# Problem Set 2: Due Thursday, October 2.

Reminder: Your two page description of the idea for your Final Paper Project is due Monday October 6.

## Problem 1

In this problem we will sketch out how to compute elevation changes by Census blocks. Remember, these elevation changes were a key input into the housing supply elasticity constructed in Saiz (2010).

On my webpage, I have posted a python script and a do file that allows you to compute the elevation within each Census block for five counties in California. Use these files, in conjunction with the US Geological Survey website (http://viewer.nationalmap.gov/viewer/) and the Census website (https://www.census.gov/geo/maps-data/data/tiger-line.html), to perform a similar analysis for a different, similarly-sized part of the United States. In particular, compute a table similar to the one that is produced in the do file (and presented at the top of the next page).

Some hints:

- On the USGS website, there is a "Download Data" button. To download the data, click on that button, choose "Click here to draw and download by bounding box," pick an appropriate area of your choice, check the "Contours" box in the menu that pops up, and then choose the "USGS Contours for <<your area>>" shape file Finally, hit the "Checkout" button which is located on the left side of the webpage.
- From the Census website, hit the button for "2010," then click on the "Download" menu, and then click on "Web Interface." Choose "Block" from the "Select a layer type menu."
- To combine the Census and USGS data you will need to open ArcGis, which you can find on WinStat.
- Check out the first six or so minutes of the following youtube video (http://www.youtube.com/watch?v=ge_U_w0PcFM ) for a tutorial of how to run a python script in ArcGis.

The table below gives the distribution of blocks' slopes, within five countries of the San Francisco Bay Area.

| County | 10th Pctile. | 25th Pctile. | 50th Pctile. | 75th Pctile. | 90th Pctile. | Blocks (Thousands) |
|---|---|---|---|---|---|---|
| Alameda | 0.026 | 0.094 | 0.147 | 0.206 | 0.266 | 63 |
| Contra Costa | 0.065 | 0.114 | 0.163 | 0.216 | 0.283 | 95 |
| Marin | 0.083 | 0.130 | 0.177 | 0.242 | 0.311 | 33 |
| San Francisco | 0.000 | 0.046 | 0.111 | 0.173 | 0.229 | 21 |
| San Mateo | 0.059 | 0.112 | 0.157 | 0.227 | 0.287 | 92 |
| Total | 0.048 | 0.106 | 0.158 | 0.217 | 0.282 | 304 |

## Problem 2

A couple of the papers discussed so far (including Blanchard and Katz 1992; Saiz 2010) apply a Bartik instrument (so named after Bartik's 1991 book). In this problem, we will explore how this variable is constructed and the extent to which it satisfies the relevance and exogeneity conditions necessary of a good instrument.

The idea behind the Bartik instrument is to measure the change in a region's labor demand that is induced by changes in the national demand for different industries' products. The series can be constructed as follows:

$$\Delta B_{i,t-k \text{ to } t} = \sum_{I \in \text{Industries}} \underbrace{\frac{\text{emp}^{I}_{i,t-k}}{\sum_{I' \in \text{Industries}} \text{emp}^{I'}_{i,t-k}}}_{\text{Term (1)}} \cdot \underbrace{\Delta \log \left( \sum_{j \in \text{Regions} \neq i} \text{emp}^{I}_{j,t-k \text{ to } t} \right)}_{\text{Term (2)}} ,$$

where $\text{emp}^{I}_{it}$ is the employment in industry $I$ in region $i$ at time $t$, and $\Delta_{t-k \text{ to } t}$ refers to taking the difference between years $t-k$ and $t$. Term (1) gives the share of employment in region $i$ that is employed in industry $I$. For example, for the auto industry this term will be relatively high in Detroit and relatively low in Madison. The second term gives the change in employment, in industry $I$, for all *other* regions. It would measure, continuing our example, how employment in the auto industry has changed in the U.S. outside of Detroit.

There are two potential data sources that you can use to compute these Bartik instruments. One is the Census County Business Patterns database, which contains info on employment in each county (or MSA) for each year between 1986 and today (the data actually go back further, but are less easily accessible). The second potential data source is the American Community Survey/Current Population Survey (both on IPUMS) that contains information on individuals' wages, the industries in which they work, and their region of residence.

In this problem, I would like you to do the following:

(a) Pick one of the papers listed above. Construct the Bartik instrument as the authors do in the paper. For a few regions, plot a time series of this variable.

(b) Run a regression of changes in regional labor demand on the left hand side and $\Delta B$ on the right hand side. What is the explanatory power of this regression? In other words, how *relevant* is the Bartik instrument as a source of labor demand shocks?

(c) Make alternative Bartik instrument data series $\Delta B_{i,t-k \text{ to } t}$ by playing around with the industry definitions and period lengths. Try a coarser industry definition (if, for example, the paper you chose applies a 2-digit industry classification, then you can recompute $\Delta B$ using a 1-digit industry classification). Also, compute an alternative $\Delta B$ series with a different period length ($k$) than the one used in your paper. How much do your answers to (a) and (b) change with these alternative definitions?

(d) As Blanchard and Katz write regarding the Bartik instrument, "This series will be valid for our purposes as long as the national growth rates are not correlated with labor supply shocks in the state [their definition of a region]" (page 25). Why, in the context of the paper you chose, may it be the case that regional supply shocks are correlated with the national growth rates? (Some potential regional supply shocks that you might want to think about are migration/immigration from other regions/countries or local changes in unemployment benefits.)

Your discussion of your (the one you chose) paper's application of the Bartik instrument should be roughly two pages long (not counting figures).

## Problem 3

Read both of the following papers. Write a referee report on one of the two.

- Ganong, Peter, and Daniel Shoag. 2013. "Why Has Regional Convergence in the U.S. Stopped?" Use the version on Daniel's home page.

- Kaplan, Greg, and Sam Schulhofer-Wohl. 2013. "Understanding the Long-Run Decline in Interstate Migration." Use the version on Greg's home page.

Elisabeth Sadoulet and Alain de Janvry have posted a useful set of guidelines on how to write a referee report, here at http://are.berkeley.edu/courses/ARE251/2004/assignments/RRGuidelines.pdf . In addition to these guidelines, keep in mind the following three points:

- First, try to relate the paper you are reviewing to the papers that we have discussed in class, in particular Barro and Sala-I-Martin (1991), Blanchard and Katz (1992), and Diamond (2013).
- Second, when writing a referee report, it's relatively easy to point out the deficiencies' in the paper's motivation, assumptions, empirical strategy, etc... . While somewhat more challenging, what is substantially more helpful is to also suggest strategies through which these shortcomings can be overcome or circumvented. Try to push yourself to write constructive comments.
- Finally, the *Guidelines* mention that a referee report should "recommend to an editor whether a paper is suitable for publication or not." You don't need to provide an explicit recommendation for this assignment.