Method and Data sections

Eric Zhao

2024-03-26

Method

Model specification

we are using the fixed-effect OLS model to explore the research question. In our model, both entity and time fixed effects are considered. We will first only include the entity effect and time effect and both included.

$Y_{\overline{takeup-rate}} = \alpha_i + \beta_1 X_{sec_{it}} + \beta_2 X_{clustering_{it}} + \beta_3 X_{vonluteering_{it}} + u_{it}$

where α_i is the sum of constant term and unobserved time-invariant heterogenities across counties : $\alpha_i = \beta_0 + \beta Z_i$

- Assumptions:
  - u_{it} is not correlated with other explanatory variables
  - $X_{1t}, X_{2t}, ...X_{nt}, ...u_{1t}, u_{2t}, ...u_{nt}$ are i.i.d. from the distribution

The base model can be expressed as a regression model containing $n-1$ dummy regressors and a constant:

$Y_{it} = \beta_0 + \beta_1 X_{sec_{it}} + \beta_2 X_{clustering_{it}} + \beta_3 X_{vonlunteering_{it}} + \gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i + \mu_{it}$

Dependent Variables

The dependent variable is the monthly recipiency rate which is also known as take-up rate of UI benefits. The calculation of this recipiency rate is following: $\frac{\text{Monthly continued claims}}{\text{Monthly Unemployed Population}}$

Notice that the continued claim is also known as insured unemployment–which is the number of people who have already filed an initial claim and who have experienced a week of unemployment and then filed a continued claim to claim benefits for that week of unemployment. Continued claims data are based on the week of unemployment, not the week when the initial claim was filed. The monthly Unemployed population in specific county is just the number of people being unemployed in specific month and county (not seasonally adjusted).

Explanatory Variables

The Explanatory variables are composed of selected social capital measures from the social capital project data repository.

The first one we are using is the level of economic connectedness by the county level. The econmic connectedness is defined as the mean level of individual EC of low-SES members of that community, as follow:

$$EC_c = \frac{\sum_{i \in L \cap c} IEC_i}{N_{Lc}}$$

where N_{Lc} is the number of low-SES individuals in community c. The IEC here represents individual economic connectedness which is calculated by:

$$IEC_i = 2\frac{H_i(g)}{d_i(g)}$$

where g is the existing social network that social captials we are limiting on. $H_i(g)$ and $d_i(g)$ are number of high SES friend and number of total friends that individual i has.

The second one is the county level cohesiveness statistics. It is defined as the average fraction of an individual's friend pairs who are also friends with each other.

$Clustering_c = \frac{\sum Clutering_i(g)}{N_c}$ where $Clustering_i(g) = \sum \frac{g_{kj}}{d_i(g)(d_i(g)-1)/2}$

$g_{kj} = 1$ denotes the existence of link between individual k and j. and $d_i(g)$ is the degree or number of friends that individual i has.

The third one is average volunteering rate by county level.Volunteering rate is defined as the share of Facebook users in the county who are a member of at least one volunteering or activism group. The researchers of social captial project start with the set of all Facebook Groups in the United States that are predicted to be about volunteering or activism based on their titles and do not have the privacy setting 'secret' enabled. To further improve this classification, they manually review the 50 largest such groups in the United States and the largest such group in each state, and remove the very small number of groups that are clearly mis-classified. Individual-level volunteering is a binary value equal to either zero or one, they also apply noise from the Laplace$(0, 1/N_\epsilon)$ distribution to protect privacy.

Expected Results

1. It would be expected that when people in low socioeconomic status are ineligible to get insured or or unwilling to take the insurance, friendship network can be an alternative source of emergent financing and the dependency can also reduce the willingness of applying further insurance. Therefore, the higher level of regional economic connectedness reduced the average take-up rate.
2. Higher level of regional clustering indicates the higher closeness of a community which each person knows each others' friends and which I think reduce the take-up rate because I expect more stable sub-groups of regional networks reflects less usage of institutional benefits but more interpersonal among small groups assuming that the number of unemployment is not correlated with regional clustering density.
3. Higher level of volunteering shows stronger inclinations toward building and sharing resources, I would expect this variable may predict more willingness in getting UI benefits from governmental programs. But it is also possible the case that higher density of "volunteering" or "activism" group in a region is correlated with lower level of unemployment which could increase the proportion of insured unemployment.

Data

Data Description

Our raw datasets consist of three major components: Social capital data, Unemployment data and Unemployment Insurance data.

The social capital statistics come from the large-scale data project The Social Capital Atlas https://socialcapital.org/ where the data are publicly available and ready for everyone to download at:https://data.humdata.org/dataset/social-capital-atlas. The dataset we are using is in the cross-section data form where first columns is each US county and other columns are specific statistics calculated or collected using privacy-protected Facebook data except population variables. The primary Sample they use to construct these statistics consists of Facebook users aged between 25 and 44 who reside in the United States, were active on the Facebook platform at least once in the prior 30 days, have at least 100 U.S.-based Facebook friends, and have a non-missing residential ZIP code as of May 28, 2022. In our study, we are focusing on the NY state so that we have subsetting the social captial datasets within the NY area.

The Unemployment data we have collected is from New York Department of Labor's Local Area Unemployment Statistics (LAUS) https://dol.ny.gov/local-area-unemployment-statistics program where it provides publicly available datasets on monthly and annual employment, unemployment, labor force, and unemployment rate data for New York State, labor market regions, metropolitan areas, counties, workforce investment regions, and municipalities of at least 25,000 people. LAUS is a joint effort between New York State and the United States Bureau of Labor Statistics (BLS) and there is also dashboard for users/customers to play arond and explore the data. Here we focus on the data from 2021-2022. The raw unemployment data we downloaded is in panel data form which includes Year, Month, area name, employed, unemployed, labor force and unemployment rate variables and total 1488 observations in total.

The UI data we have collected is from New York Department of Labor's Unemployment Insurance Data where you will be able to look at information for benefits paid, beneficiaries and initial claims by region, industry, and program.https://dol.ny.gov/unemployment-insurance-data Here we only collect the data span across 2021-2022 given that our social capital stats are calculated based on facebook users' residential ZIP code as of May, 2022. The raw data we downloaded is in a time-grouped cross-section data form across 63 counties and every month in 2021-22. The original variables include number of beneficiaries and total amount of benefits.

Variable Description Table

Variable name	Description	Unit of Measurement
takeup	Monthly take-up (recipiency) rate by county	proportion of insured unemployed people
Avg_takeup	Average of Monthly take-up (recipiency) rate by county in given year	Average proportion of insured unemployed people
ec_county	the mean level of individual EC (economic connectedness) of low-SES (for example, below-median) members of that community.	
clustering_county	The average fraction of an individual's friend pairs who are also friends with each other.	
Volunteer_rate_county	The percentage of online social network users who are members of a group which is predicted to be about 'volunteering' or 'activism' based on group title and other group characteristics.	

Summary Statistics

Table 2: Summary Statistics of Dependent and Independent Variables

	n	mean	sd	median	trimmed	mad	min	max	range
County*	1488	31.500	17.902	31.500	31.500	22.980	1.000	62.000	61.000
Year	1488	2021.500	0.500	2021.500	2021.500	0.741	2021.000	2022.000	1.000
Month*	1488	6.500	3.453	6.500	6.500	4.448	1.000	12.000	11.000
recip_rate	1454	0.502	0.181	0.487	0.488	0.183	0.125	1.100	0.975

	n	mean	sd	median	trimmed	mad	min	max	range
ec_county	1488	0.840	0.083	0.835	0.835	0.081	0.681	1.050	0.369
clustering_county	1488	0.108	0.014	0.109	0.108	0.010	0.072	0.147	0.075
volunteering_rate_county	1488	0.076	0.027	0.073	0.073	0.017	0.023	0.226	0.203

Limitation