POPULATION FIELD EXAM

2016 – 2017

Instructions

- You have 4 hours to complete this exam.
- This is a closed book examination. No written materials are allowed.
- You can use a calculator.
- There are 3 parts in this exam: Part I covers Lleras-Muney's course, Part II covers Edward Kung's course, and Part III covers Youssef Benzarti's course.
- Please answer each part in separate booklets.
- You may choose to answer two out of three parts. **75% is required in each part to receive a passing grade in the field exam.**

Part I: Population Economics (Fall 2016)

Professor Adriana Lleras-Muney

TOTAL: 100 POINTS

Answer both questions. Partial credit will be given whenever possible.

A. **(25 points) Review. Please answer the following questions:**

1. (5 points) Income and health: What is the evidence on the relationship between income and health/mortality? What are the policy implications of this debate?

2. (5 points) Explain what the synthetic cohort approach is and how it overcomes the limitations of the standard DD approach.

3. (5 points) Value of life: Why do economists argue it is important to estimate the value of life? Discuss the main approaches used to estimate the value of life

4. (5 points) Randomized experiments: What are the main advantages of randomized experiments? What are their limitations?

5. (5 points) Education and health are positively associated in many data sets. What are the possible reasons for the association, and what is the evidence in favor of each?

B. **Empirical applications**

1-(25 points) Empirical Application 1: Cutler and Miller (2005)

To estimate the effect of water cleaning technologies on mortality, Cutler and Miller estimate the following equation:

$$\ln(m_{c,t}) = \alpha + \beta_1 Filter_{c,t} + \beta_2 Chlorine_{c,t} + \beta_3 (Filter_{c,t} \times Chlorine_{c,t}) + \delta_c + \mu_t + \gamma_c l_{c,t} + \sum \rho_d d_{d,c,t} + \sum \lambda_k \ln(m_{c,t-k}) + \varepsilon_{c,t} .$$

Where the dependent variable is the log of mortality in city c and year t, Chlorine and Filter are dummies equal to one if the technology has been implemented in a given city and year. This equation also includes city dummies, year dummies, city specific trends and city-level demographics. Finally the control for lagged mortality rates.

a) (5 points) Why log mortality, is that the best functional form? How do we interpret coefficients?
b) (5 points) This specification does not contain a post*treated dummy. Explain why this is a DD. Why replace post with year dummies? Why have a dummy for each city?

c) (5 points) If we exclude city-level trends, what is the identifying assumption needed for the estimates to be unbiased?
d) (5 points) Why include other covariates? How should one chose them? Why add a city specific trend?
e) (5 points) What happens if the assumption that the timing is "exogenous" fails? How do could you check for this possibility?

2-(25 points) Empirical application 2: Card et al (2008)

To estimate the effect of having Medicare insurance on health care utilization and health outcomes Card et al estimate the following regression:

$$y_{ija} = X_{ija}\alpha_j + f_j(a) + C_{ija}\delta_j + u_{ija}$$

Where the dependent variable is a utilization (or health) measure for individual i in group j at age a, and C is dummy equal to one after age 65 (since individuals become eligible for Medicare at that age).

a) (5 points) What is the forcing variable in this regression discontinuity design? Are there any issues related to the fact that this variable is age/time? Is it an issue that age is measured in years?
b) (5 points) Is this a fuzzy or sharp design? Define both.
c) (5 points) What is the key indentifying assumption?
d) (5 points) What is the connection between RD and randomized experiments in general?
e) (5 points) What are the standard checks empirical researchers conduct to assess the robustness of these designs?

4-(25 points) Empirical application 3: Miguel and Kremer (2004)

To estimate the health externalities associated with deworming Miguel and Kremer estimate the following equation:

$$(1) \qquad Y_{ijt} = a + \beta_1 \cdot T_{1it} + \beta_2 \cdot T_{2it} + X'_{ijt}\delta + \sum_d (\gamma_d \cdot N^T_{dit}) + \sum_d (\phi_d \cdot N_{dit}) + u_i + e_{ijt}.$$

Where Y is the outcome for child i in school j at time t, T is a dummy equal to one is the school is treated at a given time, and X is a set of individual characteristics. The equation includes the number of kids treated within radius d of a given school (N^T), for different radii d, and controls for the total number of kids within a given radius (N).

a) (5 points) What assumptions are needed to estimate the externalities associated with treatment across schools? In the Manksi (1995) framework, how is the identification problem being solved?

b) (5 points) Using this equation, how do we compute the total externalities across schools? How do we compute the total intent-to-treat effects?
c) (5 points) There are also possibly externalities within schools. How could you estimate those?
d) (5 points) Explain what the effects of externalities would be on estimates of the effects of deworming from a randomized trial, if we were to simply compare the mean outcomes for treated and untreated schools.
e) (5 points) The paper finds large effects on school absenteeism but no effects on cognitive scores. How could we explain this? Does this matter for interpreting the results and for doing cost benefit analyses of the benefits of deworming?

Part II: Housing Markets and the Evaluation of Spatial Amenities (Winter 2017)

1. (40 pts) Answer the following questions in a few paragraphs. Be sure to cite relevant papers where appropriate.

   (a) Economists are often interested in estimating the value of amenities that are not explicitly traded in a marketplace, such as the value of neighborhood safety or of public education. Briefly describe how economists have attempted to do this. (10 pts)

   (b) When constructing quality of life rankings, Los Angeles often comes out near the top. Describe the intuition behind quality of life rankings and explain in a few short sentences why Los Angeles is often highly ranked. (10 pts)

   (c) Set up and describe Rosen's 1974 hedonic model (focus only on the consumer side). What can we learn from the equilibrium conditions of the model? What is Rosen's "two-step" method and what are some of its identification challenges? (10 pts)

   (d) Describe the user-cost model of house prices. What does the equilibrium relationship between house prices and rents depend on? What is missing from the user cost model? (10 pts)

2. (40 pts) Suppose you have a cross-section of data on housing transactions. The data allows you to see the street address of the home, the sale price, and the characteristics of the home, and neighborhood characteristics. You also have data on public schools, including each school's test score performance and attendance boundaries. You are interested in using this data to estimate the willingness to pay for school quality. (40 pts)

   (a) Suppose you regressed sale price on house characteristics and test scores for the school that house is assigned to. What would you expect to find? Describe some problems with this regression. (10 pts)

   (b) Come up with an alternative strategy for estimating the average willingness to pay for school quality. Make sure to write down the estimating equation. If you are deriving your identification strategy from another paper, cite that paper. How do you expect the new strategy to affect your estimates from part (a)? (10 pts)

(c) Describe some of the assumptions necessary for your answer in part (b) to be valid. How might you test these assumptions with the data that you already have? Feel free to draw the hypothetical pictures/tables you would use to test these assumptions. (10 pts)

(d) Are there remaining concerns that are not addressed in your answer to part (c)? What additional data would you need? (10 pts)

3. (40 pts) Consider a housing market with consumers $i = 1, \ldots, N$ and houses $j = 1, \ldots, N$. Each consumer chooses one house and each house gets chosen by one consumer. The utility that consumer i gets from choosing house j is:

$$V_{ij} = \alpha_i x_j - \beta_i p_j + \xi_j + \epsilon_{ij}$$

where x_j is an observed (to the econometrician) house characteristic, p_j is the price of house j, ξ_j is unobserved house quality, and ϵ_{ij} is a preference shock that is iid type 1 extreme value. α_i and β_i are preference parameters given by:

$$\begin{aligned} \alpha_i &= \alpha_0 + \sum_{k=1}^{K} \alpha_k z_{i,k} \\ \beta_i &= \beta_0 + \sum_{k=1}^{K} \beta_k z_{i,k} \end{aligned}$$

where z_{ik} are demographic characteristics of consumer i. As the econometrician, you observe x_j, p_j, z_{ik}, and d_{ij}—an indicator for whether consumer i chose house j.

(a) What is the probability that person i chooses house j? (10 pts)

(b) Let us write $\delta_j = \alpha_0 x_j - \beta_0 p_j + \xi_j$. Describe a computationally tractable strategy for estimating δ_j for each house. (10 pts)

(c) Describe a two-step procedure for estimating all the preference parameters, $\alpha_{0:K}$, $\beta_{0:K}$. (10 pts)

(d) Suppose one were to regress p_j on x_j. Interpret the coefficient on this regression. Under what conditions will it correctly estimate the average willingness to pay for x? When will it not? (10 pts)

Population Field Exam 2017 Youssef Benzarti

Econ 253A

Let us assume (for simplicity) that the current income tax has the following simple structure. Married couples can deduct $30,000 from their incomes, and singles (single or widowed adults or single parents) can deduct $15,000. Income less this deduction is called taxable income. Taxable income faces a marginal tax rate of 20% up to $100,000, and of 30% above $100,000.

a. Explain why such an income tax system generates a marriage penalty, and how it can potentially discourage labor supply of secondary earners.

Suppose that the government decides to eliminate the marriage penalty by taxing all incomes at the individual level according to the tax schedule for the singles described above (that is, married individuals would be taxed separately according to their own income).

b. Assuming no behavioral responses to this tax change, describe which couples would lose and which couples would gain from the reform. Do you think that the new tax system will be more or less redistributive that the previous one? Is the government likely to gain or to lose tax revenue from this reform (assuming again no behavioral responses)? [be as precise as you can about the statistics you would need to answer quantitatively]

c. Suppose now that the reform affects labor supply but not marriage decisions. Explain which labor supply responses this reform is likely to generate. Suppose you have cross-sectional data on family structure and incomes (for each member of the couple for married households) before and after the reform, and that male labor supply is not affected by the reform. Describe a simple differences-in- differences

methodology that would allow you to estimate the labor supply elasticity of married women with respect to the net-of-tax rate (one minus the marginal tax rate). What are the potential biases of the method you just described and how would you check the validity of the method if more data were available?

d. Suppose you now want to distinguish between the extensive labor supply responses (participation in the labor market) and the intensive labor supply responses (level of earnings on the job). Explain how cross-sectional data (before and after the reform as in question c. could allow you to estimate the extensive elasticity of labor supply of married women. Explain why, in the presence of extensive labor supply responses, cross-sectional data does not allow the estimation of intensive labor supply responses.

e. Suppose you now have access to panel data following the same individuals before and after the reform. How would you use the data to estimate the intensive labor supply elasticity of married women?

f. Suppose now that marriage decisions are potentially affected by the reform. Explain how you could use panel data before and after the reform to assess the effects on the tax reform on marriage decisions.