

ECN 301

Econometric Methods

Arjo Klamer, 2020

1 Question 1

In this exercise we will use data from Terza (2002) to investigate if abuse of alcohol has any impact on the employment status of men. This exercise is adapted from computer exercise 17.C15 in Wooldridge (2020).

```
-----
Contains data from ./alcohol.dta
-----
      obs:              9,822
-----
variable name    variable label
-----
abuse            =1 if abuse alcohol
status           out of workforce = 1;
                  unemployed = 2,
                  employed = 3
unemrate         state unemployment rate
age              age in years
educ             years of schooling
married          =1 if married
famsize          family size
white            =1 if white
exhealth         =1 if in excellent health
vghealth         =1 if in very good health
goodhealth       =1 if in good health
fairhealth       =1 if in fair health
northeast        =1 if live in northeast
midwest          =1 if live in midwest
south            =1 if live in south
centcity         =1 if live in central city of MSA
outercity        =1 if in outer city of MSA
qrt1             =1 if interviewed in first quarter
qrt2             =1 if interviewed in second quarter
qrt3             =1 if interviewed in third quarter
beertax          state excise tax, $ per gallon
cigtax           state cigarette tax, cents per pack
ethanol          state per-capita ethanol consumption
mothalc          =1 if mother an alcoholic
fathalc          =1 if father an alcoholic
livealc          =1 if lived with alcoholic
inwf             =1 if status > 1
employ           =1 if employed
-----
```

1. What fraction of the sample is employed at the time these men were interviewed? What fraction of the sample has abused alcohol.
2. Estimate a linear regression model for `employ` with the following variables as covariates: `abuse`, `age`, `agesq`, `educ`, `educsq`, `married`, `famsize`, `white`, `northeast`, `midwest`, `south`, `centcity`, `outercity`, `qrt1`, `qrt2` and `qrt3`. Use heteroskedasticity-robust standard errors.¹

¹The fact that the dependent variable is a binary variable does not violate our assumptions MLR.1–MLR.4. This is also called a linear probability model (LPM) and the errors will be heteroskedastic. We will return to this model in a later exercise.

3. The variable `abuse` might be endogenous in this setting. Argue why `mothalc` and `fathalc`, indicating whether a man's mother or father were alcoholics, respectively, could be reasonable instruments. Estimate the LPM using the GMM heteroskedasticity-robust instrumental variables model.
4. Test
 - (1) if the instruments are weak,
 - (2) if `abuse` is endogenous, and
 - (3) if the instruments are valid.
5. Compare the new parameter for `abuse` with the original parameter estimate, and conclude with respect to effect of alcohol abuse on labor market participation.
6. (Optional.) Are there other good predictors for the `abuse` variable? Use k -fold cross-validation to select the best predictor model. Using the chosen prediction specification in the GMM IV model does the parameter estimate for `abuse` change? (See, among other, Athey and Imbens (2017) for a discussion of big data and machine learning in prediction in the first stage equation.)

2 Question 2

This exercise draws heavily upon Hayashi (2000).

The relationship between the wage rate and schooling has been the subject of a large number of empirical and theoretical investigations following the pioneering study by Mincer (1958). This attention may seem puzzling because the explanation of the positive relationship seems to be obvious: education enhances the individual's productivity.

There are, however, other explanations. In the job market signaling model of Spence (1973), more educated individuals receive higher wages only because education is used as a *signal of higher ability*. Although education does not increase the individual's earning capacity, there is a correlation between the wage rate and schooling because both variables are influenced by a third variable, *ability*. One of the earliest attempts to try to isolate the effect of education on the wage rate from that of ability was the study by Griliches (1976). Well-known later studies include, among others, Blackburn and Neumark (1992) and Card (2001).

In this exercise we will estimate the type of wage equation estimated by Griliches using data from the Young Men's Cohort of the National Longitudinal Survey (NLS-Y). This cohort was first surveyed in 1966 at ages 14–24. The dataset, in the file `nls80`, is an extract from the NLS-Y used by Blackburn and Neumark (1992). A special feature of this particular dataset is that it contains two measures of ability. One measures is the score on the Knowledge of the World of Work (`KWW`) test administered by the NLS interviewers in 1966. The other measure is the IQ score that is a composed measure of various test scores obtained from the youths' school records (from 1968).

The following variables are included in this dataset:

```
-----
http://athene.umb.no/emner/pub/ECN301/data/nls80.dta
-----
obs:          935
vars:         17
-----

variable name  variable label
-----
wage          monthly earnings
```

hours	average weekly hours
iq	IQ score
kww	knowledge of world work score
educ	years of education
exper	years of work experience
tenure	years with current employer
age	age in years
married	=1 if married
black	=1 if black
south	=1 if live in south
urban	=1 if live in SMSA
sibs	number of siblings
brthord	birth order
meduc	mother's education
feduc	father's education
lwage	log(wage)

The typical wage equation estimated in the literature is the semi-log form (Card, 1995):

$$\log(\text{wage}) = \beta_1 + \beta_2 \text{educ} + \beta_3 \text{ability} + \dots + u \quad (1)$$

where *wage* is the wage rate for an individual, *educ* is the schooling in years, *ability* is some measure of ability, in addition to a series of observable characteristics such as experience, tenure and location dummies.

We will be using the same subsample as Blackburn and Neumark (1992), i.e. without any black individuals and only those for whom we have information about mother's education. (Remove these observations and make sure your working dataset has $N = 758$.)

1. Calculate means and standard deviations of all the provided variables and prepare a summary table. Also, calculate the correlation between *IQ*, *KWW* and *educ*.
2. Consider a wage equation with *educ*, *exper*, *tenure*, *south*, and *urban* as explanatory variables. However, we do not have a variable that matches the theoretical construct of *ability*. Thus, the model will either have an omitted variable problem or problems with a potentially poorly measured proxy variable, namely *IQ*.

Estimate the model using OLS both with and without the variable *IQ*.

3. If we include *IQ* in the model there is a potential problem with measurement errors. We can use instrumental variables regression to deal with that problem.
 - (1) Estimate the model using 2SLS with *meduc*, *KWW*, and *age* serving as instruments for *IQ*. Report both the first stage and the second stage results. Discuss the validity of the instruments.
 - (2) Test for endogeneity of *IQ*.
 - (3) Test for overidentifying restrictions.
4. If we omit *IQ* from the model there is a potential problem with omitted variables. We can use instrumental variables regression to deal with that problem as well.
 - (1) Estimate the model using 2SLS using *meduc*, *KWW*, and *IQ* as instruments for *educ*. Report both the first stage and the second stage results. Discuss the validity of the instruments.
 - (2) Test for endogeneity of *educ*.
 - (3) Test for overidentifying restrictions.
5. Summarize and give an overall assessment of your estimates for return to schooling in light of your findings wrt omitted variables or poor proxy variables.

References

- ATHEY, S., AND G. W. IMBENS (2017): "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31(2), 3–32.
- BLACKBURN, M., AND D. NEUMARK (1992): "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics*, 107(4), 1421–1436.
- CARD, D. (1995): "Earnings, Schooling, and Ability Revisited," *Research in Labor Economics*, 14, 23–48.
- (2001): "Estimating the Return to Schooling: Progress on some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.
- GRILICHES, Z. (1976): "Wages of Very Young Men," *Journal of Political Economy*, 84, S69–S85.
- HAYASHI, F. (2000): *Econometrics*. Princeton University Press, Princeton, NJ.
- MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66(4), 281–302.
- SPENCE, A. M. (1973): "Job Market Signalling," *Quarterly Journal of Economics*, 87(3), 355–375.
- TERZA, J. V. (2002): "Alcohol Abuse and Employment: A Second Look," *Journal of Applied Econometrics*, 17, 393–404.
- WOOLDRIDGE, J. M. (2020): *Introductory Econometrics: A Modern Approach*. Cengage Learning, Mason, OH.