

Exam #2

Economics 435

Spring 2004

1 A few issues in research design

These questions are very long, but the correct answers will be short.

a) Suppose you have a panel data set on Canadian young people starting with age 12 and going through age 30. You are interested in estimating the effect of nutritional intake at age 12 on performance in high school as measured by the grade 12 provincial Math exam score. However, you are concerned that there are unobserved family characteristics which tend to affect both nutrition when young and later performance in school. Can you use a fixed effects regression to consistently estimate the effect of nutrition on performance in the presence of these unobserved family characteristics? If so, briefly explain how you would do so. If not, explain why you cannot.

b) Suppose you have a cross-sectional data set on youth smoking rates by state. You are interested in estimating the effects of excise taxes and vending machine restrictions¹ on youth smoking. Your candidate regression function looks like this:

$$smoke_i = \beta_0 + \beta_1 * tax_i + \beta_2 * vend_i + u_i$$

where $smoke_i$ is the smoking rate in state i , tax_i is the excise tax rate on cigarettes, and $vend_i$ is a binary variable which equals one if the state bans the sale of cigarettes through vending machines and zero if it does not.

You are concerned that both taxes and vending machine restrictions are correlated with unobserved state factors, so a simple OLS regression will produce inconsistent estimates of the true effect. Fortunately, you have a candidate instrumental variable. The year before your data was gathered, the U.S. government passed the “No Youth Smoking by 2010” law. This law² set guidelines for state-level antismoking policy and financial rewards for meeting them. Specifically, an eligible state would receive \$100 per resident if they banned the sale of cigarettes through vending machines and raised their excise tax rate to \$3 per pack (which is a higher tax rate than any state had previously). The eligibility

¹By excise tax, I mean that there is a fixed per-pack tax, as opposed to a tax which is proportional to the sales price. By “vending machine restrictions” I mean laws that ban the sale of cigarettes through vending machines.

²This law does not exist; I just made it up.

rules of the program were unusual: 25 of the 50 states were randomly selected to be eligible. The rest were not eligible. This was done partly to save money, and partly to gather useful data for people like me.

Let z_i be a binary variable equaling one if state i was eligible for the program, and zero if not. Can you use z_i as an instrumental variable to consistently estimate β_1 and β_2 ? If so, briefly explain how you would do so. If not, explain why not.

2 Time series and stationarity

Let x_t and y_t be two covariance stationary stochastic processes which are independent of one another.

- a) Prove that ax_t , where a is any constant, is covariance stationary.
- b) Prove that $x_t + y_t$ is covariance stationary.
- c) Suppose that x_t is an AR(1) process

$$x_t = \epsilon_t + \rho x_{t-1}$$

where ϵ_t is white noise.

Calculate the impulse response function:

$$irf(k) = \frac{\partial x_{t+k}}{\partial \epsilon_t}$$

- d) Plot the impulse response function for $k = 0, 1, 2, 3, 4$ in the previous problem for $\rho = 0.5$, $\rho = 0$, $\rho = 1$, $\rho = 1.5$, $\rho = -0.5$, and $\rho = -1.5$. In other words there should be six separate plots.

3 Fixed effects with measurement error

Consider the following model of fixed effects with measurement error.

Suppose that we have a panel data set of observations on y_{it} and \tilde{x}_{it} (\tilde{x}_{it} is just the value of x_{it} measured with error) and we have the model:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it} \tag{1}$$

$$x_{it} = x_i + v_{it} \tag{2}$$

$$\tilde{x}_{it} = x_{it} + \epsilon_{it} \tag{3}$$

where the unobserved variables $u_{it}, x_i, v_{it}, \epsilon_{it}$ are all independent of one another, across individuals, and across time. The unobserved variable α_i is independent of u_{it} , v_{it} , and ϵ_{it} but may be correlated with x_i

$$\begin{aligned} E(x_i) &= \mu_x \\ E(\alpha_i) &= \mu_\alpha \\ E(u_{it} | \alpha_i, x_i, v_{it}, \epsilon_{it}) &= 0 \end{aligned}$$

$$\begin{aligned}
E(v_{it}|\alpha_i, x_i, u_{it}, \epsilon_{it}) &= 0 \\
E(\epsilon_{it}|\alpha_i, x_i, v_{it}, u_{it}) &= 0 \\
\text{var}(\alpha_i) &= \sigma_\alpha^2 \\
\text{var}(x_i) &= \sigma_x^2 \\
\text{var}(v_{it}) &= \sigma_v^2 \\
\text{var}(\epsilon_{it}) &= \sigma_\epsilon^2 \\
\text{corr}(\alpha_i, x_i) &= \rho_{\alpha,x}
\end{aligned} \tag{4}$$

We are interested in estimating β , the effect of x_{it} on y_{it} . However, there are two problems with doing so. First, we believe that there are individual specific fixed effects (the α_i terms) and that they may be correlated with x_{it} . Second, we suspect that our data are measured with error.

Let the OLS estimator for β be defined as

$$\hat{\beta}^{OLS} \equiv \frac{\hat{c}ov(y_{it}, \tilde{x}_{it})}{\hat{v}ar(\tilde{x}_{it})}$$

Let the FE (fixed effects) estimator be defined as:

$$\hat{\beta}^{FE} \equiv \frac{\hat{c}ov(\Delta y_{it}, \Delta \tilde{x}_{it})}{\hat{v}ar(\Delta \tilde{x}_{it})}$$

- a) Find the probability limit of $\hat{\beta}^{OLS}$ in terms of the parameters:

$$(\mu_x, \mu_\alpha, \beta, \sigma_\alpha^2, \sigma_x^2, \sigma_u^2, \sigma_v^2, \sigma_\epsilon^2, \rho_{\alpha,x})$$

- b) Find the probability limit of $\hat{\beta}^{FE}$ in terms of these parameters.

c) Suppose (for this part of the question only) that there is no fixed effect, i.e., $\text{var}(\alpha_i) = 0$. Find each of the two probability limits. Which estimator produces a smaller asymptotic bias?

d) Suppose (for this part of the question only) that there is a fixed effect, i.e., $\text{var}(\alpha_i) > 0$, but that it is uncorrelated with x_i , i.e., $\rho_{\alpha,x} = 0$. Find each of the two probability limits. Which estimator produces a smaller asymptotic bias?

e) Suppose (for this part of the question only) that there is no measurement error, i.e., $\text{var}(\epsilon_{it}) = 0$. Find each of the two probability limits. Which estimator produces a smaller asymptotic bias?

f) Consider the following two statements (difference between the two is in **boldface**):

When estimating a model using panel data in which the explanatory variable is measured with error, a fixed effects estimator will be preferable (in terms of having smaller asymptotic bias) to a simple OLS estimator when the covariance between the fixed effect and the explanatory variables is **large** relative to the measurement error.

When estimating a model using panel data in which the explanatory variable is measured with error, a fixed effects estimator will be preferable (in terms of having smaller asymptotic bias) to a simple OLS estimator when the covariance between the fixed effect and the explanatory variables is **small** relative to the measurement error.

Based on your results, which of the two statements is correct?