Exam #2

Economics 435: Quantitative Methods

Spring 2006

1 Ugly criminals

A recent paper by Mocan and Tekin ("Ugly criminals," NBER working paper #12019) describes its findings as follows:

> "Using data from three waves of Add Health[1] we find that being very attractive reduces a young adult's (ages 18-26) propensity for criminal activity and being unattractive increases it for a number of crimes, ranging from burglary to selling drugs."

For the purposes of this problem we will simplify the study quite a bit. Our outcome variable is:

$$crime \equiv \text{number of crimes committed by the respondent between 1994 and 2000.}$$

Our primary explanatory variable is:

$$ugly \equiv \begin{cases} 1 & \text{if respondent is described by the interviewer in 1994 as "unattractive" or "very unattractive"} \\ 0 & \text{if respondent is described by the interviewer in 1994 as "attractive" or "very attractive"} \end{cases}$$

Let x be some additional explanatory variable that affects a person's propensity for engaging in criminal activity, and let our structural model be:

$$crime = \beta_0 + \beta_1 ugly + \beta_2 x + u$$
$$E(u|ugly, x) = 0$$

where β_1 is the effect of unattractiveness on criminal activity. We will assume for the entirety of this question that this structural model is correct. We will also assume that we have a random sample of data on $(crime_i, ugly_i)$ for $i = 1, 2, \ldots, n$, and that $ugly_i$ varies in the data.

a) Suppose that we estimate a simple linear regression of $crime$ on $ugly$. Let $\hat{\beta}_1$ be the OLS coefficient from our regression:

$$\hat{\beta}_1 = \frac{c\hat{o}v(crime, ugly)}{v\hat{a}r(ugly)}$$

Suppose that the omitted variable x is parental income. Provide a reasonable guess on the likely sign of β_2 and the likely sign of $cov(x, ugly)$. Also provide an argument for why your guess is correct.

b) Assuming your guesses in part (a) of this question are correct will, do you expect that plim $(\hat{\beta}_1)$ be greater than β_1 or less than β_1?

---

[1] "Add Health" is just a commonly used abbreviation for a panel data survey known officially as *The National Longitudinal Study of Adolescent Health*. "Longitudinal" is another word for panel data, and "three waves" means that we have data on each student in 1994, 1996, and 2000.

c) Suppose that attractiveness has no causal effect on subsequent criminality, i.e., $\beta_1 = 0$. Based on your result in (b), are we more likely to conclude that being unattractive increases the likelihood of committing crimes (as Mocan and Tekin conclude), or to conclude that it decreases the likelihood of committing crimes?

d) Suppose that we have no omitted variables (i.e., $\beta_2 = 0$), but may have measured attractiveness imperfectly. Let "true" unattractiveness is defined as:

$$trulyUgly \equiv \begin{cases} 1 & \text{if respondent would be described by a majority of Canadians} \\ & \text{in 1994 as "unattractive" or "very unattractive"} \\ 0 & \text{if respondent would be described by a majority of Canadians} \\ & \text{in 1994 as "attractive" or "very attractive"} \end{cases}$$

Also suppose that the interviewer's classification of respondents as attractive or unattractive is partly a matter of personal taste. To be exact, suppose that each interviewer randomly classifies some attractive people as unattractive, and randomly classifies some unattractive people as attractive.

Under these specific assumptions, will an OLS regression of *crime* on $(ugly, x)$ tend to lead researchers to conclude a positive effect of (true) unattractiveness on crime even if there is no effect? Supposing that there really is a causal effect leading from unattractiveness to a life of crime (i.e., $\beta_1 > 0$) will an OLS regression of *crime* on $(ugly, x)$ tend to lead researchers to overestimate or underestimate the effect of (true) unattractiveness on crime?

2   Peer effects

Suppose that we have school-level data on exam scores (specifically the Foundation Skills Assessment or FSA exam) for BC Grade 4 students from 2001-2003. We will use i to indicate the school, and t to indicate the year ($t = 1$ in 2001, $t = 2$ in 2002, etc.). The relevant variables are:

$$y_{it} = \text{average score of low-income students in school } i \text{ in year } t \text{ on the FSA exam}$$
$$x_{it} = \text{proportion of Grade 4 students in school } i \text{ in year } t \text{ that are low-income}$$

We are interested in measuring the impact of schoolmate[2] characteristics (in this case, income) on a student's own academic performance.

Our "basic" structural model is:

$$y_{it} = \alpha_i + \beta_1 x_{it} + u_{it}$$
$$E(u_{it}|x_{i1}, x_{i2}, x_{i3}, \alpha_i) = 0$$

where α_i is a school-specific fixed effect (which we could call "school quality")

Our "enhanced" structural model is:

$$y_{it} = \alpha_i + \delta_i t + \beta_1 x_{it} + u_{it}$$
$$E(u_{it}|x_{i1}, x_{i2}, x_{i3}, \alpha_i, \delta_i) = 0$$

where school quality is now $\alpha_i + \delta_i t$. Note that our enhanced structural model allows for each school's quality level to have a time trend.

---

[2]Note that our outcome variable is the average score of low-income students rather than the average score of students in general. This helps to separate the effect of having poor classmates from the effect of *being* poor.

Consider the following estimators of β_1

$$
\begin{aligned}
\hat{\beta}_1^{OLS} &\equiv \frac{\hat{cov}(x_{it}, y_{it})}{\hat{var}(x_{it})} \\
\hat{\beta}_1^{FE1} &\equiv \frac{\hat{cov}(x_{it} - x_{i1}, y_{it} - y_{i1})}{\hat{var}(x_{it} - x_{i1})} \\
\hat{\beta}_1^{FE2} &\equiv \frac{\hat{cov}(\tilde{x}_i, \tilde{y}_i)}{\hat{var}(\tilde{x}_i)}
\end{aligned}
\tag{1}
$$

where:

$$
\begin{aligned}
\tilde{x}_i &\equiv x_{i,1} + x_{i,3} - 2x_{i,2} \\
\tilde{y}_i &\equiv y_{i,1} + y_{i,3} - 2y_{i,2}
\end{aligned}
$$

a) Suppose that parents value school quality, and they are willing to spend money (for example, by buying houses in areas with good schools) to obtain it. Does this imply that the correlation between school quality and x_{it} is likely to be positive or negative?

b) How many years of data do we need in order to calculate $\hat{\beta}_1^{OLS}$? To calculate $\hat{\beta}_1^{FE1}$? To calculate $\hat{\beta}_1^{FE2}$?

c) Suppose that the "basic" model applies. Find the probability limits of each of the three estimators of β_1, and identify which of them are consistent.

d) Suppose that the "enhanced" model applies. Find the probability limits of each of the three estimators of β_1, and identify which of them are consistent.