

ECON 8686 Problem Set #1- Differences-in-Differences, Panel Data

This problem set is based on:

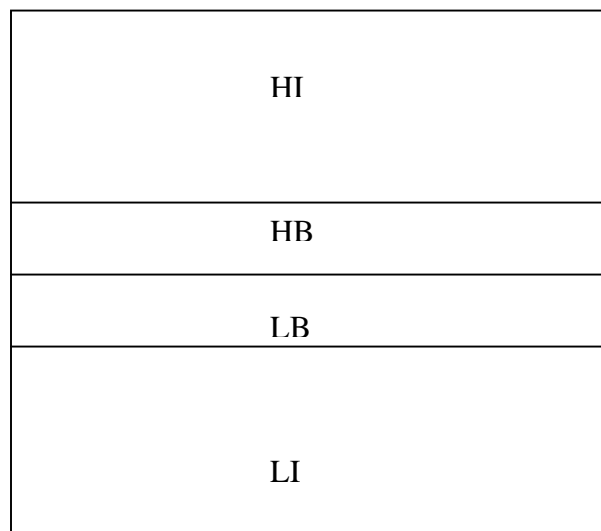
McKinnish, Terra. 2005. "Importing the Poor: Welfare Magnetism and Cross-Border Welfare Migration." *Journal of Human Resources* 40(1):57-76.

The research question is whether poor families in the U.S. migrate to states with higher welfare benefits. The main empirical difficulty is to show that migrants to high-benefit states are moving for welfare benefits, rather than other state amenities (such as strong labor markets) that tend to be positively correlated with welfare benefits.

Most studies of welfare migration, including this one, focus on the Aid to Families with Dependent Children (AFDC) program. AFDC was a welfare program that provided cash payments to low-income single mothers. (It was reformed and renamed Temporary Aid to Need Families (TANF) in 1996.) Even though it was a federal welfare program, states set their own benefit levels, generating sizeable differences in generosity across states. An important feature for the purpose of this paper is that benefit levels are set at the state level. They do not vary by county within a state.

Part I. Difference-in-Differences Strategy

Assume that the costs of between-state migration are lower for individuals located close to state borders. Consider the simple example for a country with two states illustrated below. The top state is the high-benefit state and the bottom state is the low-benefit state. Area HB contains the counties of the high-benefit state that border on the other state, and area LB is likewise defined for the low-benefit state. Areas HI and LI are the interiors of the two states. If the assumption of differential migration costs is correct, then, the border counties in area HB should disproportionately draw migrants from the border counties in area LB.



Welfare participation should therefore be higher in the border counties of the high-benefit state (H2) relative to the interior counties of the high-benefit state (HI), and welfare participation should be lower in the border counties of the low-benefit state (LB) relative to the interior counties of the low-benefit state (LI). Allowing W to be a measure of welfare participation, we can generate the following differences-in-differences formula:

$$(a) [\bar{W}_{HB} - \bar{W}_{HI}] - [\bar{W}_{LB} - \bar{W}_{LI}]$$

Both interior and border counties in high-benefit states should have higher welfare participation (because participation is more attractive in high benefit states), but the difference between border and interior counties should reflect the differential due to higher welfare migration into border counties. The within-state comparison nets out other state-level unobserved variables.

Consider the following regression model:

$$(b) W_{cs} = \alpha_0 + \alpha_1 B_{cs} + \alpha_2 BenDiff_{cs} + \alpha_3 B_{cs} * BenDiff_{cs} + \alpha_4 H_s + \varepsilon_{cs}$$

where B is an indicator variable that equals one if the county is on the state border, $BenDiff$ equals the neighbor's welfare benefit minus own state's welfare benefit (positive if the neighboring state has higher benefits). H is an indicator for the high benefit state.

- 1) If there is welfare-motivated migration, what should be the sign of α_3 ? Explain.
- 2) Evaluate $E[(\bar{W}_{HB} - \bar{W}_{HI}) - (\bar{W}_{LB} - \bar{W}_{LI})]$ as a function of the parameters and variables in equation (b). In other words, show the relationship between equations (a) and (b). (Hint: notice $BenDiff$ for the high-benefit state equals $-BenDiff$ for the low-benefit state)

Part II. Cross-Sectional Estimation

The data set `Welfmig1.dta` contains observations for all counties in the 48 continental states for 1990. The variables are:

`AFDCExp`: log of per capita AFDC expenditures in county

`AFDCBen`: monthly AFDC benefit in state (benefit for a family of 4, in 100s of dollars)

`Neardist`: Distance from county to the closest neighbor state

`NeighborBen`: monthly AFDC benefit in the closest neighboring state (in 100s of dollars)

`State`: State ID Code

`AFDCExp`, `Neardist`, and `NeighborBen` vary at the county level. `AFDCBen` and `State` vary only at the state level.

1) Estimate:

a) `reg AFDCExp AFDCBen`

b) `reg AFDCExp NeighborBen`

and

c) `reg AFDCExp AFDCBen NeighborBen`

Can the results of a, b or c be interpreted as evidence of welfare migration? Explain

2) Add state fixed-effects. First, use the command:

`tab state, gen(std)`

Then:

`reg AFDCExp AFDCBen NeighborBen std2-std48`

Do you obtain a coefficient estimate for AFDCBen? Should you?

(Be careful. STATA will very helpfully drop variables in the case of perfect collinearity, but this can fool you into thinking you are identified when you are not. You can confirm that the coefficient on AFDCBen is not identified in the state fixed-effects model with the alternative estimation:

```
xtreg AFDCExp AFDCBen NeighborBen, fe i(state)
```

Likewise, if you estimate

```
reg AFDCExp NeighborBen std2-std48
```

notice that STATA no longer drops any of the state indicators)

3) Create a variable that is the difference between neighbor state's benefits and own state's benefits:

```
gen BenDiff=NeighborBen-AFDCBen
```

Now compare:

```
reg AFDCExp NeighborBen std*
```

to:

```
reg AFDCExp BenDiff std*
```

Explain the results.

[Note: if you use the more efficient std* to include the state dummies, you must make sure you do not have any other variables that start with "std"!]

4) We will define a border county as one with a county centroid within 25 miles of a state border.

Create the variables:

```
gen Border25=(neardist<=25)
```

Estimate:

```
a) reg AFDCExp Border25
```

```
b) reg AFDCExp Border25 std*
```

Interpret the coefficient on Border25. Explain how Border25 is identified and how this differs between the regression without state fixed-effects (a) and with state fixed-effects (b).

Give this some careful thought before you answer!

5) Now:

```
gen BorderBen25=NeighborBen*Border25
```

and estimate:

```
reg AFDCExp Border25 NeighborBen BorderBen25
```

Interpret the coefficient on BorderBen25. If there is an unobserved state characteristic that is correlated with welfare benefits, will it bias the coefficient on BorderBen25?

6) Estimate

```
reg AFDCExp Border25 NeighborBen BorderBen25 std*
```

Using your answer to question #4 above, discuss the difference in identification of the coefficient on BorderBen25 in this model versus the one without state fixed-effects in question #5.

7) Create the variables:

```
gen Border2550=(neardist>25&neardist<=50)
```

```
gen BorderBen2550=NeighborBen*Border2550
```

Estimate:

```
reg AFDCExp Border25 Border2550 NeighborBen BorderBen25 BorderBen2550 std*
Interpret the results.
```

8) Cluster the standard errors at the state level:

```
reg AFDCExp Border25 Border2550 NeighborBen BorderBen25 BorderBen2550 std*,
cluster(state)
```

What has happened to the coefficients? To the standard errors?

Part III: Panel Data Estimation

Welfmig2.dta contains all the same variables as Welfmig1.dta, but for the years 1970-1990, and includes the variable *Year*.

1) Estimate `reg AFDCExp AFDCBen std*`. Is the coefficient on *AFDCBen* identified?

2) Add year effects to the model. Notice that the coefficient on *AFDCBen* increases. Can you tell a story for why the model without year indicators understates the effect? Here are some hints: a) What is happening to *AFDCBen* over time? b) Single mothers are becoming much more common over the 70's and 80's. What is happening to *AFDCExp* over time?

3) Estimate the model:

```
reg AFDCExp AFDCBen Border25 NeighborBen BorderBen25 std* yr*, cluster(state)
Interpret the coefficient on BorderBen25
```

4) Replace state fixed-effects and year fixed-effect with state-year fixed-effects (ignore clustering, see note below). Is the coefficient on *AFDCBen* identified? What about on *NeighborBen*?

Some different ways to do state-year effects: (don't need report them all in your homework, just one, but should experiment a little to confirm different ways give you the same thing).

```
a) gen std1yr=std1*year
   tab std1yr, gen(st1y)
```

```
   .
   gen std48yr=std48*year
   tab std48yr, gen(st48y)
```

Estimate:

```
global s "st1y2-st1y21 st2y1-st2y21 . . . st48y1-st48y21"
```

```
reg AFDCExp Border25 NeighborBen BorderBen25 $s
```

b)

```
gen id=state*100+(year-1900)
```

```
tab id, gen(sty)
```

Estimate:

```
reg AFDCExp Border25 NeighborBen BorderBen25 sty2-sty1008
```

or

```
reg AFDCExp Border25 NeighborBen BorderBen25 sty*, nocons
```

or

```
xtreg AFDCExp Border25 NeighborBen BorderBen25, fe i(id)
```

[Note 1: you will likely need to increase your memory size and your matrix size to create so many indicator variables and include them in a model:

```
set mem 100000
```

```
set mat 1100]
```

[Note 2: xtreg does not allow the cluster() command]

5) Does the welfare migration effect change over time? Create an indicator variable for the years 1980-90 and generate appropriate interaction terms to include in the regression to test this hypothesis.

Use this result to explain why your coefficient on **BorderBen25** is so much bigger in question #6 in part II compared to question #3 in part III.