\Leftrightarrow \sigma_y^2 = \mathrm{Var}(\mathbf{x}\boldsymbol{\beta}) + \sigma_u^2. && [\text{ because } E(x_i u) = \mathrm{Cov}(x_i, u) = 0] \end{aligned}$$

□

b. For a random draw i from the population, write $y_i = \beta_0 + \mathbf{x}_i\boldsymbol{\beta} + u_i$. Evaluate the following assumption, which has been known to appear in econometrics textbooks: "$\mathrm{Var}(u_i) = \sigma^2 = \mathrm{Var}(y_i)$ for all i."

Answer:

It looks like we are drawing finite sample $\mathbf{x}_i, y_i$ from the population. Write the variance of y_i as follow

$$\begin{aligned} \mathrm{Var}(y_i) &= \mathrm{Var}(\beta_0 + \mathbf{x}_i\boldsymbol{\beta} + u_i) \\ &= \mathrm{Var}(\beta_0) + \mathrm{Var}(\mathbf{x}_i\boldsymbol{\beta}) + \mathrm{Var}(u_i) + \mathrm{Cov}(\mathbf{x}_i\boldsymbol{\beta}, u_i) + \mathrm{Cov}(\beta_0, u_i) + \mathrm{Cov}(\mathbf{x}_i\boldsymbol{\beta}, \beta_0) \qquad (1) \end{aligned}$$

The condition for the given statement "$\mathrm{Var}(u_i) = \sigma^2 = \mathrm{Var}(y_i)$ for all i." to be true is when all of the terms in (1) except $\mathrm{Var}(u_i)$ are zero. We are given that $E(u) = 0$ and $E(x_j u) = 0$ which lead us to say that the covariance terms are zero from the result in part a. Now we are left with $\mathrm{Var}(\mathbf{x}_i\boldsymbol{\beta})$. For it to be zero we need to make assumption which may not make sense for example for $\mathbf{x}_i$ to be non random which is a contradiction to the random draw that we generate, or that the $\boldsymbol{\beta}$ is zero, which is strange. It also depends on the sample size that we draw, we may find out that some of the variance changes depends on the size of the sample. If we use OLS to estimate the $\widehat{\boldsymbol{\beta}}$ we need large enough sample to converge to the actual $\boldsymbol{\beta}$, thus in this case we may also treat $\widehat{\boldsymbol{\beta}}$ as random.

c. Define the population R-squared by $\rho^2 \equiv 1 - \sigma_u^2/\sigma_y^2 = \mathrm{Var}(\mathbf{x}\boldsymbol{\beta})/\sigma_y^2$. Show that the R-squared, $R^2 = 1 - SSR/SST$, is a consistent estimator of ρ^2, where SSR is the OLS sum of squared residuals and SST$= \sum_{i=1}^{N}(y_i - \bar{y})^2$ is the total sum of squares.

Answer:

We need to show that $\mathrm{plim}(R^2) = \rho^2$. Recall that

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{N} \widehat{u}_i^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}.$$

Also, recall that

$$\begin{aligned} &\mathrm{plim}(\frac{1}{N}\sum_{i=1}^{N} \widehat{u}_i^2) = \mathrm{E}(u_i^2) = \sigma_u^2, \\ &\mathrm{plim}(\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2) = \mathrm{E}((y_i - \mathrm{E}(y_i))^2) = \sigma_y^2. \end{aligned}$$

Now take plim of R^2, we have

$$\begin{aligned}
\text{plim}(R^2) &= \text{plim}\left(1 - \frac{\sum_{i=1}^N \widehat{u}_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2}\right) \\
&= \text{plim}\left(1 - \frac{\frac{1}{N}\sum_{i=1}^N \widehat{u}_i^2}{\frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2}\right) \\
&= 1 - \frac{\text{plim}(\frac{1}{N}\sum_{i=1}^N \widehat{u}_i^2)}{\text{plim}(\frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2)} && \text{[by Slutsky's theorem]} \\
&= 1 - \sigma_u^2/\sigma_y^2.
\end{aligned}$$

□

d. Evaluate the following statement: "In the presence of heteroskedasticity, the R-squared from an OLS regression is meaningless" (This kind of statement also tends to appear in econometrics texts.)
Answer:
From part c we know that R^2 is a consistent estimator for ρ^2. The value of R^2 really depends on the value of the variance of u and y. Note that this is not a conditional variance, $\text{Var}(u|x)$, so even if we impose heteroskedasticity condition it have no effect on $\text{Var}(u)$, the consistency of R^2 will hold. Thus, the statement in the problem is not correct.

Chapter 5

Problem 5.1

In this problem you are to establish the algebraic equivalence between 2SLS and OLS estimation of an equation containing additional regressor. Although the result is completely general, for simplicity consider a model with a single (suspected) endogenous variable:

$$\begin{aligned}
y_1 &= \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + u_1, \\
y_2 &= \mathbf{z}\boldsymbol{\pi}_2 + v_2.
\end{aligned}$$

For notational clarity, we use y_2 as the suspected endogenous variable and $\mathbf{z}$ as the vector of all exogenous variables. The second equation is the reduced form for y_2. Assume that $\mathbf{z}$ has at least one more element than $\mathbf{z}_1$. We know that one estimator of $(\boldsymbol{\delta}_1, \alpha_1)$ is the 2SLS estimator using instruments $\mathbf{x}$. Consider an alternative estimator of $(\boldsymbol{\delta}_1, \alpha_1)$: (a) estimate the reduced form by OLS, save the residuals $\widehat{v}_2$; (b) estimate the following equation by OLS:

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \rho_1 \widehat{v}_2 + error. \tag{5.52}$$

Show that the OLS estimates of $\boldsymbol{\delta}_1$ and α_1 from this regression are identical to the 2SLS estimators. (Hint: Use the partitioned regression algebra of OLS. In particular, if $\widehat{y} = \mathbf{x}_1\widehat{\boldsymbol{\beta}}_1 + \mathbf{x}_2\widehat{\boldsymbol{\beta}}_2$ is an OLS regression, $\widehat{\boldsymbol{\beta}}_1$ can be obtained by first regressing $\mathbf{x}_1$ on $\mathbf{x}_2$, getting the residuals, say $\ddot{\mathbf{x}}_1$, and then regressing y on $\ddot{\mathbf{x}}_1$; see, for example, Davidson and MacKinnon (1993, Section 1.4). You must also use the fact that $\mathbf{z}_1$ and $\widehat{v}_2$ are orthogonal in the sample.)
Answer:
Based on the hint, denote $\mathbf{x}_1 = (\mathbf{z}_1, y_2)$, $\widehat{\boldsymbol{\beta}}_1 = (\widehat{\boldsymbol{\delta}}_1, \widehat{\alpha}_1)'$, $\mathbf{x}_2 = \widehat{v}_2$, and $\widehat{\beta_2} = \widehat{\rho}_1$. By partitioning out the regression following the hint. We can do the following steps

1. Regress $\mathbf{x}_1$ on $\mathbf{x}_2 = \widehat{v}_2$ and save the residual $\ddot{\mathbf{x}}_1$.
By construction $\mathbf{x}_1$ consists of $\mathbf{z}_1$ and y_2. Now we need to investigate the residuals that we will get. Note that the suspected endogenous variable is y_2, and $\mathbf{z}_1$ is exogenous that is $\text{E}(\mathbf{z}_1'\widehat{u}_1) = \text{E}(\mathbf{z}_1'\widehat{v}_2) = 0$. Thus from regressing $\mathbf{z}_1$ on $\widehat{v}_2$, the residuals is $\mathbf{z}_1$ itself. Also from the first stage regression we will have $y_2 = \widehat{y}_2 + \widehat{v}_2$. Thus from regressing y_2 on $\widehat{v}_2$ the residuals is $\widehat{y}_2$. Thus we get the residuals $\ddot{\mathbf{x}}_1 = (\mathbf{z}_1, \widehat{y}_2)$.

2. Regress y_1 on $\ddot{\mathbf{x}}_1$.
Now the second step, we will regress y_1 on $\ddot{\mathbf{x}}_1 = (\mathbf{z}_1, \widehat{y}_2)$. It turns out the regression that we will perform is the same as our second stage regression in 2SLS. Thus the $\widehat{\boldsymbol{\beta}}_1$ that we will get by partialling out is the same as if we do it by 2SLS.

□

Problem 5.3

Consider the following model to estimate the effects of several variables, including cigarette smoking, on the weight of newborns:

$$\log(bwght) = \beta_0 + \beta_1 male + \beta_2 parity + \beta_3 \log(faminc) + \beta_4 packs + u, \tag{5.54}$$

where *male* is a binary indicator equal to one if the child is male, *parity* is the birth order of this child, *faminc* is family income, and *packs* is the average number of packs of cigarettes smoked per day during pregnancy.

a. Why might you expect *packs* to be correlated with u?
Answer:
There may be other factor that affect the weight of new born that correlates with the average number of packs of cigarettes smoked per day that is omitted and thus contained in u. For example, women who smoke cigarettes while pregnant is more likely to regularly check their baby's health to doctor during pregnancy, or if we assume people who smoke on average also drink coffee, then the intake of caffeine might be correlated with the average packs of cigarettes smoked per day. Other example is the level of states attention to smoking behaviour, if the states care about the bad effect of smoking they tend to implement regulation to reduce smoking behaviour for general population, such as introduce high taxes for cigarettes. So states may also be correlated with packs and thus contained in u.

b. Suppose that you have data on average cigarette price in each woman's state of residence. Discuss whether this information is likely to satisfy the properties of a good instrumental variable for packs.
Answer:
It is tempting to say that it might be a good IV, since the average cigarette price should have negative correlation with average cigarette consumption by the law of demand. This satisfy the relevance condition for good IV. However, the cigarette price is also affected by the state policy such as tax and other instrument as mentioned in part a, that may be contained in u. Thus, it may fail the exogeneity condition for good IV.

c. Use the data in BWGHT.RAW to estimate equation (5.54). First, use OLS. Then, use 2SLS, where *cigprice* is an instrument for *packs*. Discuss any important differences in the OLS and 2SLS estimates.
Answer:
OLS Regression

```
. reg lbwght male parity lfaminc packs

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     12.55
       Model |  1.76664363         4  .441660908   Prob > F        =    0.0000
    Residual |    48.65369     1,383  .035179819   R-squared       =    0.0350
-------------+----------------------------------   Adj R-squared   =    0.0322
       Total |  50.4203336     1,387  .036352079   Root MSE        =    .18756

------------------------------------------------------------------------------
      lbwght | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        male |   .0262407   .0100894     2.60   0.009     .0064486    .0460328
      parity |   .0147292   .0056646     2.60   0.009     .0036171    .0258414
     lfaminc |   .0180498   .0055837     3.23   0.001     .0070964    .0290032
       packs |  -.0837281   .0171209    -4.89   0.000    -.1173139   -.0501423
       _cons |   4.675618   .0218813   213.68   0.000     4.632694    4.718542
------------------------------------------------------------------------------
```

2SLS Regression using *cigprice* as instrument for *packs*

```
. ivreg lbwght male parity lfaminc (packs = cigprice)
Instrumental variables 2SLS regression
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =      2.39
       Model |  -91.350027         4 -22.8375067   Prob > F        =    0.0490
    Residual |  141.770361     1,383  .102509299   R-squared       =         .
-------------+----------------------------------   Adj R-squared   =         .
       Total |  50.4203336     1,387  .036352079   Root MSE        =    .32017

------------------------------------------------------------------------------
      lbwght | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       packs |   .7971063   1.086275     0.73   0.463    -1.333819    2.928031
        male |   .0298205    .017779     1.68   0.094    -.0050562    .0646972
      parity |  -.0012391   .0219322    -0.06   0.955     -.044263    .0417848
     lfaminc |    .063646   .0570128     1.12   0.264    -.0481949    .1754869
       _cons |   4.467861   .2588289    17.26   0.000     3.960122    4.975601
------------------------------------------------------------------------------
Instrumented: packs
 Instruments: male parity lfaminc cigprice
```

The estimated effect of *packs* from the OLS and 2SLS are very different. The OLS estimate suggests that smoking one more pack of cigarettes decrease birthweight by 8.37% and is statistically very significant. On the other hand, the 2SLS estimate suggests that additional pack of cigarettes smoked increase birthweight 79.7% and is not significant, which does not make sense at all. Again, as stated in part b, it may be because *cigprice* is not a good IV.

d. Estimate the reduced form for *packs*. What do you conclude about identification of equation (5.54) using *cigprice* as an instrument for *packs*? What bearing does this conclusion have on your answer from part c?
Answer:
Reduced form Regression for *packs*

```
. reg packs male parity lfaminc cigprice
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     10.86
       Model |  3.76705108         4   .94176277   Prob > F        =    0.0000
    Residual |  119.929078     1,383  .086716615   R-squared       =    0.0305
-------------+----------------------------------   Adj R-squared   =    0.0276
       Total |  123.696129     1,387  .089182501   Root MSE        =    .29448

------------------------------------------------------------------------------
       packs | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        male |  -.0047261   .0158539    -0.30   0.766    -.0358264    .0263742
      parity |   .0181491   .0088802     2.04   0.041     .0007291    .0355692
     lfaminc |  -.0526374   .0086991    -6.05   0.000    -.0697023   -.0355724
    cigprice |    .000777   .0007763     1.00   0.317    -.0007459    .0022999
       _cons |   .1374075   .1040005     1.32   0.187    -.0666084    .3414234
------------------------------------------------------------------------------
```

From the reduced form, we can finally see that the effect of *cigprice* on *packs* is not statistically significant and also the sign is positive which is against my expected sign explained in part b. In fact, now the *cigprice* also fails the relevance condition for a good IV.

Problem 5.5

One occasionally sees the following reasoning used in applied work for choosing instrumental variables in the context of omitted variables. The model is

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \gamma q + a_1.$$

where q is the omitted factor. We assume that a_1 satisfies the structural error assumption $\mathrm{E}(a_1|\mathbf{z}_1, y_2, q) = 0$, that $\mathbf{z}_1$ is exogenous in the sense that $\mathrm{E}(q|\mathbf{z}_1) = 0$, but that y_2 and q may be correlated. Let $\mathbf{z}_2$ be a vector of instrumental variable candidates for y_2. Suppose it is known that $\mathbf{z}_2$ appears in the linear projection of y_2 onto $(\mathbf{z}_1, \mathbf{z}_2)$, and so the requirement that $\mathbf{z}_2$ be partially correlated with y_2 is satisfied. Also, we are willing to assume that $\mathbf{z}_2$ is redundant in the structural equation, so that a_1 is uncorrelated with $\mathbf{z}_2$. What we are unsure of is whether $\mathbf{z}_2$ is correlated with the omitted variable q, in which case $\mathbf{z}_2$ would not contain valid IVs. To "test" whether $\mathbf{z}_2$ is in fact uncorrelated with q, it has been suggested to use OLS on the equation

$$y = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \mathbf{z}_2\boldsymbol{\psi}_1 + u_1, \tag{5.55}$$

where $u_1 = \gamma q + a_1$, and test $\mathrm{H}_0 : \boldsymbol{\psi}_1 = 0$. Why does this method not work?
Answer:
We want to test the hypothesis that $\mathbf{z}_2$ is uncorrelated with q. Suppose $\mathbf{z}_2$ is uncorrelated with q then $\mathbf{z}_2$ is exogenous in (5.55). The endogeneity problem is still exist in (5.55) since y_2 is correlated with q. It means if we perform a regression for (5.55), i.e., y on $(\mathbf{z}_2, y_2, \mathbf{z}_1)$ the OLS estimator will not be consistent. In other words, the coefficient that we get from OLS, that is $(\widehat{\boldsymbol{\delta}}_1, \widehat{\alpha}_1, \widehat{\boldsymbol{\psi}}_1)$, will not be consistent. Taking the inconsistent estimator for testing a hypothesis will risk us having an invalid conclusion, thus this method will not work.

Non-textbook Problem

Show that IV estimation can be implemented as the 2SLS procedure.
Answer:
Suppose we have the following model

$$y = \mathbf{x}\boldsymbol{\beta} + u,$$

with $\mathbf{x} = (1, x_1, \ldots, x_K)$, with x_K as the endogenous variable. Now assume we use m instrumental variables $(z_1, \ldots, z_m)$ for x_K. Denote $\mathbf{z} = (1, x_1, \ldots, x_{K-1}, z_1, \ldots, z_m)$ as the regressor for the reduced form. From the first stage we will get the fitted value, $\widehat{x}_K$. The first stage will yield the following estimates

$$x_K = \underbrace{\widehat{\delta}_0 + \widehat{\delta}_1 x_1 + \ldots + \widehat{\delta}_{K-1} x_{K-1} + \widehat{\theta}_1 z_1 + \ldots + \widehat{\theta}_m z_m}_{\widehat{x}_K = \mathbf{z}\widehat{\boldsymbol{\delta}}} + \widehat{r}_K$$

Now denote $\widehat{\mathbf{x}} = (1, x_1, \ldots, x_{K-1}, \widehat{x}_K)$ as the regressor for the second stage regression. The second stage regression will yield the following estimates

$$y = \widehat{\mathbf{x}}\widehat{\boldsymbol{\beta}} + error.$$

The estimates from the 2SLS will be the OLS estimator from the second stage, that is

$$\widehat{\beta}_{2SLS} = (\widehat{\mathbf{x}}'\widehat{\mathbf{x}})^{-1}(\widehat{\mathbf{x}}'y).$$

Recall the IV estimator

$$\widehat{\beta}_{IV} = (\widehat{\mathbf{x}}'\mathbf{x})^{-1}(\widehat{\mathbf{x}}'y).$$

We need to show that the IV estimator is the same as the 2SLS estimator. From the first stage we have $\widehat{x}_K = \mathbf{z}\widehat{\boldsymbol{\delta}}$. We know that the OLS estimator for the first stage is $\widehat{\boldsymbol{\delta}} = (\mathbf{z}'\mathbf{z})^{-1}(\mathbf{z}'x_K)$. Thus we have

$$\widehat{x}_K = \mathbf{z}\widehat{\boldsymbol{\delta}} = \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}(\mathbf{z}'x_K).$$

Since $x_1, \ldots, x_{K-1}$ is in $\mathbf{z}$. We will have

$$\widehat{\mathbf{x}} = \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}(\mathbf{z}'\mathbf{x}).$$

Now, we only need to show that $\widehat{\mathbf{x}}'\widehat{\mathbf{x}} = \widehat{\mathbf{x}}'\mathbf{x}$ to prove that $\widehat{\beta}_{2SLS} = \widehat{\beta}_{IV}$. Denote $p_z = \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'$ and we have that $\widehat{\mathbf{x}} = p_z\mathbf{x}$. Note that p_z is symmetric and is an idempotent matrix with the property $p_z p_z = p_z$. Now we have

$$
\begin{aligned}
\widehat{\beta}_{2SLS} &= (\widehat{\mathbf{x}}'\widehat{\mathbf{x}})^{-1}(\widehat{\mathbf{x}}'y) \\
&= (p_z\mathbf{x})'(p_z\mathbf{x})(\widehat{\mathbf{x}}'y) \\
&= (\mathbf{x}'p_z)(p_z\mathbf{x})(\widehat{\mathbf{x}}'y) \\
&= (\mathbf{x}'p_z\mathbf{x})(\widehat{\mathbf{x}}'y) \\
&= ((p_z\mathbf{x})'\mathbf{x})(\widehat{\mathbf{x}}'y) \\
&= (\widehat{\mathbf{x}}'\mathbf{x})(\widehat{\mathbf{x}}'y) \\
&= \widehat{\beta}_{IV}.
\end{aligned}
$$

□