Brian Kang

Econ 483

Project Proposal

Introduction + Question

Observe rising cities in the Pacific Northwest, notably Seattle, WA and Portland, OR. During the past decade, their boom in technological and digital engineering (HiTech) accompanied a rapid population growth and local economic success, placing these urban cities on list of international hubs for development and investment.

As much as the robustness of these locations attract business and population, these MSAs are experiencing difficulties in resolving the negative spillovers caused by a dramatic increase in demand, and nearby suburban communities like Redmond and Spokane, WA are also affected by the repercussions of scarcity. Housing and rent prices nearly double since 2012. The homelessness crisis worsened. Traffic congestion aggravated.

I would like to first focus on the homelessness crisis, which brings up my question. ***"What is the effect of residency away from workplace on poverty?"*** With the few negative spillovers listed above, it is logical to think that those who cannot afford the livelihood in more expensive locations are crowded out to less expensive locations. However, this does not imply the location of labor must necessarily change. If an individual simply changes the residency to a nearby outside city, the workplace need not be changed, although this is not the rule of thumb.

It is also logical to think that those whom are crowded out, depending on characteristics like age, race, and physical amenity preferences, concentrate around close cities. Hence creating neighborhoods that are specific age or race conglomerated. With such environment, I assume these individuals hold similar occupation and income. If this is true, those who are in modern-day poverty may be located in a location with high poverty concentration. This problem should be addressed.

Methodology: Data + Model

I aim to use micro level cross-sectional data. Datasets are collected from American Community Survey (ACS), Public Use Microdata Sample (PUMS), Current Population Survey (CPS), and Survey of Income and Program Participation (SIPP), which all are subsidiaries under the greater U.S. Census. These datasets include both household and personal information such as household demographics, income, wage, location of residence, travel time to work, location of migration, income, etc. ACS and PUMS will constitute the majority of this research.

The combined dataset from 2018 constitutes 3,214,539 observations x 286 variables, where the majority of the variables are dummy variables (flag variables) or recoded variables from past versions of the census. So essentially the useful information of this dataset is about 3 million by 130 variables large. High dimensionality is not likely to become an issue for my study with such a large dataset, which now leads to my initial econometric model:

(variable names are listed below)

$$POVPIP = \beta_0 + \beta_1 * COW + \beta_2 * JWMNP + \beta_3 * JWAP + \beta_4 * JWDP + \beta_5 * JWTR$$
$$+\beta_6 * MIG + \beta_7 * ESR + \beta_8 * POWSP + \beta_9 * HouseholdInfo$$
$$+\beta_{10} * (COW * MIG) + \beta_{11} * (JWMNP * MIG) + \beta_{12} * (JWTR * MIG)$$
$$+\vec{\beta_A} * OtherDemographicInfo + \vec{\beta_B} * OtherMSAChars + \epsilon$$

The response variable is Income-to-Poverty ratio defined as an individual's income divided by the poverty price determined by household characteristics by the CPS. The higher the value, the more income compared the set poverty level. I will use this variable as a descriptor for income inequality. The explanatory variables are those related to the physical residency location of an individual's house and work, which includes the labor status of the individual, travel to work, location of work, household information, etc. "OtherDemographicInfo" includes anything that may affect the individual's decision to be living there he or she resides, so characteristics like age, education level, race, sex, etc. "OtherMSAChars" will involve macro data related to the individual's decision making, such as housing prices, population size of location, location area, per capita GDP, etc.

This OLS model is just a beginning. If I find other good data, I plan to look into on state-by-state cases for poverty, not just analyzing the general U.S. trend. However, this may not be possible because to observe specific locational trends I have to rely on PUMA data, which seems to require a different programming tool. After this, if possible, I would like to look into the change in trend over time since the dataset I have only has data up to the past 5 years (ACS). Some other econometric models I might use include the LASSO for variable selection, decision trees for the most significant features, and neural nets for the best prediction. I considered bootstrapping, but I did not see the significance to this because the dataset is already large in magnitude as it already is.

Potential variables that will or may be considered for my econometric model:

- Division: geological region in US (west coast and mountain region)
    - (PUMA: would help if I knew how to use it)
- Region: more general region code
- ST: state
- **COW: Class of worker**
- **JWMNP: travel time to work**
- **JWAP: time of arrival at work**
- **JWDP: time of departure for work**
- **JWTR: means of transportation to work**
- **MIG: mobility status**
- NWAB: temporary absence from work
- NWAV: available for work
- NWLA: on layoff from work
- NWLK: looking for work
- NWRE: informed of recall for work
- *ADJINC: adjustment factor for income*
- *INTP: interest, dividends, and net rental income (use ADJINC to adjust to constant dollars)*
- *OIP: ALL other income (use ADJINC to adjust to constant dollars)*
- *PAP: public assistance income (use ADJINC to adjust to constant dollars)*
- *RETP: retirement income (use ADJINC to adjust to constant dollars)*
- *SEMP: self-employment income (use ADJINC to adjust to constant dollars)*
- *SSIP: supplementary security income (use ADJINC to adjust to constant dollars)*
- *SSP: social security income (use ADJINC to adjust to constant dollars)*
- *WAGP: wage or salary income*
- DRIVESP: number of vehicles calculated from JWRI
- **ESR: employment status**
- FOD1P: field of degree
- INDP: industry (not sure if this variable may be significant)
- MIGPUMA: PUMA (use MIGSP for certain variables)
- *MIGSP: Migration state or country*
- OCCP: occupation code
- PERNP: total person's earnings (use ADJINC to adjust to constant dollars)
- PINCP: total person's income (use ADJINC to adjust to constant dollars)
- POBP: place of birth
- **POVPIP: income-to-poverty ratio**
    - (= income/poverty bracket (defined by U.S. Census per year))
- POWPUMA: PUMA for work location (use POWSP for certain variables)
- **POWSP: place of work state or country**
- SOCP: standard occupational classification code
-
- **Extra features**: kids, marriage status, household condition, relationship, education, etc.
- Error term