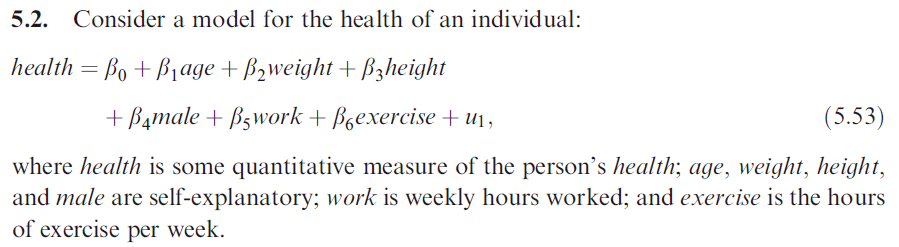
Yi Ma

Homework 9



1. Why might you be concerned about exercise being correlated with the error term?

The error term involves unobserved factors that might be correlated with the variable *exercise* and also determine the health of an individual. For example, income is likely to be correlated with exercise. People with higher income can afford exercise-related activities such as getting gym membership or hiring a personal trainer. Therefore, higher income might induce higher frequency of exercise. Also, people with higher income are more aware of their health status because of more access to health providers so higher income is likely to promote better health.

1. Discuss whether *disthome* and *distwork* are likely to be uncorrelated with the error term.

If the sample selection follows a randomized setting, meaning that people do not systematically choose the location of their homes and jobs relative to health clubs, it is reasonable to believe that these two factors are uncorrelated with each other. However, the location of health clubs might depend on the location of residential areas or the location of nearby companies. Therefore, the location of health clubs may not be exogenous.

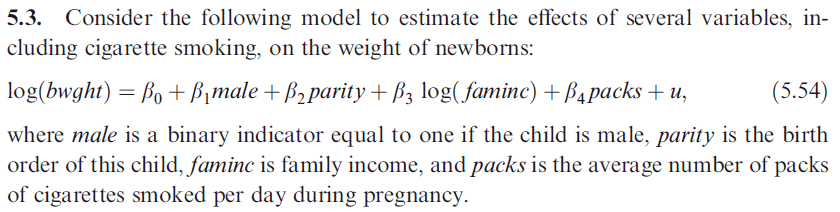
1. Write down the reduced variables in equation 5.53 with the exception of *exercise*. Write down the reduced form of exercise, and state the conditions under which the parameters of equation are identified.

The reduced form for *exercise* is

We need at least one of or to be different from zero for the parameters of equation are identified.

1. How can the identification assumption in part c be tested?

We can do an F test with . Ideally, we want to reject the null so that the parameters of equations are identified.



1. Why might you expect packs to be correlated with u?

There are unobserved factors in u that might be correlated with the number of packs of cigarette smoked and also determine the weight of newborns. For example, diet can be one of the unobserved factors because women who smoke more during pregnancy are likely to follow an unhealthy diet that may have an effect on the weight of newborns.

1. Discuss whether average cigarette price is likely to satisfy the properties of a good instrumental variable for packs.

In order for average cigarette price to be a good instrumental variable. It has to satisfy two conditions: relevance and exogeneity. Average cigarette price is relevant to *packs* since a higher average price is likely to reduce the number of packs smoked. However, average cigarette price might fail the exogeneity condition because

1. Use OLS, then 2SLS.

OLS output

lm(formula = log(bwght) ~ male + parity + log(faminc) + packs,

data = mydata)

Residuals:

Min 1Q Median 3Q Max

-1.63729 -0.08845 0.02034 0.12271 0.84409

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.675618 0.021881 213.681 < 2e-16 \*\*\*

male 0.026241 0.010089 2.601 0.00940 \*\*

parity 0.014729 0.005665 2.600 0.00942 \*\*

log(faminc) 0.018050 0.005584 3.233 0.00126 \*\*

packs -0.083728 0.017121 -4.890 1.12e-06 \*\*\*

2SLS output

ivreg(formula = log(bwght) ~ male + parity + log(faminc) + packs |

. - packs + cigprice, data = mydata)

Residuals:

Min 1Q Median 3Q Max

-2.19538 -0.06910 0.07829 0.19077 0.89686

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.467861 0.258829 17.262 <2e-16 \*\*\*

male 0.029821 0.017779 1.677 0.0937 .

parity -0.001239 0.021932 -0.056 0.9550

log(faminc) 0.063646 0.057013 1.116 0.2645

packs 0.797106 1.086275 0.734 0.4632

There seems to be a huge difference between OLS and 2SLS in the estimated effect of *packs* on *bwght*. With OLS, one more pack of cigarettes will reduce bwght by about 8.4%, holding other constant and the t-test is statistically significant on the coefficient. The 2SLS estimate on the coefficient of *packs* is not significant and it has opposite sign.

1. Estimate the reduced form for *packs*.

lm(formula = packs ~ male + parity + log(faminc) + cigprice,

data = mydata)

Residuals:

Min 1Q Median 3Q Max

-0.36386 -0.11365 -0.08285 -0.04761 2.36602

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.1374075 0.1040005 1.321 0.1866

male -0.0047261 0.0158539 -0.298 0.7657

parity 0.0181491 0.0088802 2.044 0.0412 \*

log(faminc) -0.0526374 0.0086991 -6.051 1.85e-09 \*\*\*

cigprice 0.0007770 0.0007763 1.001 0.3171

Now that *cigprice* is not statistically significant and the coefficient shows that *cigprice* has a positive effect on *packs* which is not expected. Therefore, it is reasonable to say that *cigprice* fails as an IV for *packs* because its correlation with *packs* is not statistically significant and it does not have an expected sign with regards to *packs*.