

# Exam D0M61A Advanced econometrics

19 January 2009, 9–12am

### Question 1 (5 pts.)

Consider the wage function

 $w_i = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta'_3 h_i + \varepsilon_i,$ 

where  $w_i$  is the log-wage of individual i,  $S_i$  is years of schooling,  $E_i$  is years of experience,  $h_i$  is a vector of other observed characteristics of i, and  $\varepsilon_i$  is an error term. Using labour market data from 1976 from the U.S. National Longitudinal Survey of Young Men (Card, 1995), the following equations were estimated by OLS, with standard errors in parentheses (2040 observations were used):

$$w_{i} = 4.8159 + 0.0825S_{i} + 0.0436E_{i} + \text{residual}$$
(1)  
(0.0824) (0.0046) (0.0028)  

$$R^{2} = 0.1450 \qquad s = 0.3864 \qquad SSR = 304.256$$
  

$$w_{i} = 4.8920 + 0.0732S_{i} + 0.0427E_{i}$$
(0.0813) (0.0045) (0.0027)  

$$- 0.171BLACK_{i} + 0.155SMSA_{i} - 0.086SOUTH_{i} + \text{residual}$$
(2)  
(0.024) (0.019) (0.018)  

$$R^{2} = 0.2093 \qquad s = 0.3719 \qquad SSR = 281.359$$

Here  $BLACK_i$ ,  $SMSA_i$  and  $SOUTH_i$  are dummy variables indicating whether the individual was black, lived in a metropolitan area and lived in the south.

1.  $(\frac{1}{2} \text{ pt.})$  Comment on the following assertion: "From a prediction perspective, where one wants to predict someone's wage given his level of schooling, experience and other observed characteristics, there is no point in worrying about endogeneity."





2.  $(\frac{1}{2} \text{ pt.})$  Give a precise interpretation of the coefficient estimate associated with  $S_i$  in (1).

3. (1 pt.) Assuming that the conditions for exact inference are satisfied, test the joint hypothesis that the coefficients associated with  $BLACK_i$ ,  $SMSA_i$  and  $SOUTH_i$  are zero. (For hypotheses tests, always formally state the null hypothesis, the alternative hypothesis, the test statistic, the chosen significance level, the critical value and the outcome. Choose the significance level to be as informative as possible or, even better, compute the *p*-value.)

4. (1 pt.) Comment on the following assertion: "The above estimates of  $\beta_1$  are not helpful for an individual who wants to estimate his *returns to schooling*, even when all individuals have the *same* returns to schooling." Start your comments with a sensible definition of the returns to schooling for any individual *i* (in terms of *i*'s wage only!).

 $\mathbf{2}$ 



Motivated by a desire to estimate the returns to schooling, equation (2) was reestimated with GMM, using the following instruments: a constant,  $BLACK_i$ ,  $SMSA_i$ ,  $SOUTH_i$ , and also  $AGE_i$  (age),  $KWW_i$  (score on Knowledge of the World of Work test),  $IQ_i$  (score on IQ test) and  $NEARCOL_i$  (dummy variable indicating whether *i* lived near a college in 1966). The results are as follows (with heteroskedasticity-robust standard errors in parentheses):

- $w_{i} = 4.4504 + 0.1054S_{i} + 0.0425E_{i}$ (0.1239) (0.0081) (0.0027)  $- 0.1372BLACK_{i} + 0.1405SMSA_{i} - 0.0834SOUTH_{i} + \text{residual} \quad (3)$ (0.0260) (0.0196) (0.0187)  $R^{2} = 0.1792 \qquad s = 0.3789 \qquad SSR = 292.069$
- 5.  $(\frac{1}{2} \text{ pt.})$  The estimate of the returns to schooling has increased. Which critical assumption related to the effect of ability do we have to make for this estimate to be reliable? What is the likely sign of the bias if this assumption fails to be true?

- 6.  $(\frac{1}{2} \text{ pt.})$  Why is it that, next to  $S_i$ , also  $E_i$  is considered to be endogenous here?
- 7.  $(\frac{1}{2} \text{ pt.})$  What is the motivation behind the use of  $NEARCOL_i$  as an instrument?
- 8.  $(\frac{1}{2} \text{ pt.})$  The *J*-statistic equals 7.958. Test the hypothesis that the instruments are jointly valid.



#### Question 2 (3 pts.)

Cameron and Trivedi (*Microeconometrics: Methods and Applications*, Cambridge University Press, 2005) write on p. 68:

"If this conditional mean is linear in x, so that  $E[y|x] = x'\beta$ , the parameter  $\beta$  has a structural or causal interpretation [...] This permits meaningful policy analysis of effects of changes in regressors on the conditional mean."

Comment on this statement.

#### Question 3 (6 pts.)

Consider a random sample of n workers, drawn from the population of workers who become unemployed in a given country, in a given period (say, January 2008). Let  $y_i$  be the length of worker *i*'s unemployment spell, i.e. the time elapsed before *i* finds a new job. We seek to explain  $y_i$  by *i*'s characteristics, denoted as  $x_i$  (a vector). Assume  $y_i \ge 0$  is a random variable whose conditional density, given  $x_i$ , is

 $f(y_i|x_i;\lambda_i) = \lambda_i \exp(-\lambda_i y_i), \quad \text{where } \lambda_i = \exp(\beta_0 + \beta' x_i)$ 

and  $\beta_0$  and  $\beta$  (a vector) are parameters to estimate.

1. (1 point) Suppose that one of the variables in  $x_i$  is the amount of education *i* has taken (expressed in years, say). What is the likely sign of the parameter (say  $\beta_j$ ) associated with this variable? Motivate your answer.

4



2. (2 points) Given the observations  $(y_1, x_1), ..., (y_n, x_n)$ , construct the loglikelihood function of  $(\beta_0, \beta)$ .

- 3. (2 points) Consider the following complication. Suppose that for some workers in the sample, the observed unemployment spell is incomplete due to the fact that these workers have not found a new job by the time the observation period ends (say, 31 December 2008). To account for this, let  $y_i$  be *i*'s complete (but possibly unobserved) unemployment spell, and let  $z_i$  be the observed, possibly incomplete, unemployment spell. Here  $z_i$  is related to  $y_i$  by
  - $z_i = y_i$  if i's unemployment spell is complete

 $z_i < y_i$  if *i*'s unemployment spell is incomplete

and we know whether or not  $z_i$  is a complete or an incomplete unemployment spell. That is, we observe the variable  $d_i$  defined as

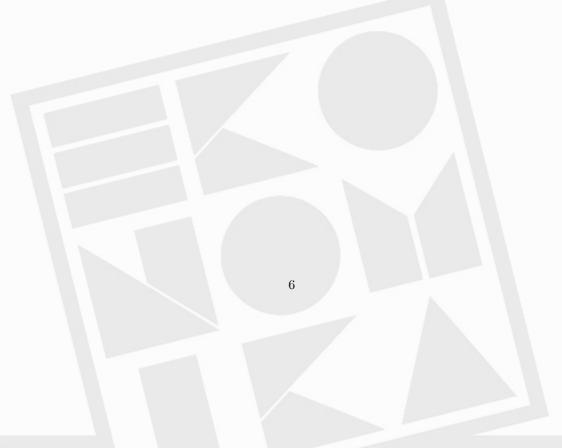
> $d_i = 0$  if *i*'s unemployment spell is complete  $d_i = 1$  if *i*'s unemployment spell is incomplete.

Given the observations  $(z_1, x_1, d_1), ..., (z_n, x_n, d_n)$ , construct the log-likelihood function of  $(\beta_0, \beta)$ .

5



4. (1 point) Suppose n = 1 and suppose there are no covariates, so we only observe  $(z_i, d_i)$  for a single individual *i*. If  $d_i = 1$ , what is the maximum likelihood estimate of *i*'s expected unemployment duration? [You may want to reflect on the following equivalent question. Suppose you execute a command on your computer, without the slightest idea of how much time it will take before execution stops. A natural model for the waiting time before execution stops is the exponential distribution (recall it is memoryless). Suppose you wait for 30 seconds and execution hasn't stopped. What is the maximum likelihood estimate of the expected waiting time? What is it after only waiting for 0.5 seconds?]



Dekenstraat 2, 3000 Leuven - info@ekonomika.be



## Question 4 (6 pts.)

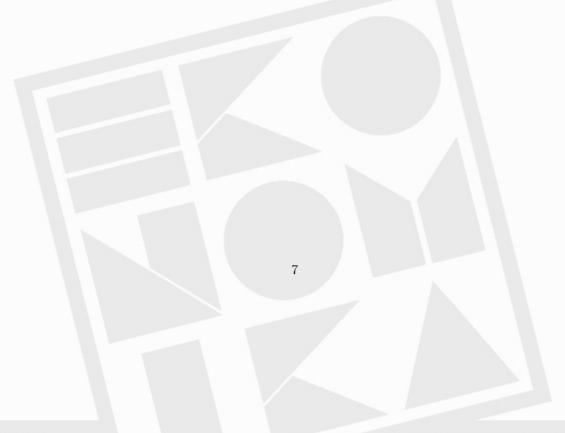
1. (3 points) Suppose you wish to estimate the population correlation between X and Y, defined as

$$r = \frac{E\left[(X - E(X))(Y - E(Y))\right]}{\sqrt{E(X - E(X))^{2}E(Y - E(Y))^{2}}}$$

Suppose you only have 10 observations,  $(X_1, Y_1), ..., (X_{10}, Y_{10})$ . It is well known that the sample correlation,

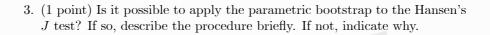
$$\widehat{r} = \frac{\sum_{i=1}^{10} \left(X_i - \overline{X}\right) \left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{10} \left(X_i - \overline{X}\right)^2 \sum_{i=1}^n \left(Y_i - \overline{Y}\right)^2}}$$

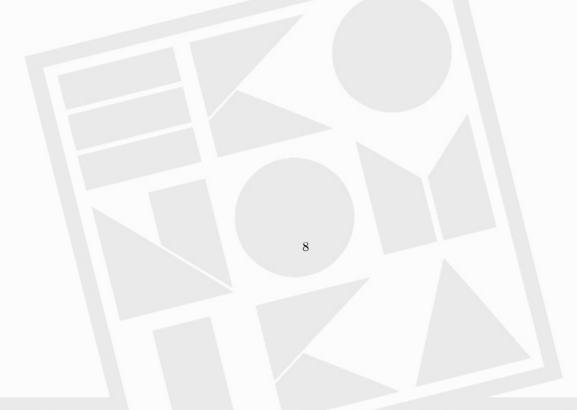
is a biased estimator of r. Can you use the bootstrap to reduce the bias of  $\hat{r}$ ? Describe your procedure in detail. How do you solve the (potential) problem that a sample correlation is only defined when the denominator is non-zero?





2. (2 points) Describe how you obtain a bootstrap 95% confidence interval for the IV estimator  $\hat{\delta} = (X'Z)^{-1}X'y$  in the just-identified case (using the notation of the course notes).





Dekenstraat 2, 3000 Leuven - info@ekonomika.be