Midterm Exam

Economics 522

28 October 2020

Submitting Your Answers:

Please submit your answers to me by the end of the day Sunday, November 1st. This is an open-book exam; you are free to use class notes and other resources. However, *please answer in your own words*: responses that are simply copied from other sources will not receive credit. Your answers can be provided in a LaTeX or Word document, or submitted as photos of what you've written down on paper (in this case, make sure that the full answer can be seen!).

Each question counts for one-third of the score for the exam as a whole. To be eligible for full credit, you must provide *explanations, in words* along with your mathematical calculations.

1. Consider a model in which workers i are randomly assigned (using a random number generator) to one of two job training programs. Having been given their random assignment, workers then choose whether to participate in the program to which they've been assigned, or instead, to participate in the other program. Let $D_i = 1$ if worker i chooses to participate in Program 1; and let $D_i = 0$ if she chooses to participate in Program 2. After the training ends, the level of worker i's wages is given by

$$Y_i = \alpha + \delta_i D_i + \epsilon_i$$

In this equation, δ_i is a worker-specific coefficient (a random variable), and $\operatorname{E}\delta_i$ is its expected value, representing the expected difference in wages that comes from participating in Program 1 rather than Program 2.

   (a) Show that the standard instrumental variables estimator, using random assignment as the instrument, does *not* consistently estimate $\operatorname{E}\delta_i$ in general. Are there any special cases in which $\operatorname{E}\delta_i$ can be consistently estimated?

   (b) What information about $\operatorname{E}\delta_i$ can be obtained from the (inconsistent) instrumental variables estimator? Please define any additional notation you use in your answer, and be specific about any additional assumptions you make.

2. A simple structural model of wages $Y_{i,t}$ for worker i at time t is

$$Y_{it} = \alpha + t \cdot \beta + D_{it}\delta + u_i + \epsilon_{it}.$$

This model contains a constant, a time trend, a dummy variable D_{it} which takes the value 1 if the worker has chosen to participate in a training program on or before time t, and an error-components disturbance term. You have two time-points of data for each worker: Time $t = 0$ denotes the period before the program was offered; time $t = 1$ denotes the period after the program was offered.

If we think of the u_i component as representing the worker's motivation, we would suspect that more motivated workers (those with higher values of u_i and also higher wages, other things being equal) are the kind of people who are more likely to take advantage of opportunities for job training. That is, it seems likely that u_i and D_{it} will be positively correlated.

(a) Discuss how to estimate the program effect δ by the method of difference-in-differences. What are the important assumptions of this method?

(b) How would you estimate the standard errors of the limiting distribution of $\sqrt{n}(\hat{\delta} - \delta)$ so that the estimates are robust to heteroskedasticity and serial correlation in ϵ_{it}?

(c) Suppose the model is altered to $Y_{it} = \mathbf{X}'_{i,t}\gamma + D_{it}\delta + u_i + \epsilon_{it}$. Discuss how to estimate δ, the program effect, for this model.

3. Consider a nonlinear model $Y_i = \phi(\mathbf{X}_i, \theta) + \epsilon_i$ for which you know the functional form of $\phi()$ but not the true value of the θ parameter. The θ vector is of dimension k and we let θ_0 denote its true value. You can assume that the ϕ function is differentiable in θ.

   The estimation problem is that $\mathrm{E}(\epsilon_i|\mathbf{X}_i) \neq 0$, that is, one or more of the $\mathbf{X}_i$ covariates is correlated with the disturbance term. Luckily, you have k valid instruments $\mathbf{Z}_{i,1}$ and have an additional $m - k$ variables $\mathbf{Z}_{i,2}$ that may also be valid instruments. All of these instruments and potential instruments are contained in $\mathbf{Z}_i = [\mathbf{Z}_{i,1}, \mathbf{Z}_{i,2}]'$, a vector of total length m.

   The data series $\{(\mathbf{X}_i, \mathbf{Z}_i, \epsilon_i)\}$ is independent (over i) but not necessarily identically distributed (that is, the series is inid).

   Please answer the following

   (a) How would you estimate θ using the Generalized Method of Moments approach?

   (b) What are the first-order conditions for $\hat{\theta}$, the GMM estimator?

   (c) Assuming that all of the $\mathbf{Z}_i$ variables are valid instruments, what is the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$?

   (d) How would you estimate the variance matrix of the limiting distribution?

   (e) How would you test the validity of the instrumental variables?