

Problem Set 1, Economics 272C, Professor Bitler

Due date: Wednesday February 5.

Problem 1. Welfare magnets

This problem is based on

McKinnish, T. “Importing the Poor: Welfare Magnetism and Cross State Migration.” *Journal of Human Resources* **40**:1, 2005, 57–76. As we may briefly discuss in class, the question of interest is whether or not people are induced to move by high welfare benefits.

The premise behind the McKinnish paper is that individuals living near state borders face lower costs of migration to adjacent states than do individuals living in interior counties. Thus, if the nearest neighbor state is a high benefit state, and my state is a low benefit state, my border counties should have lower welfare participation relative to my interior counties, and that state should have higher participation in border counties than interior counties. This leads to the following specification:

$$W_{cst} = \alpha_0 + \alpha_1 \cdot B_{cs} + \alpha_2 \cdot D_{cst} + \alpha_3 \cdot B_{cs} \cdot D_{cst} + \alpha_4 \cdot S_s + \alpha_5 \cdot \tau_t + \epsilon_{cst},$$

where W_{cst} is welfare expenditures per capita in my county, B_{cs} is an indicator for being on the border, D_{cst} is the difference in my neighbor’s benefits and my benefits, and S_s and τ_t are state and year dummies.

Download the compressed dataset `Welfmig2.dta.gz` from my website. This has the following variables in it at the county level for 1970–1990 from the McKinnish paper.

- `neardist`: Distance from my county centroid to nearest neighbor state border
- `state`: State FIPS code
- `year`: year
- `AFDCExp`: $\ln(\text{per capita county expenditures on AFDC})$
- `AFDCBen`: $\ln(\text{state monthly AFDC benefit level})$

- NeighborBen: $\ln(\text{neighbor state monthly AFDC benefit level})$

Create a variable that is the difference in benefits between the nearest neighbor state and my state. Estimate a cross sectional relationship of the early literature (regress expenditures on the difference in benefits). Cluster at the state level (why does this make sense)? Does this regression suggest that there is welfare migration?

- Add state fixed effects. What does the coefficient on the difference in benefits represent now?
- Add year fixed effects. What does the coefficient on the difference in benefits represent now? Why did it change?
- Is there within state variation in the difference in benefits ($NeighborBen - AFDCBen$) in 1970? In 1990?
- Create a dummy for being a border county as being a county whose centroid is within 30 miles of an adjacent state. Does every state have a border county? Which ones are excluded? What if the cutoff were 50 miles? 25 miles?

Create a variable that is the difference in neighbor state benefits minus my benefits, quantity interacted with being in a border county. Do so for the 30 and 50 and 25 mile border county definitions.

- Now consider McKinnish's main model above. Control for state FE, year FE, an indicator for being a border county within 30 miles of the border, the difference in benefits with the nearest state, and the interaction of border county being within 30 miles and the difference in benefits. What does the coefficient on the border county/difference in benefits interaction suggest? What variation is identifying it? Should we be concerned that the main border county coefficient is positive? Would we be concerned if the main border county coefficient were significant? Does it matter if you select a 50 mile border county definition? A 25 mile one?

- g. Replace the state and year fixed effects with state by year fixed effects. What does the coefficient on the border county/difference in benefits interaction suggest for the 30 mile definition? 50 mile definition? 25 mile definition? What variation is identifying it?
- h. Drop the states with no variation in border counties for each of the distances (this is the closest thing to the McKinnish specification in table 4, but without the county fixed effects). What does the coefficient on the border county/difference in benefits interaction suggest? Anything else to note about differences between this specification and table 4 of the published paper?
- i. Do one linear specification (that is, instead of the dependent variable being the log of benefits per capita, the dependent variable should be the level). Does the implied effect of a \$100 change in benefits change a lot from the log specification?

Problem 2. Income and consumption

This problem is based on Meyer, B. and J. Sullivan. “Changes in the Consumption, Income, and Well-Being of Single Mother Headed Families.” *American Economic Review* **98**:5, 2008, 2221–2241. This paper looks at changes in these variables for women likely to be at risk of being on welfare across time. We are going to use this data to look at some different relationships and also to help you learn about bootstrapping.

Download the compressed dataset `extract-ms.dta.gz` from the website. This is a subset of the final data set from the Consumer Expenditure Survey used in table 1 of the Meyer and Sullivan paper. The variables in the extract are as follows:

- `newid`: Id (recall there are 100 copies of the data for each observation, which should be identical except for the imputed housing and vehicle consumption)
- `period`: Time period, 93–95 or 97–00 or 01–03

- `finca_tax_needs2`: Post-tax income plus food stamps
- `finlwt`: Final weight for family get annual means)
- `tflow2_needs`: Total consumption (quarterly, so multiply by 4 to
- `main`: = 1 if in main single mom sample

Note that the reason for having 100 copies per observation is to allow you to bootstrap the variances for inference because the authors have imputed various components of consumption. Thus, both the within-decile averages and decile cutoffs are estimates. See Appendix 2 of NBER working paper 11976 for details. You should adjust your calculations for this.

If you need a brief tutorial on the bootstrap, see Professor Brownstone's JEP piece with Robert Valetta (Fall 2001), or see the discussion in the Cameron and Trivedi textbook.

- First, verify that there are 100 observations for each person. What share of people likely have imputed values for the components of consumption (use the weights)?
- Restricting your sample to the main sample used by the authors, calculate the deciles of the income and consumption distributions by time period (use weights). Replicate the point estimates in the top panel of table 1, and get confidence intervals by bootstrapping (you won't exactly match the CIs in the paper because they were calculated differently in the paper). (Remember that the income variable is annual but the consumption variable is quarterly.)
- Now we will look at the joint distribution of consumption and income in this sample over the entire period. Create a table that is 10 by 10, where the value in the j th row and i th column contains the share of the overall distribution that is in the j th decile of consumption and i th decile of income.
- Test whether income and consumption are independent for the sample of single mothers over the entire period.

e. Grouping deciles 5-10 together, create a 5 by 5 table for the ratio of consumption in 97–2000 to consumption in 93–95 by consumption and income decile. Do the same for the ratio of income in 97–00 to income in 93–95. Do you draw substantively different conclusions than Meyer and Sullivan drew from their table 1? (No need for confidence intervals.)

Problem 3. Taxes and labor supply

Hint: See the discussion of non-linear budget sets. Suppose the economy is made up of workers with identical preferences, but half have wages of \$10 per hour and half have wages of \$20 per hour. Imagine preferences are given by the following indirect utility function:

$v(w_n, y_{nli}) = \exp(-0.04 \cdot w_n) \cdot (y_{nli} - 12.5 \cdot w_n - 412.5)$, where w_n is net wages after taxes per hour, and y_{nli} is non-labor income.

Assume that non-labor income for everyone in the economy is zero.

Suppose the tax system is given by the following:

$$T(Y) = 0.10 \cdot Y, Y < 90$$

$$T(Y) = 9 + 0.5 \cdot (Y - 90), Y \geq 90.$$

- a. Using the indirect utility function, find the labor supply function for everyone in the economy. (This will be a function of the wage rate and non-labor income. It should be positive.)

- b. Find the optimal hours for persons with wages of \$10 per hour and and \$20 per hour.

- c. What is the total government revenue with this system?

- d. Suppose the administration changes, and that taxes are now 0 in the lower tax bracket. How does this change the budget set for each type?