Better Location or Stronger Network? Welfare Participation and Spatial Distribution of Networks*

Su, Danyang

October 17, 2016

Abstract

This paper presents a simple framework to study the effect of social networks on welfare participation in the presence of residential sorting. By modeling both an individual's welfare and her residential decisions, I am able to distinguish the direct network effect from the indirect effect caused by endogenous spatial distribution of networks. Empirical estimation shows a considerably larger estimate of network effects when residential sorting is controlled for. I propose a mechanism that explains underestimation in models that do not take into account of residential sorting. There are two key components in the mechanism: first, people who are likely to participate in welfare programs also tend to live in places with better access to welfare benefits. Second, these people belong to the minority in a social network and have different residential preferences than do people of the majority. As a result, neighborhoods that facilitate welfare participation tend to have weaker networks. I propose a generalized Roy's Model to illustrate the mechanism. The model generates a key testable prediction: If network strength and welfare facility of a neighborhood are indeed negatively correlated, then by eliminating variations in the levels of welfare facilities among neighborhoods, people who choose to live in neighborhoods with stronger networks are not negatively affected by worse welfare facilities, and the distribution of welfare participants should concentrate to neighborhoods with stronger networks. Counterfactual analysis is consistent with the prediction.

*I am especially grateful to Joseph Hotz for great the support in shaping this project. I thank Christopher Timmins and Patrick Bayer for comments on network and sorting model. I am grateful to Peter Arcidiacono and Arnaud Maurel for helpful comments. The result presented in this paper is preliminary and may be changed in the future.

1 Introduction:

Economists have long recognized the persistent disparities and segregation among ethnic groups in the United States (see, for example, Donohue III and Heckman, 1991; Altonji and Blank, 1999; Hellerstein, Neumark, and McInerney, 2008; Hellerstein, McInerney, and Neumark, 2011). In addition, it has been widely documented over the past decade that ethnic groups as a form of social network exert significant influence on their members (Borjas, 1992; Bertrand, Luttmer, and Mullainathan, 2000; Bandiera, Barankay, and Rasul, 2009; Beaman, 2012). Most researchers agree that having common ethnicity is crucial for members of an ethnic network to interact with each others, and for the network to influence social behaviors of its members[1]. However, social networks may also exert more subtle influences. For example, an ethnic group with largely homogeneous members may well exhibit different network effects than does one whose members are heterogeneous. To date, the empirical literature on social networks has paid little attention to the multiplicity of channels through which a network may influence its members.

In this paper, I define a network in terms of ethnicity that spans over multiple neighborhoods. The influence of the network in each neighborhood is adjusted by the relative network intensity in that neighborhood, so two neighborhoods may have different network effects only through different network intensities. This definition of social networks allows me to distinguish the two channels of network effects: a social network may exert a direct effect on its members through direct exposure to the network in a neighborhood. It may also exert an indirect effect through the spatial distribution of the network. If people sort over neighborhoods, the network density in a neighborhood may be endogenously determined by the demographics of the network.

I use a sorting model to control for endogenous network distribution over neighborhoods. The empirical model assumes that each individual in the social network simultaneously makes a decision on the neighborhood of residence and whether to participate in welfare programs. The decision depends on both individual characteristics and neighborhood characteristics. The model allows construction of an instrument for network density in each neighborhood, so I am able to disentangle the direct effect of a social network from the indirect effect. At the same time, it preserves the same structure of a baseline model that does not account for endogenous network distribution. Therefore, I am able to make sensible comparison between the two models.

Using census data for households in California, I find that, the baseline model that does not distinguish between the direct and indirect effect of a social network generally underestimates the direct social effect. After controlling for endogenous network distribution, the estimate of direct social effect on welfare participation is about three times larger. The result is robust when I apply instruments to the model. It implies that network density and unobserved welfare facilities are negatively correlated. I propose a mechanism to explain the scenario: the majority of the population in a social network are less inclined to take welfare benefits or to live in neighborhoods that facilitate welfare participation. As a result, for people of the minority in the network who are more likely to take up welfare

[1] For example, Topa [2001], Ioannides and Datcher Loury [2004], and more recently, Beaman and Magruder [2012], study the role of social networks in helping people find job matches. Sacerdote [2001] finds evidence that the GPA of a student is positively affected by her roommate's GPA. Borjas [1995] and Bertrand et al. [2000] look at ethnic groups and find that networks contribute to the persistence of disparities among ethnic groups. Montgomery [1991] provides early theoretical framework to study network effect, which has been extended by Calvo-Armengol and Jackson (2004, 2007).

benefits, the mismatch between network strength and welfare access among neighborhoods poses a trade-off: they either live in neighborhoods with strong social networks and benefit from the improved information flowing through the network, or in ones with better access to welfare benefits that does not channel through direct network effects. A model not taking into account this trade-off between the direct effect (though agents' exposure to local networks) and indirect network effect (through the trade-off caused by the spatial distribution of the network) inevitably underestimates the direct effect of social networks. I call the bias "minority bias".

I formalize the mechanism with a generalized Roy's Model. The model generates a testable prediction: if network density and access to welfare benefits of a neighborhood are negatively correlated, by eliminating the difference in the welfare access among neighborhoods, people who receive greater network effects are not penalized by weaker welfare access. As a result, the direct effect of a social network plays a bigger role and increase the share of welfare participants in neighborhoods with stronger networks. Counterfactual simulations on the ethnic groups in the dataset are generally consistent with the prediction.

The result also implies an interesting policy implication. Researchers have been keenly interested in how does relocation of households away from poor neighborhoods affect their immediate as well as long-term well beings (see, for example, Katz et al. [2001] and Kling, Liebman, and Katz [2007]). Most studies on relocation programs define a network at the neighborhood level, so a person moving to a neighborhood is exposed to a different network as well. In This paper place a caution on wide application of such programs: if social networks extends beyond a neighborhoods, then influence of the network persists even if a person moves to a new neighborhood. Therefore, it is crucial for the success of these relocation programs to move people not only to neighborhoods with better demographics but also weaker ties to old network.

This paper makes several contributions to the literature. First, it contributes to the nascent literature of endogenous networks. While the endogeneity of network formation and network effects has long been recognized in the literature, until recently empirical works have tended to exploit exogenous designs to address the issue rather than to study the role it plays as part of the network effect[2]. A notable exception is Schmutte [2016]. Schmutte [2016] studies a dynamic setting and allows individuals searching jobs through social networks over time. A worker makes the trade-off between the benefit and cost of searching jobs through a job referral network, and the effectiveness of the search depends upon the overall job market condition. In his paper, the endogeneity stems from the temporal variation of external factors such as economic status. This paper instead studies endogeneity caused by spatial distribution of social networks. If social networks differ in spatial distribution, then their members may receive different indirect network effects through different trade-off patterns between network strength and welfare facility of a neighborhood.

This paper also makes contributions to the literature of the estimation of the network effect. Following the seminal works by Borjas [1992, 1995] and Bertrand et al. [2000], most papers in the literature have placed much emphasis on exploring the direct network effects. They often identify a

[2]For example, Kling et al. [2007] study the Moving To Opportunity program, which randomly assigned poor families to neighborhoods with better social economic status. Bayer and Timmins [2007] propose a plausible source of exogeneity in network distribution, which is based upon the assumption that residents are not able to sort at the census block level, so the neighborhood at the census block level is exogenous, and can be used to control for endogenous network distribution.

shared identity (e.g. having common ethnicity, living in the same neighborhood, or studying in the same classroom), and use various forms of average demographics (e.g. test score) to measure a network and estimate its effects. On the other hand, distributions of these demographics, which are also intrinsic properties of a network, are acknowledged but rarely investigated in a rigorous manner. The lack of progress is due to the difficulty of developing a tractable frame work to capture the effect of a network through its distribution. Beaman [2012] is among the first to empirically study how does the change in demographics of a social network may affect its effect on individual outcomes. Based on the theory developed by Calvo-Armengol and Jackson [2004], she studies data obtained from a refugee resettlement program and finds evidence that the a network may actually hurt the chance of finding a job for its members if there are more newly arrived refugees, because more people in the network are competing for the job openings available. In her study, she categorizes members in an ethnic network by the years they spent in US. In this paper, I build upon the framework of Bayer and Timmins [2007] and allow for a more general characterization of the demographic distribution of a network. I explore the role of the demographic distribution in estimating direct network effects.

Finally, this paper extends the application of the discrete choice equilibrium framework. Berry [1994] and Berry et al. [1995] introduced the canonical BLP model and BLP instruments to address the correlation between endogenous variables (e.g. product price) and unobserved product characteristics. Bayer and Timmins [2007] adapt the framework to estimating the social interaction effects, and propose an generally applicable instrumenting strategy. They construct the instrument using observable, exogenous neighborhood characteristics. By exploiting the fact that endogenous variables are not affected by any particular network if there are many small networks interacting with each other in the market equilibrium, I am able to further relax the assumption of the instrument and allow inclusion of endogenous and unobserved neighborhood characteristics in the estimation.

2 Model:

In this section, I first analyze a baseline model taken from previous literature, which treats the social network in a neighborhood as exogenously given. I point out the model's weakness, and then proceed to extend it to one that controls for endogenous network distribution. Finally, I lay out the full model that allows an instrument strategy to identify the direct network effects on welfare.

2.1 Baseline Model:

The empirical framework is very similar to the model developed by Bertrand et al. [2000]. The decision of an agent i of network j in neighborhood k to participate in welfare programs is governed by the following linear probability function:

$$\mathbb{P}(Welf_{ijk}) = \alpha_0 + \alpha_1 Netw_{jk} + \eta z_{ij} + \lambda_k^1 + \mu_j + \varepsilon_{ijk} \tag{1}$$

where z_{ij} is a vector of the observed individual characteristics, λ_k^1 are the neighborhood fixed effects (e.g. easier access to welfare benefits), μ_j are the ethnicity fixed effects, and ε_{ijk} is an idiosyncratic shock. Note that under a logit specification, μ_j and ξ_{jk}^1 do not capture the observed component of

$\beta_1^* z_{ij}^* x_k^*$ perfectly without imposing strong assumptions on the distribution of z_{ij}^* and x_k^*. Therefore the model is subject to sensitivity analysis. Similar to Borjas(1992,1995), Bertrand et al. [2000], Bandiera et al. [2009], and Dustmann et al. [2016], I define networks in terms of ethnicity. It is likely that a network may also be defined over geographic affinity or similar demographics. Following the convention in the literature, I impose the restriction that only one type of network exists, namely the ethnic network. In particular, I follow Bertrand et al. [2000] and measure $Netw_{jk}$ as the product of the relative density of people of language group j living in neighborhood k (CA_{jk}) and the average welfare participation rate of the language group j ($\bar{Y}_j$):

$$Netw_{jk} = CA_{jk} \times \bar{Y}_j$$

where CA_{jk} is defined as:

$$CA_{jk} = \ln\left(\frac{C_{jk}/L_j}{A_k/T}\right)$$

The numerator term, C_{jk}/L_j, is the population speaking j in location k as a percentage of total population speaking j, and the denominator term, A_k/T, is the percentage of population living in k over total population. The specification of CA_{jk} captures the fact that if a location has higher concentration of a particular language group than that of the overall population, then it tends to provide better contact availability and stronger links among people of that language group. As a result, the welfare knowledge transmits among people faster, and the welfare attitudes of the language group are more influential on an individual's welfare decision due to more salient exposure to social networks in her neighborhood. Note that this definition assumes the attitude is uniform at the ethnic group level. Therefore it is restrictive in that people living in neighborhoods with diverse demographics shares the same attitude or information. But on the other hand, it avoids potentially large measurement errors if one use $\bar{Y}_{jk}$ instead, and given that I focus on California rather than the entire United States, people of different neighborhoods are more likely to interact with each other.

This measurement of social networks greatly improves upon previous study by allowing identification of the marginal network effect after controlling for neighborhood and ethnic group fixed effects, but its success hinges on one of the two crucial assumptions: (1) neighborhoods have the same effects on people of different characteristics, or (2) the spatial distribution of a social network across neighborhoods is exogenously determined. If condition (1) holds, then λ_k^1 fully captures the effect of a neighborhood on welfare, regardless of residential sorting of individuals. If condition (2) holds, the concentration of a neighborhood is exogenously determined ($\varepsilon_{ijk} \perp Netw_{jk}$), and α_1 is identified.

In reality, neither assumption holds. On the one side, A person's interaction with her neighborhood may also depends on her own characteristics. For example, Borjas [1999] documents that the welfare participation rate of immigrants are much more sensitive to changes in welfare benefits than that of their US-born counterparts. On the other side, it has been established that people sort over locations based on their own characteristics (see, for example, Bayer et al. [2007]). Therefore, a person whose residential preference is similar to that of the majority is likely to live in a neighborhood with strong networks, and condition (2) is violated. Therefore, if neighborhood characteristics have heterogeneous effect on both an individual's residential preference and welfare participation, equation (1) would fail to identify the network effect.

2.2 Extended Model:

A simple fix to the problem is simply controlling for the heterogeneous effect. Consider the following extension:

$$\begin{aligned}
\mathbb{P}(Welf_{ijk}) &= \alpha_0 + \alpha_1 Netw_{jk} + \eta^* z_{ij}^* + \beta_{ij}^* x_k^* \\
&\quad + \lambda_k^1 + \mu_j + \varepsilon_{ijk} \\
\beta_{ij}^* &= \beta_0^* + \beta_1^* z_{ij}^*
\end{aligned} \tag{2}$$

where z_{ij}^* and x_k^* are vectors of both observed and unobserved individual and neighborhood characteristics, respectively. β_{ij}^* is the heterogeneous coefficient. I assume β_{ij}^* takes a linear form. Equation (2) can be rewritten as:

$$\begin{aligned}
\mathbb{P}(Welf_{ijk}) &= \alpha_0 + \alpha_1 Netw_{jk} + \eta^* z_{ij}^* + \beta_1^* z_{ij}^* x_k^* \\
&\quad + \beta_0^* x_k^* + \lambda_k^1 + \mu_j + \varepsilon_{ijk}
\end{aligned} \tag{3}$$

Equation (3) fully determines an individual's likelihood of welfare participation, but for econometricians, the unobserved components of z_{ij}^* and x_k^* renders equation (3) unidentified. One common solution is to impose distributional assumptions on the unobserved components and estimate a random coefficient model. Another approach, employed in this paper, draws insights from the nature of λ_k^1 and μ_j. Note that λ_k^1 is essentially the collection of all the fixed neighborhood k's attractiveness of observed and unobserved neighborhood characteristics, so λ_k^1 mechanically captures observed and unobserved fixed effect of x_k^*. μ_j captures unobserved individual characteristics at the language group level. For example, if college graduates are less likely than high school graduates to participate in welfare programs, then a language group with a higher proportion of college graduates tends to have a smaller overall participation rate than one with lower proportion of college graduates. To econometricians who do not observe individual education level, it is as if members of one language group tends to be more likely to participate in welfare programs than the other. By the same reasoning, I capture the unobserved component of $\beta_1^* z_{ij}^* x_k^*$ with an additional term ξ_{jk}^1 (in equation (4)), the language-group-neighborhood fixed effects for people who live in neighborhood k. Note that ξ_{jk}^1 in neighborhood k varies with language group index j, because two language groups may have different demographics of residents in the same neighborhood. The final empirical model to be estimated is:

$$\mathbb{P}(Welf_{ijk}) = \alpha_0 + \alpha_1 Netw_{jk} + \eta z_{ij} + \beta_1 z_{ij} x_k + \lambda_k^1 + \mu_j + \xi_{jk}^1 + \varepsilon_{ijk} \tag{4}$$

where z_{ij} and x_k are the observed components of z_{ij}^* and x_k^*, respectively. The difference between equation (4) and equation (1) is the addition of the terms $\beta z_{ij} x_k$ and ξ_{jk}^1. The first term captures the heterogeneous welfare effects of neighborhood characteristics that depend on individual characteristics. The second term captures the unobserved heterogeneous welfare effect of neighborhood characteristics aggregated at the ethnic group level[3]. ξ_{jk}^1 can also be interpreted as the net network

[3]Note that under a logit specification, μ_j and ξ_{jk}^1 do not capture the observed component of $\beta_1^* z_{ij}^* x_k^*$ perfectly without imposing strong assumptions on the distribution of z_{ij}^* and x_k^*. Therefore the model is subject to sensitivity analysis.

facility in neighborhood k for language group j[4].

Equation (4) is now identifiable with some additional assumptions. First of all, if I assume that ξ^1_{jk} is uncorrelated to $Netw_{jk}$, α_1 is easily identified[5]. This assumption implies that after controlling for observed $z_{ij}x_k$, the remaining unobserved individual characteristics and neighborhood characteristics does not affect welfare participation and residential sorting at the same time. The assumption holds if unobserved characteristics affects welfare participation but has no tangible effect on the overall geographic distribution of social networks. Two locations may have very different levels of welfare access, but are otherwise very similar in all other characteristics. If welfare benefits concerns few people in the network. The two neighborhoods may have very similar network densities. In other words, the effect of ξ^1_{jk} on residential sorting is dominated by the combined effect of other neighborhood characteristics.

In general, the assumption that ξ^1_{jk} is not correlated to $Netw_{jk}$ does not hold. Massey and Denton [1993] document that people who are more likely to take up welfare benefits may also choose to live in places (usually in poorer neighborhoods) that enhance their chance of welfare participation, thus contributing to the spatial segregation and in equality in the United States. It is likely that an underlying spatial sorting process affect both network distribution and welfare participation. Therefore, I have to find an instrument that is correlated to $Netw_{jk}$ but is not correlated to unobserved heterogeneity that affects either welfare participation or residential sorting. It turns out it is easier to find an instrument that is exogenous in both dimensions. In order to construct the instrument, I combine equation (4) with a residential sorting process. In the next subsection, I will first lay out the model, and discuss the instrumenting strategy in section (3).

2.3 Full Model:

In this section, I develop a simple static framework in which an agent chooses a neighborhood as well as whether to participate in welfare programs[6]. The agent's preference towards a neighborhood depends on the match between her own characteristics (e.g. education) and the neighborhood characteristics (e.g. average employment rate)[7]. This specification assumes that people only make residential and welfare participation decisions. Therefore, it excludes other variables entering the choice set. This assumption is violated if certain individual characteristics are neighborhood dependent (e.g. employment) and become a part of the decision set. While the framework in this paper can be easily extended to allow additional choices, I focus only on choices over welfare participation in addition to residential decisions. Upon settling down in a neighborhood, the agent's decision to participate in welfare programs is affected by the social network she is exposed to, as well as the interplay between her characteristics and those of the neighborhood in which she lives.

[4]Note that ξ^1_{jk} also mechanically captures fixed effects of language group-neighborhood-level characteristics. Suppose, for person i from language group j, if her likelihood of welfare participation in a neighborhood k also depends on the average unemployment rate of the language group j in neighborhood k, then ξ^1_{jk} mechanically absorbs the effect. Therefore, I do not separately address them in this paper.

[5]One can simply identified $\alpha_1 Netw_{jk} + \lambda^1_k + \mu_j + \xi^1_{jk}$ for each jk, then identified α_1.

[6]For example, McKinnish [2007] finds that people migrate across state borders in response to changes in welfare benefits.

[7]Because I focus on ethnic minorities in the United States, which constitutes only a small percentage of all residents in a neighborhood, I assume that the neighborhood does not vary much from relocation of any particular minority. The clear advantage of this assumption is that it avoids multiple equilibria that often plague nonlinear choice models.

I begin with the agent's residential decision. There are $\mathscr{K}$ locations and $\mathscr{J}$ language groups. Let u_{ijk0} be the utility agent i of language group $j \in \mathscr{J}$ derive from living in neighborhood $k \in \mathscr{K}$ if she is not participating in welfare programs. Her utility is:

$$\begin{aligned} u_{ijk0} &= \beta z_{ij} x_k \\ &\quad + \lambda_k + \xi_{jk} + \varepsilon_{ijk0} \end{aligned} \tag{5}$$

where z_{ij} and x_k are defined as in the baseline model. $\beta z_{ij} x_k$ captures the effects of observed individual and neighborhood characteristics on residential sorting. Following the argument for converting equation (2) to equation (4), I construct λ_k as the overall attractiveness of neighborhood k, so a higher λ_k means neighborhood k is appealing to people across language groups. Similarly, ξ_{jk} is the attractiveness of neighborhood k to language group j. If two language groups have different distribution of z_{ij}^*, then they may have different ξ_{jk} for each k.

In addition, if agent i participates in welfare programs when she is living in neighborhood k, her utility is:

$$\begin{aligned} u_{ijk1} &= \alpha_0 + \alpha_1 Netw_{jk} + \eta z_{ij} + \beta z_{ij} x_k + \beta_1 z_{ij} x_k \\ &\quad + \lambda_k + \lambda_k^1 + \mu_j + \xi_{jk} + \xi_{jk}^1 + \varepsilon_{ijk0} + \varepsilon_{ijk1} \\ &= u_{ijk0} + \alpha_0 + \alpha_1 Netw_{jk} + \eta z_{ij} + \beta_1 z_{ij} x_k \\ &\quad + \lambda_k^1 + \mu_j + \xi_{jk}^1 + \varepsilon_{ijk1} \end{aligned} \tag{6}$$

Equation (6) shows that if person i choose to participate welfare programs in neighbor k, then in addition to utility derived from living in k, she also derives utility from participation. Note that equation (6) is similar to equation (4). Essentially, I embed the structure of the welfare choice model (i.e. equation (4)) in a neighborhood choice model (i.e. equation(5)). Combining equation (5) and equation (6), the final utility agent i of language group j receives if she lives in neighborhood k and choose participation decision $p \in \{0,1\}$ is:

$$\begin{aligned} u_{ijkp} &= \beta z_{ij} x_k + 1_p \beta_1 z_{ij} x_k + 1_p \eta z_{ij} \\ &\quad + 1_p \alpha_0 + 1_p \alpha_1 Netw_{jk} \\ &\quad + \lambda_k + 1_p \lambda_k^1 + 1_p \mu_j + \xi_{jkp} + \varepsilon_{ijkp} \end{aligned} \tag{7}$$

where $\xi_{jkp} = \xi_{jk} + 1_p \xi_{jk}^p$. Agent i of j chooses (k,p) if and only if $u_{ijkp} \geq u_{ijk'p'}$ for all $(k', p') \in \mathscr{K} \times \{0,1\}$.

This specification has one clear advantage: it minimizes the variations among model specifications. Define Q_{ijkp} as the probability of individual i of language group j living in neighborhood k and making welfare participation decision p. Note that ε_{ijkp} follows a extreme type I distribution, then the probability of i participating in welfare programs conditioned on her living in k has a analytical form:

$$\frac{Q_{ijk1}}{Q_{ijk1} + Q_{ijk0}} = \frac{\exp(\alpha_0 + \alpha_1 Netw_{jk} + \eta z_{ij} + \beta_1 z_{ij} x_k + \lambda_k^1 + \mu_j + \xi_{jk}^1)}{\exp(\alpha_0 + \alpha_1 Netw_{jk} + \eta z_{ij} + \beta_1 z_{ij} x_k + \lambda_k^1 + \mu_j + \xi_{jk}^1) + 1}$$

which is essentially the logit specification of equation (3). Therefore, the full model preserves the core structure of the baseline model.

3 Identification and Estimation:

3.1 Outline of Identification:

The key parameter of interest is α_1, the coefficient on network effects in equation (7). However, network effects are not easily identifiable in the presence of unobserved heterogeneity ξ_{jkp}, which also causes failure in identification of α_0, λ_k, λ_k^1, and μ_j. Here I briefly discuss the identification strategy. First of all, I follow the procedure laid out by Berry [1994] and rewrite equation (7) as:

$$u_{ijkp} = \beta z_{ij} x_k + 1_p \beta_1 z_{ij} x_k + 1_p \eta z_{ij} + \delta_{jkp} + \varepsilon_{ijkp} \tag{8}$$
$$\delta_{jkp} = 1_p \alpha_0 + 1_p \alpha_1 Netw_{jk} + \lambda_k + 1_p \lambda_k^1 + 1_p \mu_j + \xi_{jkp} \tag{9}$$

where δ_{jkp} is the mean utility of location k for agent i of language group j if she chooses (k,p)[8]. Essentially, equation (9) simply collect terms that are not directly identified.

In order to identify α_1 on the right hand side of equation (9), I first have to identify δ_{jkp} on the left hand side, along with β, β_1, and η from equation (8). I assume that ε_{ijkp} observes a type I extreme value distribution, so that the choice probability has analytical expressions [McFadden, 1972]. Denote $V_{ijkp} = \beta z_{ij} x_k + \beta_1 1_p z_{ij} x_k + 1_p \eta z_{ij} + \delta_{jkp}$, then the probability of agent i choosing joint decision (k,p) is simply:

$$Q_{ijkp} = \exp\left(V_{ijkp}\right) / \sum_{k \in \mathscr{K}} \left[\exp\left(V_{ijk0}\right) + \exp\left(V_{ijk1}\right)\right] \tag{10}$$

Equation (10) resembles a standard multinomial logistic model, with the addition of language group index j. Therefore, I am able to use standard procedure to identify β, β_1, η, and δ_{jkp} up to a constant for each $j \in \mathscr{J}$. Because language groups do not interfere with each other, I make normalization $\delta_{j10} = 0$ for each $j \in \mathscr{J}$ to fully identify δ_{jkp}.

Now having identified δ_{jkp}, I rewrite equation (9) as:

$$\delta_{jkp} = Const + 1_p \alpha_0 + 1_p \alpha_1 Netw_{jk} + \lambda_k + 1_p \lambda_k^1 + 1_p \mu_j + \xi_{jkp} \tag{11}$$

where $Const$ is a constant term[9]. Since δ_{jkp} is identified in the first stage, equation (11) resembles a standard linear model. Note that equation (11) inherits all the benefits of the baseline model (equation (1)). If ξ_{jkp} is assumed to be exogenous, then α_1 is directly identifiable. However, as discussed earlier, both ξ_{jkp} and $Netw_{jk}$ may be affected by the same underlying sorting process. Therefore, it is crucial to instrument for $Netw_{jk}$ in order to identify α_1.

[8] See Berry [1994] for existence and uniqueness of δ.

[9] A constant term is needed here so that $\mathbb{E}(\xi_{jkp}) = 0$ regardless of the normalization process.

3.2 Instrument:

I adopt the simulated instrument method developed by Timmins and Murdock [2007] and Bayer and Timmins [2007]. The intuition of their instrument strategy is that the share of a ethnic group in a neighborhood is an equilibrium outcome, namely the demand for a particular neighborhood depends not only on the characteristics of the neighborhood itself, but also on how these characteristics compared to those of alternative neighborhoods. If one assumes that the distribution of characteristics of one neighborhood is independent of those of the others, then the utility of living in a neighborhood are not directly affected by the attractiveness of living in other neighborhoods. Therefore, characteristics of other neighborhoods may serve as an instrument for the share of social network in an neighborhood. Bayer and Timmins [2007] argue that while any functional form that captures the comparison between utilities of living in alternative neighborhoods is a valid instrument, a natural choice is the simulated spatial distribution of social networks based only on observed individual and neighborhood characteristics.

The assumption of using only exogenous neighborhood characteristics is restricting, because many variables in social studies are intrinsically endogenous. In this paper, I am able to directly control for unobserved neighborhood and language group characteristics through fixed effects, so I may relax the assumption on using only exogenous neighborhood characteristics for generating instruments and allows the characteristics of a neighborhood to be overall endogenous but exogenous for each language group. The intuition is a simply extension of the argument by Bayer and Timmins [2007]. Again consider their proposed instrument. If there is only one social network, then I am not able to control for neighborhood fixed effects. Instead, I use observed neighborhood characteristics of other neighborhoods $k^{'}$, namely $\{x_{k'}\}_{k'\in\mathscr{K},k'\neq k}$, to form the instrument for neighborhood k. The instrument strategy requires that $\{x_{k'}\}_{k'\in\mathscr{K},k'\neq k}\perp\lambda_k$, because if $\{x_k\}_{k\in\mathscr{K}}$ are endogenous, then the unobserved neighborhood characteristics, λ_k, may also affect $x_{k'}$ in equilibrium. Instead, if multiple language groups are involved, then the same requirement holds for each language group. I can decompose the network specific heterogeneity into λ_k, the overall attractiveness of a neighborhood, and ξ_{jk}, the net attractiveness for language group j. As a result, I may modify the assumption as $\{x_{k'},\lambda_{k'}\}_{k'\in\mathscr{K},k'\neq k}\perp\xi_{jk}$.

This assumption is violated if a neighborhood's net attractiveness to an ethnic group is correlated to the overall attractiveness of other neighborhoods. For example, if an ethnic group j constitutes the the majority of U.S. population, then its residential preference heavily influences the overall attractiveness of a neighborhood. As a result, if the ethnic group strongly prefer one neighborhood over all others, then it tends to have both a high ξ_{jk}, and affects $x_{k'}$ or $\lambda_{k'}$ for all $k^{'}\neq k$. In this paper, I exclude the only possible candidate, the English speakers, and given the ethnic diversity in the United States, such violation is unlikely to occur. This assumption improves the power of the instrument in an important way, in that it justifies inclusion of commonly used neighborhood characteristics that are intrinsically endogenous (e.g. average income level and employment rate).

The instrument strategy is not applicable if ethnic groups only sort on ethnic group-specific neighborhood characteristics. For example, if a Chinese only concerns about the unemployment rate, and other demographics of the local Chinese community in a neighborhood, then each language group is essentially facing the sorting problem in the Bayer and Timmins's framework, and I would have to apply stronger assumptions to implement the instrument. In other words, language groups need to interact with each other indirectly through general market equilibrium for the instrument to work.

The procedure of constructing the instrument closely follows Bayer and Timmins [2007]. First of all, I reconstruct the individual specific neighborhood attractiveness in the absence of unobserved heterogeneity and social network effects:

$$\begin{aligned} \tilde{u}_{ijkp} &= \tilde{\beta} z_{ij} x_k + \tilde{\beta}_1 1_p z_{ij} x_k + \tilde{\eta} 1_p z_{ij} \alpha_1 \\ &\quad + 1_p \tilde{\alpha}_0 + \tilde{\lambda}_k + \tilde{\lambda}_k^1 + \tilde{\mu}_j + \varepsilon_{ikp} \end{aligned} \tag{12}$$

The coefficients with "~" sign in equation (12) are estimates from naive estimation that assumes exogenous ξ_{jkp}. I then proceed to calculate the simulated likelihood for each choice for each individual ($\tilde{Q}_{ijkp}$), and generate the simulated share of network in a neighborhood as instruments. Bayer and Timmins [2007] argue that one can use estimates from the original estimation as an initial guess. Note that even though the estimates from the naive estimation may be biased, the instrument remains valid. For example, suppose elementary school quality is a favorable amenity (i.e. the coefficient on elementary school quality is positive), then a neighborhood with a good elementary school tends to attract more residents. If the estimate of the coefficient from the naive estimation is negative, the simulated share of network nevertheless tends to be correlated, albeit negatively, to the actual share. As Bayer and Timmins [2007] argue, using estimates from the naive estimation and a variation of the true utility specification are simply a convenient and tractable choice of functional form.

3.3 Estimation:

In this subsection, I briefly discuss challenges in my estimation strategy and my approach to address them. First of all, inclusion of language-group-neighborhood-participation fixed effects (δ_{jkp}) greatly increases the number of parameters to be estimated. If the number of parameters is small, one may simply use a standard MLE method to estimate all the first stage parameters, including δ_{jkp}. But if the parameter set is large, then the estimation process is computationally taxing. Therefore, I adopt a mixed estimation strategy that nests a GMM estimator within a MLE framework and apply the techniques developed by Berry et al. [1995] and Berry et al. [2004] to estimate δ_{jkp}. For any guess of β, β_1 and η, I am able to obtain optimal δ_{jkp} through contraction mapping:

$$(\delta_{jkp})_{n+1} = (\delta_{jkp})_n + \ln Y_{jkp} - \ln Y_{jkp}((\delta_{jkp})_n, \beta_1, \beta_1^p, \eta)$$

where Y_{jkp} is the observed share of each choice node (k,p) for members of the language group j, and $Y_{jkp}(\cdot)$ is the simulated share from the n-th guess $(\delta_{jkp})_n$. Then I can use standard MLE procedure to estimate β_1, β_1^p and η:

$$\min_{\beta,\eta} \sum_i \ln(Q_{ijkp})$$

where Q_{ijkp} is defined in equation (10). Once I find all the first stage parameters, I then turn to equation (11) and estimate α_1.

The second challenge is related to the structural assumption on idiosyncratic shock ε_{ijkp}. I assume ε_{ijkp} obeys the extreme type I distribution, which implies none zero proportion for each choice node. In reality, due to the limitation on the size of the data, some choice nodes are inevitably not observed.

Therefore, I focus on a neighborhood set within a smaller geographic region, and only on large ethnic groups so it is possible to observe each choice node for each ethnic group. I discuss the data selection procedure in more detail.

4 Data and Summary Statistics:

4.1 Data:

I use the 5 percent 1990 Census Public Use Micro Sample (PUMS) in California with MSA as the geographic unit. I limit the choice set within California for three reasons: first of all, with the exception of the Spanish speaking population, no other ethnic group is observed over every MSA in the United States. Studies show that people are much more constrained across states then within a state (see Molloy et al. [2011] for a recent review), and only about 10% of sample observations in California lived in a different state in 1985. Therefore, reducing the choice set to MSAs within a state remains fairly representative of the actual choice set for residents of California. Second, California has the most diverse ethnic community in the United State, so it is more likely to find multiple ethnic groups that span all major MSAs in California. Finally, because the definition of a social network is based on ethnicity rather than geography, a natural concern arises: whether ethnic identity is a good indicator of common network. Bertrand et al. [2000] justifies the use of ethnicity in the United States. In this paper, the argument is more convincing because people of the same language group in a smaller geographic region are more likely to form a network. For example, a Chinese living in Los Angeles might have little contact with the Chinese community in New York, but she is much more likely to interact with other Chinese people in California. Therefore, using ethnicity to identify network membership is more sensible if the geographic region is smaller. I further restrict the sample to observations in the eight most populous MSA in California that accounts for about 80% of all residents. To construct location level characteristics, I use the MSA code of respondents to identify their residential choice, and calculate the mean value of individual characteristics in that location, including English speaking sample observations.

The sample population consists of non-institutionalized female samples in their adulthood (over age 20) and younger than 55[10] who do not speak English at home. I identify language group through questions in the survey. If a person writes down a language other than English to the question "Does this person speak a language other than English at home? What is this language?", her language group is identified. Many languages have low frequency in the survey, and only a few languages cover all the major MSAs in California. To maintain a viable sample size, I exclude language groups that have less than 1000 people, including both genders of all ages, or have less than 500 observations in the selected female sample. There remains 16 language groups with 90,534 observations. In addition, I drop those ineligible language groups that do not span all the choice nodes, including Korean, German, Persian, Armenian, Hindi, Arabic, and Portuguese. 80,885 observations remain in the sample.

The main variable of interest in my study is the welfare use. The decision over welfare participation is extracted from the variable "Income from public assistance". If the individual reports any

[10] I excludes females over age 55 they are more likely to receive Supplemental Security Income (SSI). I am not able to distinguish SSI from other forms of welfare benefits, so I drops female samples over 55 years old.

positive number, I assume she participates in at least one welfare program. The mean welfare used by a language group is aggregated from women between 20 and 55 years old of that language group. Appendix (8.3) lays out detailed information about data construction and variable definition.

4.2 Summary Statistics:

Table (1) summarizes the neighborhood characteristics for each MSA. I calculate MSA level characteristics by aggregating the individual characteristics of all observations in the survey in that MSA, including the English speaking population as well. Poverty level is defined as the ratio of household income to the poverty threshold in 1990[11]. Professional rate is defined as the proportion of people who hold a managerial or professional occupation in 1990[12]. The proportion of managerial/professional occupations, the unemployment rate, and the black/non-black ratio are calculated over the entire sample population of both genders, including English speaking population. For child presence rate, college degree rate, proportion of single mothers, and welfare participation rate, I average over all females between 20 and 55 years old. Note that there is remarkable heterogeneity across MSAs in the sample. For example, San Francisco and San Jose enjoy a considerable margin in poverty level and professional rate over other MSAs, whereas Los Angeles-Long Beach and Riverside-San Bernardino suffer from higher unemployment rate. Demographically, black people are not evenly distributed in California. For example, Oakland having the highest concentration of black population, about six times as high as that in Orange County. Single mothers are also unevenly distributed across MSAs, with Orange County, San Francisco, and San Jose having relatively fewer single mother than other MSAs. Finally, welfare participation rates across MSAs are low, between 2% and 8%. Like other characteristics, it also exhibits spatial variations. Not surprisingly, MSAs with high poverty level, high managerial and professional jobs, and low concentration of single mothers tend to have low welfare participation rate. Bayer and Timmins [2007] caution that the instrumenting strategy performs poorly if there is little variation in neighborhood characteristics. Overall, table (1) documents considerable heterogeneity across MSAs, which is necessary to apply the instrument developed in this paper.

Table (2) summarizes selected language group characteristics. Foremost of all, the language groups in my study exhibit substantial variations in characteristics, suggesting diversity in the demographics among language groups. For example, over half of Tagalog speaking females in the sample hold at least a college degree, whereas less than 9% of Mon-Khmer speaking females graduate from colleges. Second, table (2) illustrates that welfare participation rates vary substantially across language groups. About 40% of Mon-Khmer speaking females in the sample participate in welfare programs, whereas about only 1% of Tagalog speaking females do so in the sample. Third, welfare participation rates seem to be correlated to certain social economic status of the language groups. For example, language group with high college graduation rate, fewer single mothers, and are more fluent in English tends to have lower welfare participation rate. This pattern concurs findings in the literature. However, such correlation is not robust. For example, the Thai people in the sample have higher college graduation rate and fewer single mothers than their Spanish counterparts, but they nev-

[11] I construct my own poverty level because the poverty level provided in the data is top-coded at 500% that accounts for a non-trivial number of observations.

[12] For details, refer to Appendix (8.3)

ertheless are more likely to take up welfare benefits. Such inconsistency implies other unobserved characteristics may also affect welfare participation. Finally, the overall welfare participation rates across language groups are low, suggesting that welfare participants constitutes only a small portion of the social network.

Table (3) summarizes the characteristics of welfare participants and non-participants across all language groups. First of all, there is a clear pattern of heterogeneity in characteristics between the two groups. The vast majority of people receiving welfare are single mother and tend to give birth to one more child than female samples who do not participate in welfare programs. On the other hand, it is very unlikely for a college graduate to participate in welfare programs ($\sim 2.5\%$). Secondly, table (2) and Table (3) together illustrate that people who are likely to participate in welfare programs belong to the distinct minority in the sample. Specifically, only about 10% of female samples are single mothers, but almost half of welfare participants are single mothers. The result indicates that welfare participants are not evenly distributed in a social network, but have different characteristics than the majority non-participants.

Figure (2) illustrates the distribution of social networks. First of all, language groups show diverse patterns of spatial distribution. Chinese people mostly live in a single MSA, San Francisco, Vietnamese people are divided between San Jose and Orange County, and Japanese people tend to be more evenly distributed across the major MSAs in California. Such diversity implies the attractiveness of a neighborhood is language group-dependent. Otherwise, one should expect to see similar pattern across language groups. Even if two language groups are similar in observed demographics, they may still differ in unobserved demographics that affect the their sorting patterns. For example, the Chinese community and the Japanese community shares similar demographics, but they exhibit very different patterns of spatial distribution. Overall, it is possible that residential sorting and differences in observed or unobserved demographics together contribute to the spatial distributional patterns shown in the data. Figure (2) also implies that there might be variation in location specific attractiveness. For example, San Francisco is in general very popular among many language groups, whereas Riverside region is a less pleasant place to live for most language groups.

Finally, figure (3) demonstrates the distribution of welfare participants of each language group across MSAs. Note that overall, there are lower percentage of people taking up welfare benefits in the popular San Francisco area, as indicated by the darker blue color in the figure. In contrast, Riverside-San Bernardino and Sacramento seem to be more welfare friendly. This pattern may suggest that, in addition to attraction, neighborhoods may also exert fixed effect on welfare participation. Interestingly, figure (2) and figure (3) together suggest that there is a mismatch between network strength and welfare participation rates: MSAs (e.g. San Francisco) with stronger networks tend to have lower welfare participation rate for each language group. Similar to findings in figure (2), figure (3) also finds suggestive evidence of variation in the influence of MSAs on welfare participation. For example, Japanese welfare participants mostly live in San Diego, whereas welfare participants of most other language groups prefer to live in Sacramento.

5 Result:

In this section, I first present estimates from the baseline model, then compare it with the estimates obtained from the full model that accounts for endogenous network distribution.

5.1 Benchmark Estimation Results:

Table (4) shows results from estimating equation (1), using different sampling criteria. Column (I) shows the results from logit estimation of the baseline model (equation (1)), Column (II) estimates the same model with ineligible language groups, and Column (III) includes all language groups in the United States[13] with more than 1000 sample observations, and more than 500 female observations between 20 and 55 years old. Estimation using only the main sample yields the average marginal effect to be 0.045. If I include ineligible language groups as well, the estimated effect is larger, but it becomes smaller if the language group is defined over the entire United States. In general, the result shows an overall significantly positive network effect on welfare and is consistent with findings in the literature.

Table (4) also explores the effects of other covariates on welfare participation. The likelihood of welfare participation rises with the number of children, suggesting that children are a major family burden, and a major incentive to take up welfare programs. Welfare participation is also rising with age but at slower rate. One possible explanation is that women are most likely to take up welfare benefits during the prime childbearing age (25~35), and rely less on welfare programs when their children grow up. Graduating from a college and having good command of English both decrease the chance of participating in welfare programs, consistent with the fact that welfare participants tend to be the disadvantage people in a society, who have fewer opportunities of getting better paid jobs. A married woman is less likely to take up welfare benefits, potentially because she now has access to her husband's resources.

Finally, being a single mother raises the chance of taking up welfare benefits, and the magnitude of the effect is the largest among the effects of all the binary covariates. This outcome highlights the role single motherhood plays in welfare participation. In the United States, many welfare programs are designed to help single mothers, therefore, it is not surprising that I both observe a disproportional share of single mothers among welfare participants and a large effect in estimation. Overall, the estimation results on individual characteristics is consistent with the difference in characteristics between welfare participants and non-participants as seen in table (4).

5.2 First Stage Estimation Result:

In this section, I present estimates of coefficients of individual effect on welfare participation (η), and heterogeneous neighborhood effects on residential preference (β) and on welfare participation (β_1) in equation (8). Table (5) shows estimates of β. The variables in the first column are neighborhood characteristics, and the variables in the first row are individual characteristics, so the estimate in cell (m,n) represents the coefficient of $z_{ij}^{n}x_{k}^{m}$. Overall, table (5) suggests that people do sort over neighborhoods

[13] I drop MSAs with no observations in either welfare participation or no participation. In practice, I drop 861 (out of 390,230) observations and 15 (out of 276) MSAs.

based on their individual characteristics as well as the neighborhood characteristics. A neighborhood with higher unemployment rate attracts single mothers but alienates families with children at home and women good at English. Note that a woman with children at home who is not a single mother must be married to a man. As a result, she as part of a household may prefer to live in a neighborhood where the husband has easier access to jobs. In contrast, single mothers in the United States belong to the most impoverished group, who tend to live in the poorest neighborhood, usually with high crime rate and unemployment rate (McLanahan and Booth [1989]). Women with better English skills are more likely to obtain a desired job. As a result, they prefer to live in neighborhoods with better employment opportunities.

Estimates of the effect of occupation composition reveal the heterogeneous job preference among people in the sample. First all of, women with good English skills sorting to MSAs with a more managerial/professional job sectors, presumably because they are better qualified for such jobs. Second, women with children are less likely to live in these MSAs. A possible explanation is that having a professional or managerial job, which is usually full time, interferes with women's role in childcare. Finally, MSAs with a more managerial/professional job sectors are also more attractive to single mothers. This result is somewhat puzzling, because the career aspiration of single mothers are usually impeded by the pressure to raise children(Kimmel, 1998). The puzzle may be explained by skill complementarities. Eeckhout et al. [2014] find evidence that large cities attract both high and low-skilled workers. Note that only the largest MSAs are included in the sample, so it is possible that an MSA with a thriving managerial/professional sector also offers more low-skill job opportunities, which is attractive to single mothers.

Interestingly, richer neighborhoods tend to be less attractive to women with children present and better English skills. A possible interpretation is that conditioned on having the same unemployment rate, a richer neighborhood usually has higher living costs. An another possible reason is that the sample consists of people who are used to speak to their native tongue at home and socialize with their compatriots. Consequently, they are less likely to be culturally integrated into the United States than those who speak English at home. Lang [1986] argues that when sharing common language is crucial in working together, the welfare cost is borne by the minority. Therefore, it is not surprising that these people are sorted to less poorer MSAs.

Finally, note that the heterogeneous effect of living in a neighborhood with more/less black people are either not significant or much smaller than other neighborhood characteristics. While this result, at first glance, may be different from the consensus in the literature of racial discrimination[14], it merely means the effect does not vary much with individual characteristics. Recall that the fixed effect of a neighborhood characteristics on residential sorting is absorbed by neighborhood fixed effect λ_k, so a neighborhood with more/less black people may still be more/less attractive to people, although the effect is likely to be homogeneous. A small heterogeneous effect of the racial composition compared to that of other neighborhood characteristics also implies that members of the ethnic groups in the sample are more actively sorting over characteristics that are likely to directly affect their well beings (e.g. poverty level, unemployment rate, and distribution of professional/managerial jobs). If racial composition does affect people's residential preference, it does so likely through attitudes towards

[14] See Altonji and Blank [1999] for an overview.

black people or a general perception of living in a black neighborhood, rather than how does it affect each individual.

Table (6) shows estimates of β_1. First of all, living in an MSA with a strong managerial/professional sector discourages use of welfare benefits for women with good English skills. The effect is both significant and and large. Intuitively, a women with good English skills is better qualified for high skill jobs, and she is more likely to find one such job in an MSA with a strong managerial/professional sector. As a result, she is in less need for welfare benefits. Second, high poverty level discourages welfare participation of single mothers. The reason is that poor single mothers are more likely to sort away from richer MSAs than do poor non single mothers who may have support from husbands. Consequently, single mothers who stayed are on average better off than non single mothers. Median earning of single mothers in the three richest MSAs (i.e. San Jose, San Francisco, and Orange County) is \$10,000, and that of non-single mothers is \$9,252.5. In comparison, the numbers are \$4,800 for single mothers and \$5,500 for non-single mothers in the three poorest MSAs (i.e. Riverside-San Bernardino, Sacramento, and San Diego). I do not find significant effect of other neighborhood characteristics on welfare participation. Overall, table (6) suggests that the interaction between certain neighborhood characteristics and individual characteristics plays a role in an individual's welfare decision. Recall that people who participate in welfare programs belong to a distinct group than the majority (see table (3)), then table (5) and (6) together implies that they have different neighborhood preferences, which may also affect their likelihood of welfare participation.

Table (7) shows the estimates of η. The signs of the estimate are in general consistent with those from the baseline model. Interestingly, the effects of single motherhood, children's presence, and English Fluency are not significant. Note that table (6) allows the effects of these covariates to be neighborhood dependent, and I find significant effects of single motherhood and English fluency on welfare when I allow them to be dependent upon neighborhood characteristics. This result suggests that being a single mother or not good at English by themselves may not directly induce a woman to participate in welfare programs. Instead, they do so in a more subtle way by impeding women's access to alternative resources such as income from a well paid job. If policy makers want to better identify people who need welfare benefits most, it is more efficient to allocate more resources in neighborhoods that induce welfare participation.

5.3 Second Stage Estimation Result:

Table (8) shows estimates of the average marginal network effects. Specification (I) assumes ξ_{jkp} is exogenous. The average marginal network effect is 0.22. If there is an 1% increase (in absolute level) in the average welfare participation rate, then the welfare participation probability of its members is increased by 0.22%. This result is considerably larger than the baseline estimate 0.045, suggesting underestimation if one does not take into account of endogenous network distribution[15]. Specification

[15] One concern is that the underestimation is simply the result of the attenuation bias, caused by mis-specification of logit models. Note that the estimates in table (8) is not estimates of coefficients, but of average marginal effects. Wooldridge [2010] (p470-572) shows that estimates of average marginal effects is consistent in probit models even if the coefficients themselves are affected by attenuation bias. Interpretation of results from logit models is more complicated, but simulation shows that the bias on average marginal effects caused purely by mis-specification is second-order compared to other sources of bias (Mood, 2010).

(II) and (III) relax the assumption of exogenous ξ_{jkp} and implement instruments to estimate network effects. Specification (II) instruments on the network strength (CA_{jk}) but treats the average welfare participation rate of a language group ($\bar{Y}_j$) to be exogenous. The estimated average marginal network effect is 0.16, still significantly larger.

Finally, specification (III) instruments on both CA_{jk} and $\bar{Y}_j$. If people choose neighborhood and welfare at the same time, then the unobserved heterogeneity is likely to be correlated to average participation rate as well. Interestingly, the result of estimation is similar to that of specification (II). A possible interpretation is that the correlation with $\bar{Y}_j$ is much weaker than that with CA_{jk}. Recall that in table (2), most language groups have low average welfare participation rate. As a result, most people essentially only make residential choice. For the small share of people who are likely to consider the whole choice set, their decisions are unlikely to affect the spatial distribution of social networks, which is the major source of variation in the network measure. In the third row of table (8), I report the p-value (in percentage terms) that the estimate of average marginal network effect is smaller than that in the baseline model (0.045). In all specifications, I am able to reject the hypothesis at 10% level. The results from various specifications indicate underestimation in the baseline model.

The second row of table (8) reports coefficient estimates from auxiliary linear regressions of networks on simulated networks. First of all, the estimates in both specifications are significant. In row four of table (8), I report the F-stats of the auxiliary regressions and reject the weak instrument hypothesis. Second, both instruments are strongly correlated to actual network measures. The estimates in specification (III) is larger than that in specification (II), indicating better match after I instrument for both CA_{jk} and $\bar{Y}_j$. But the improvement is relatively small, echoing the above argument that instrumenting only on CA_{jk} is sufficient if welfare participation is very low.

6 Discussion:

6.1 Network Strength vs. Welfare Facility:

The estimates of network effects in the full model in section (5) implies under-estimation of the baseline model that does not take into account endogenous network distribution. If people hold heterogeneous preferences over neighborhood characteristics, the downward bias implies that those who are more likely to participate in welfare programs tend to live in neighborhoods with weaker social networks. The seeming counter-intuitive result actually highlights the more subtle mechanism through which a social network may affect its members. It has been widely documented that people who are more likely to participate in welfare programs tend to live in neighborhoods with better welfare access (see, for example, Blank [1988]; also see Moffitt [1992] for a review). Therefore, when they decide on residence, they not only consider the network strength, but also other aspects of a candidate neighborhood.

To understand how is residential sorting correlated to the measure of network on welfare, first note that welfare participants belongs to a distinct minority. On the one side, both the estimation result from the model and the actual data shows a low welfare participation rate. Figure (1) shows the distribution of welfare participation probability. If I define people with 10% or more likelihood as potential participants, figure (1) nevertheless indicate that this groups remains the absolute minority with only about

13% of the sample population. The result is consistent with actual welfare participation rate that falls in a $2\% - 10\%$ range across the language groups in this paper. It is also consistent with overall pattern of welfare participation (Trippe et al., 1992).

On the other side, the characteristics of welfare participants are notably different from those of the non-participants. As implied by table (3), if people sort over neighborhoods, and welfare-inclined people also prefers to live in neighborhoods that facilitate welfare participation, then it is likely that the welfare-inclined minority has very different residential preferences than does the majority. As a result, most people in a social network lives in neighborhoods that are less welfare friendly, and for the minority who are likely to participate in welfare programs, there is a trade-off between better welfare facility and stronger networks.

I adopt a general form of the Roy's model to illustrate this mechanism. To make analysis tractable, instead of a finite number of neighborhood choices as in standard Roy's models, I assume a continuous measure of locations. For illustration purpose, I have only one ethnic group in my model[16].

An individual choose which place to live. Her preference has the following parametric structure:

$$U_{k(i)} = v_i + \gamma_i \tag{13}$$

where $v_i \in \{V_L, V_H\}$, and γ_i follows a symmetric distribution so that asymmetry solely comes from v_i. Here I abstract from the utility maximization in standard selection models. Instead, a person chooses location k if k best suit her preference or type[17]. v_i captures the notion of minority/majority in a social network with $\mathbb{E}(v_i = V_L) = \alpha > 0.5$ (i.e. V_H is the minority), and γ_i captures the fact that an individual may have heterogeneous preference over many aspects, which can be summarized as a single type variable γ_i. I make the normalization such that $\mathbb{E}(v_i) = \mathbb{E}(\gamma_i) = 0$. Because $\mathbb{E}(v_i) = \mathbb{E}(\gamma_i) = 0$, I can write $V_H = (1-\alpha)w$ and $V_L = -\alpha w$ for some non-negative value w. Note that if $w = 0$, the population distribution is fully symmetric.

Now consider the geographic distribution of the language group. Let $f(\cdot)$ be the density function for γ_i, then the share s_k of the population in location k with $U_k = x$ can be written as:

$$s_k = g(x) = \alpha f(x - [1-\alpha]w) + (1-\alpha) f(x + \alpha w)$$

In plain words, s_k is the weighted sum of the V_H and V_L types people with type $v_i + \gamma_i = U_k$.

After an individual decides on the neighborhood to live, her probability of welfare participation is governed by the following empirical model:

$$y_{ik(i)} = \beta_0 + \beta_1 Netw_{k(i)} + \eta z_i + u_{k(i)} + \varepsilon_{ik(i)} \tag{14}$$

where ε_{ik} is the exogenous idiosyncratic shock, and z_i is the observed individual characteristics. $Netw_{k(i)}$ is a measure of social networks for welfare at neighborhood $k(i)$ chosen by i. I define $Netw_{k(i)}$

[16] It is important to note that while multiple language groups are needed for identification, they are not interfering with each other in determining an agent's decision.

[17] Alternatively, one can think that the person's utility function is $\min_k(U_k - v_i - \gamma_i)$, in the presence of a continuous measure of locations. The intuition is that if the choice set is large enough, there is always one choice that matches perfectly with an individual's preference.

as product of density and average welfare participation:

$$Netw_{k(i)} = s_{k(i)} \times \mathbb{E}(y_{ik(i)})$$

In addition, $y_{ik(i)}$ is also affected by the unobserved location characteristics $u_{k(i)}$, with $\mathbb{E}(u_{k(i)}) = 0$[18]. I assume places with a higher U induces people to participate in welfare programs, so $cov(u_{k(i)}, U_{k(i)}) > 0$. More specifically, I impose a parametric form: $u_k = U_k + \omega_k$, with $\omega_k \perp U_k$. Note that the people with V_H is the minority in the population, so V_H can be interpreted as the welfare-inclined type in the network. This interpretation captures that fact that the welfare-inclined minority in a social network are more likely to live in places that are overall more "welfare friendly". Now consider equation (14), if $Netw_{k(i)}$ and $u_{k(i)}$ are correlated, then a naive estimation of equation (14) without taking into account of the sorting process is biased. The direction of the bias depends on $\mathbb{E}(s_{k(i)}u_{k(i)})$. If $\mathbb{E}(s_{k(i)}u_{k(i)}) \gtreqless 0$, then $\hat{\beta}_1 \gtreqless \beta_1$. With the assumption imposed on v_i and ξ_i, I derive the following proposition:

Proposition 1. $\mathbb{E}(s_{k(i)}u_{k(i)}) \gtreqless 0$ *if and only if* $\alpha \gtreqless 0.5$.

Proof: See Appendix.

Proposition (1) shows that if an individual who is more likely to take up welfare benefits belongs to the minority in a social network ($\alpha < 0.5$), then the overall network strength of the neighborhood an individual lives in is negatively correlated to the welfare facility of the neighborhood. Intuitively, the spatial heterogeneity of network strength is an equilibrium outcome that arises from the distribution of $v_i + \gamma_i$ and is independent of the type for each i in equilibrium. But because the minority group has different location preferences than the majority, the relative shares of each group vary negatively with $u_{k(i)}$. In other words, people of the minority group are more likely to be observed in neighborhood k with high u_k, which tend to be less populated.

Proposition (1) illustrates the obstacle facing members of the welfare-inclined minority: the smaller is the group, the more eccentric it is to others in the network. In terms of residential sorting over neighborhoods measured by welfare facility, this eccentricity alienates the two groups away from each other, and members of the minority are penalized for living in less populous neighborhoods. Proposition (1) highlights the trade-off between the welfare facilities and the network strength of a neighborhood: a person either choose a more welfare-friendly neighborhood, or one with a stronger network.

6.2 Counterfactual Analysis:

The above model also generates a testable prediction. If the welfare-inclined minority group suffers from the penalty caused by the trade-off, then by removing the welfare facility of all the neighborhoods, the direct network effect on welfare participation should be more prominent, and the distribution of welfare participants should concentrate in neighborhoods with stronger networks. Formally, I adopt the following variation of equation (14):

$$y_{ik(i)} = \beta_0 + \beta_1 Netw_{k(i)} + \eta z_i + \varepsilon_{ik(i)} \tag{15}$$

[18] Note that in this model, there is only one ethnic group, so I am not able to distinguish between the overall neighborhood fixed effects and ethnic group specific neighborhood fixed effects.

but maintain the same underlying rule of residential sorting, as specified in equation (13). Equation (15) imposes a restriction that welfare participation of i does not depends on the location characteristics of the neighborhood she chooses to live. Given equation (15), the following proposition holds:

Proposition 2. *Suppose $\alpha < 0.5$, and an individual's location preference does not affect welfare participation, namely, $u_i = 0$. Denote $\hat{k}(i)$ as the counterfactual location choice made by individual i, then*

$$\mathbb{E}(s_{\hat{k}(i)}|y_{i\hat{k}(i)} = 1) > \mathbb{E}(s_{k(i)}|y_{ik(i)} = 1)$$

Proof: See Appendix.

Proposition (2) shows that conditioned on person i is participating in welfare programs, she is more likely to live in neighborhoods with stronger networks if residential sorting does not affect welfare participation. Intuitively, if potential welfare participants do not have to make the trade-off between stronger networks and more welfare friendly neighborhoods, they tend to live in neighborhoods favored by the majority. Note that $u_{k(i)}$ does not enter equation (15). Consequently, a person's location preference u_i does not affect her welfare participation, and the notion of minority is irrelevant to welfare participation.

I test this prediction through counter-factual analysis. In particular, I simulate the choice probability of each individual in the sample, using a modified specification of equation (7):

$$\begin{aligned} u_{ijkp} \quad = \quad & \beta z_{ij} x_k + 1_p \eta z_{ij} \\ & + 1_p \alpha_0 + 1_p \alpha_1 Netw_{jk} \\ & + \lambda_k + 1_p \lambda_k^1 + 1_p \mu_j + \xi_{jk0} + \varepsilon_{ijkp} \end{aligned}$$

which is the same as equation (7) but without $1_p \beta_1 z_{ij} x_k$ and $\xi_{jk1} - \xi_{jk0}$. Essentially, I remove the heterogeneous welfare facility associated with each neighborhood for each individual. I compare the spatial distribution of welfare participants of the counterfactual simulation against that of the observed data. The left panel of figure (4) shows the direction of the change in the distribution of welfare participants after I remove the heterogeneous welfare facility. It is the same figure as figure (2), with the addition of triangles indicating that where there is an increase of welfare participation for the network in the MSA. For the majority of the language groups, the most populous MSAs have witnessed an increase in the share of welfare participants. A notable exception is the Spanish speaking population, with the share of welfare participants is decreasing in Los Angeles, the MSA with the biggest Spanish community. To further investigate whether such change is salient, I also measure the change in the share of welfare participants in each MSA. The right panel shows the magnitude of change. Note that for the changes in MSAs with the strongest networks, most of them are both positive and sizable (e.g. Italian, French, Chinese, and Vietnam). The change in San Francisco for the Tagalog speaking population is not negative, but not salient. In general, results from the counterfactual simulation is consistent with the prediction of the model.

6.3 Direct Approach to Address Minority Bias:

If the indirect effect of a social network (via its spatial distribution) is entangled with the direct network effect, one may be interested in finding a simple adjustment in the baseline model to account for the indirect network effect. Note that in this paper, the minority bias that causes underestimation in the baseline model arises from the fact that the welfare inclined minority is penalized for having different residential preferences than the majority. Therefore, an intuitive strategy is to investigate directly the effect of living in places with weak networks. In this subsection, I include a measure of network strength in the baseline model. In particular, I estimate the following variant of equation (1):

$$\mathbb{P}(Welf_{ijk}) = \alpha_0 + \alpha_1 Netw_{jk} + \alpha_2 CA_{jk} + \eta z_{ij} + \lambda_k^1 + \mu_j + \varepsilon_{ijk} \tag{16}$$

Specification (16) makes a restricting assumption: the marginal effects of living in places with weak network are the same across language groups. Note that the language groups in this paper exhibit diverse socio-demographics, so depending on a language group's overall attitude towards welfare participation, living in neighborhoods with weak networks may impose varying effect on its members. For example, about 40% of Mon-Khmer versus 1.23% of Tagalog speaking population were participating in welfare programs in 1990. As a result, the minority from the former group is less eccentric than that of the latter, and a neighborhood with a strong network of Mon-Khmer speaking people is less penalizing in welfare facility as well. Even if two language groups have identical welfare participation rates, the effect of network strength may also be different; they may exhibit different level of trade-off between network strength and welfare facility. In general, the assumption of homogeneous effect of network strength fails to hold if language groups exhibit substantial variations in eccentricity (measured by welfare facility) of their welfare inclined minorities.

Table (9) reports estimates of marginal effect of $Netw_{jk}$ and CA_{jk} from equation (16). Similar to table (4), column (I)-(III) in table (9) report estimates of average marginal effects for the primary language groups in California, the extended set of language groups in California, and all language groups in the United States, respectively. The estimate for network in column (I) is similar to the IV estimates in table (8), the difference not being significant. The result implies the assumption of homogeneous effect of network strength seems to hold for the language groups in the study. Column (II) investigate the effect of including ineligible language groups, the estimate is slightly larger. The two estimates together suggest that the assumption of homogeneous effect of network strength seems to hold for language groups in California. Note that estimates for network strength are negative both for specifications, supporting the argument that the welfare inclined minority is penalized for their eccentricity in residential preference.

The estimate of marginal network effect in column (III), however, remains much smaller. In this case, including CA_{jk} does not improve the estimation by much. This discrepancy between the estimates in California and in the United States is consistent with fact that the validity of the assumption is weaker at the national level than at the state level. Recall that there are only eight MSAs in my estimation. If the choice set is small, then the spatial distribution among social networks are more similar. To see this argument, suppose there are only two MSAs to choose from, then ethnic groups with vastly different socio-demographics may have similar distribution between the two MSAs. As more MSAs join the choice set, the effect of demographics on location preference becomes more pronounced, and ethnic groups start to develop different patterns of spatial distribution. As a result, the validity of the assumption is weakened. Estimates of network strength are significant larger in the first two specifications than that in the third specification, also implying heterogeneity in penalty facing minorities of language groups across the nation. However, because the estimation framework does not allow me to estimate over the U.S. sample, I am not able to rule out the possibility that the true average marginal effect is not significantly different from 0.047. In general, inclusion of a measure of network strength may ease the bias, but its validity requires further investigation.

7 Conclusion:

In this paper, I distinguish the direct effect and the indirect effect of social networks on welfare participation. Social networks affect people's welfare decision directly through network exposure: a network's attitude towards welfare participation, measured by the average welfare participation of the people in the network, is more effective on an individual's welfare decision if there is greater contact intensity (or a stronger network) among members of the network in the neighborhood where she lives. In addition, social networks also have more subtle effect through endogenous network distribution. People have heterogeneous residential preferences, and if welfare inclined people prefer to live in more welfare friendly neighborhoods, their residential choices are endogenous, which are determined by the match between their individual characteristics and the location characteristics. At the same time, residential sorting of members of a network and the network's demographic composition together determines its spatial distribution, which in turn affects the level of contact intensity a person faces. In other words, the network intensity an agent is exposed to is not exogenously assigned, but rather endogenously determined by the demographics of the network.

I develop a framework to disentangle the direct network from the indirect effect by modeling an agent's joint decision on residence and welfare participation. This framework allows me to adapt the instrument strategy developed by Bayer and Timmins [2007]. The adapted instrument improves on the applicability of the approach of Bayer and Timmins [2007] by allowing endogenous variable to be included as covariates. I find strong evidence of residential sorting, and, to a less extent, heterogeneous effect of neighborhood characteristics on individual's participation. Estimation with the proposed instruments yields a much higher estimate of the direct network effect at $0.15 \sim 0.16$, compared to 0.045 in the baseline model that does not account for endogenous network distribution. I am able to reject the null hypothesis that the direct network effect is less than or equal to 0.045 at the 10% level. The comparison implies underestimation of direct network effect in the baseline model.

Next I propose a mechanism to explain the estimation. If welfare inclined people belong to the minority in a social network, then they are penalized for having different residential preferences than the majority. There are two key components in the model: first, welfare inclined agents tend to live in neighborhoods that facilitate welfare participation. Second, they belong to the minority, so their preferred neighborhoods tend to have weaker networks. As a result, a person of the welfare inclined minority has to make a trade-off, she either lives in a neighborhood with better welfare facility, or in one that provides better information through stronger a network. This trade-off marked causes the downward bias in the baseline estimation, which I call the minority bias. The model also generates a testable prediction: if the penalty for living in weaker neighborhood is lifted, one should see that the distribution of welfare participants concentrate towards to neighborhoods with stronger network. The result from counterfactual analysis is consistent with the prediction. I also discuss a simple remedy to the minority bias in the baseline model and discuss its validity. I include the direct measure to the network intensity to proxy for the penalty facing the minority. If I focus on California, I obtain estimates similar to those from models that control for endogenous network distribution. However, this correction measure seems to fail if I estimate the model at the national level.

Social networks are formed by people. Therefore, properties of a social network include not only the common identity shared by members of the network, but also the diversity in many other dimensions. The understanding of such diversity is crucial to understanding the workings of a social network on its members, but it has received little attention so far. This paper explores one channel through which the diversity affects people's welfare decision, namely the endogenous spatial distribution of a social network. However, the conclusion may extend beyond welfare participation, and the channel beyond spatial dimensions. A fuller understanding of network effects awaits further studies.

8 Appendix:

8.1 Proof of Proposition (1):

Proof. I will first prove the case when $\alpha < 0.5$. The proof for $\alpha > 0.5$ is similar. To prove proposition (1), note that:

$$\begin{aligned}
\mathbb{E}(s_{k(i)}u_{k(i)}) &= \mathbb{E}(s_{k(i)}U_{k(i)}) + \mathbb{E}(s_{k(i)}\omega_{k(i)}) \\
&= E(s_{k(i)}U_i) \\
&= \int ug(u)g(u)du \\
&= \int ug^2(u)du \\
&= \alpha^2 \int uf^2(u-[1-\alpha]w)du + (1-\alpha)^2 \int uf^2(u+\alpha w)du \\
&\quad +2\alpha(1-\alpha)\int uf(u-[1-\alpha]w)f(u+\alpha w)du \qquad (17)
\end{aligned}$$

The second equality holds because $\omega_k \perp U_k$, and $s_{k(i)}$ is a function of $\omega_{k(i)}$. Denote $h_1 = \int f^2(u-[1-\alpha]w)du$, $h_2 = \int f^2(u+\alpha w)du$, and $h_3 = \int f(u-[1-\alpha]w)f(u+\alpha w)du$, then equation (17) can be rewritten as

$$
\begin{aligned}
\mathbb{E}(s_{k(i)}u_i) &= h_1\alpha^2 \int u \frac{f^2(u-[1-\alpha]w)}{h_1} du \\
&+ h_2(1-\alpha)^2 \int u \frac{f^2(u+\alpha w)}{h_2} du \\
&+ h_3 2\alpha(1-\alpha) \int u \frac{f(u-[1-\alpha]w)f(u+\alpha w)}{h_3} du \\
&= h_1\alpha^2(1-\alpha)w - h_2(1-\alpha)^2\alpha w + h_3\alpha(1-\alpha)(1-2\alpha)w \\
&= (h_1-h_3)\alpha(1-\alpha)(2\alpha-1)w \qquad (18)
\end{aligned}
$$

The second equality holds because the three integrands are essentially expectations of a symmetric distribution with respect to $(1-\alpha)w$, $-\alpha w$, and $\frac{(1-2\alpha)}{2}w$, respectively. The third inequality holds because $h_1 = h_2$. As a result, $\mathbb{E}(s_{k(i)}u_i) < 0$ if $h_1 > h_3$. By Cauchy-Schwarz Inequality:

$$
\begin{aligned}
\int f(u-[1-\alpha]w)f(u+\alpha w)du &\leq \left[\int f^2(u-[1-\alpha]w)du\right]^{\frac{1}{2}} \left[\int f^2(u+\alpha w)du\right]^{\frac{1}{2}} \\
&= \left[\int f^2(u-[1-\alpha]w)du\right]^{\frac{1}{2}} \left[\int f^2(u-[1-\alpha]w)du\right]^{\frac{1}{2}} \\
&= \int f^2(u-[1-\alpha]w)du
\end{aligned}
$$

Because $f(\cdot)$ is a probability density function, equality holds if and only if $w = 0$, which implies $\alpha = 0.5$. Therefore, if $\alpha < 0.5$, $\mathbb{E}(s_{k(i)}u_i) < 0$. If $\mathbb{E}(s_{k(i)}u_i) < 0$, equation (18) implies $(h_1-h_3)(2\alpha-1) < 0$ and $\alpha \neq 0.5$. I have shown that $h_1 > h_3$ when $\alpha \neq 0.5$, so $\alpha < 0.5$. By similar procedure, $\mathbb{E}(s_{k(i)}u_i) > 0$ if and only if $\alpha > 0.5$.

If $\alpha = 0.5$, then there is no asymmetry in u_i. $g(u) = f(u)$, and

$$
\begin{aligned}
\mathbb{E}(s_{k(i)}u_i) &= \int u f^2(u)du \\
&= \int f^2(u)du \times \int u \frac{f^2(u)}{\int f^2(u)du} du \\
&= \int f^2(u)du \times 0 \\
&= 0
\end{aligned}
$$

The third equality exploits the mean of a symmetric distribution with respect to 0 is 0. If $\mathbb{E}(s_{k(i)}u_i) = 0$, then from equation (18), $(h_1-h_3)(2\alpha-1) = 0$. Because $h_1 = h_3 \Leftrightarrow \alpha = 0.5$, $(h_1-h_3)(2\alpha-1) = 0 \Leftrightarrow \alpha = 0.5$. Therefore, $\mathbb{E}(s_{k(i)}u_i) = 0$ if and only if $\alpha = 0.5$. □

8.2 Proof of Proposition (2):

Proof. Note that

$$
\begin{aligned}
\mathbb{E}(y_{ik(i)}Netw_{k(i)}) &= \mathbb{P}(y_{ik(i)} = 1)\mathbb{E}(Netw_{k(i)}|y_{ik(i)} = 1) \\
&= \mathbb{E}(y_{ik(i)})^2\mathbb{E}(s_{k(i)}|y_{ik(i)} = 1)
\end{aligned}
$$

Taking expectation of both sides of equation (14), $\mathbb{E}(y_{ik(i)})$ can be expressed as:

$$\mathbb{E}(y_{ik(i)}) = \frac{\beta_0 + \eta\mathbb{E}(z_i)}{1-\beta_1\mathbb{E}(s_{k(i)})}$$

Because $\mathbb{E}(y_{ik(i)})$ is not affected by u_i, so $\mathbb{E}(y_{ik(i)}) = \mathbb{E}(y_{i\hat{k}(i)})$. To prove the result, I only need to show $\mathbb{E}(y_{ik(i)}Netw_{k(i)}) < \mathbb{E}(y_{i\hat{k}(i)}Netw_{\hat{k}(i)})$. Multiplying both sides of equation (14) by $Netw_{k(i)}$ and taking expectations, I have

$$\mathbb{E}(y_{ik(i)}Netw_{k(i)}) = \beta_0\mathbb{E}(Netw_{k(i)}) + \beta_1\mathbb{E}(Netw_{k(i)})^2 + \eta\mathbb{E}(z_iNetw_{k(i)}) + \mathbb{E}(u_iNetw_{k(i)})$$

Because individuals preferences towards locations are not affected by with or without $u_{k(i)}$ in equation (13), I have:

$$\begin{aligned}\mathbb{E}(y_{ik(i)}Netw_{k(i)}) - \mathbb{E}(y_{i\hat{k}(i)}Netw_{\hat{k}(i)}) &= \mathbb{E}(u_iNetw_{k(i)}) \\ &= \mathbb{E}(y_{ik(i)})\mathbb{E}(u_is_{k(i)})\end{aligned}$$

If $\alpha < 0.5$, by proposition (1), $\mathbb{E}(u_is_{k(i)}) < 0$, so $\mathbb{E}(s_{\hat{k}(i)}|y_{i\hat{k}(i)} = 1) > \mathbb{E}(s_{k(i)}|y_{ik(i)} = 1)$. $\square$

8.3 Data Construction:

(To be finished)

References

Joseph G Altonji and Rebecca M Blank. Race and gender in the labor market. *Handbook of labor economics*, 3:3143–3259, 1999.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, 2009.

Patrick Bayer and Christopher Timmins. Estimating equilibrium models of sorting across locations*. *The Economic Journal*, 117(518):353–374, 2007.

Patrick Bayer, Fernando Ferreira, and Robert McMillan. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638, 2007. URL `http://www.nber.org/papers/w13236`.

Lori Beaman and Jeremy Magruder. Who gets the job referral? evidence from a social networks experiment. *The American Economic Review*, 102(7):3574–3593, 2012.

Lori A Beaman. Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the us. *The Review of Economic Studies*, 79(1):128–161, 2012.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

Steven Berry, James Levinsohn, and Ariel Pakes. Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, 112(1): 68–105, 2004.

Steven T Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

Marianne Bertrand, Erzo FP Luttmer, and Sendhil Mullainathan. Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055, 2000.

Rebecca M Blank. The effect of welfare and wage levels on the location decisions of female-headed households. *Journal of Urban Economics*, 24(2):186–211, 1988.

George J Borjas. Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics*, 107 (1):123–150, 1992.

George J Borjas. Ethnicity, neighborhoods, and human capital externalities. *THE AMERICAN ECONOMIC REVIEW*, 1995.

George J Borjas. Immigration and welfare magnets. *Journal of labor economics*, 17(4):607–637, 1999.

Antoni Calvo-Armengol and Matthew O Jackson. The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454, 2004.

Antoni Calvó-Armengol and Matthew O Jackson. Networks in labor markets: Wage and employment dynamics and inequality. *Journal of economic theory*, 132(1):27–46, 2007.

John J Donohue III and James Heckman. Continuous versus episodic change: The impact of civil rights policy on the economic status of blacks. *Journal of Economic Literature*, 29(4):1603–1643, 1991.

Christian Dustmann, Albrecht Glitz, Uta Schönberg, and Herbert Brücker. Referral-based job search networks. *The Review of Economic Studies*, 83(2):514–546, 2016.

Jan Eeckhout, Roberto Pinheiro, and Kurt Schmidheiny. Spatial sorting. *Journal of Political Economy*, 122(3):554–620, 2014.

Judith K Hellerstein, David Neumark, and Melissa McInerney. Spatial mismatch or racial mismatch? *Journal of Urban Economics*, 64(2):464–479, 2008.

Judith K Hellerstein, Melissa McInerney, and David Neumark. Neighbors and co-workers: the importance of residential labor market networks. 29(4):659–695, 2011.

Yannis M Ioannides and Linda Datcher Loury. Job information networks, neighborhood effects, and inequality. *Journal of economic literature*, 42(4):1056–1093, 2004.

Lawrence F Katz, Jeffrey R Kling, and Jeffrey B Liebman. Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, 116(2):607–651, May 2001.

Jean Kimmel. Child care costs as a barrier to employment for single and married mothers. *Review of Economics and Statistics*, 80(2):287–299, 1998.

Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

Kevin Lang. A language theory of discrimination. *The quarterly journal of economics*, pages 363–382, 1986.

Douglas S Massey and Nancy A Denton. *American apartheid: Segregation and the making of the underclass*. Harvard University Press, 1993.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. Technical report, 1972.

Terra McKinnish. Welfare-induced migration at state borders: New evidence from micro-data. *Journal of public Economics*, 91(3):437–450, 2007.

Sara McLanahan and Karen Booth. Mother-only families: Problems, prospects, and politics. *Journal of Marriage and the Family*, pages 557–580, 1989.

Robert Moffitt. Incentive effects of the us welfare system: A review. *Journal of Economic Literature*, 30(1):1–61, 1992.

Raven Molloy, Christopher L Smith, and Abigail Wozniak. Internal migration in the united states. *The Journal of Economic Perspectives*, 25(3):173–196, 2011.

James D Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American economic review*, 81(5):1408–1418, 1991.

Carina Mood. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, 26(1):67–82, 2010.

Bruce Sacerdote. Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, (1):681–704, May 2001.

Ian M Schmutte. Labor markets with endogenous job referral networks: Theory and empirical evidence. *Labour Economics*, 42, 2016.

Christopher Timmins and Jennifer Murdock. A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics and management*, 53(2): 230–249, 2007.

Giorgio Topa. Social interactions, local spillovers and unemployment. *The Review of Economic Studies*, 68(2):261–295, 2001.

Carole Trippe, Pat Doyle, and Andrew Asher. Trends in food stamp program participation rates: January 1976 to 1990. Technical report, U.S. Department of Agriculture, Food and Consumer Service, Alexandria, VA, 1992.

Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

Table 1: Summary Statistics: MSA-level Statistics

MSA	Sample Size	Poverty Level	Unemployment %	Professional %	Black %	Child Present	Single Mom %	Welfare %
Orange County	122,362	557.29	4.21	23.91	1.58	0.45	8.24	2.78
Los Angeles-Long Beach	429,407	459.79	6.52	20.36	9.77	0.46	11.51	5.71
Oakland	102,428	503.13	4.78	25.39	12.74	0.45	10.93	5.64
Riverside-San Bernardino	115,615	393.47	6.29	17.11	5.94	0.56	11.44	6.95
Sacramento	70,946	420.43	5.15	21.70	6.22	0.49	11.90	7.76
San Diego	118,620	414.25	4.95	21.44	5.68	0.47	10.39	5.41
San Francisco	73,097	562.58	4.20	26.93	6.10	0.35	7.44	3.16
San Jose	74,928	564.27	4.13	27.57	3.20	0.44	8.51	3.59

Note: The table is generated from the 1990 PUMS 5% dataset for California. I drop the institutionalized sample. Poverty Level is calcualted at the household level. Employment rate, the rate of managerial and professional occupations, and the ratio of black residents are calculated at the individual level of both genders between 20 and 55 years old. The children present rate, single mother rate, and welfare takeup rate is calculated at the individual level of females between 20 and 55 years old.

Table 2: Summary Statistics: Language-level Statistics

Language	Sample Size	Age	Child Present	College %	Married	No. Kid(>0)	Single Mom %	English Fluency	Welfare %
Italian	816	37.81	0.39	25.74	0.65	2.20	5.76	0.96	2.57
French	1,572	37.19	0.36	40.08	0.57	2.15	7.12	0.95	2.74
Spanish	57,746	34.08	0.56	6.98	0.61	2.82	13.17	0.61	5.76
Chinese	7,671	35.95	0.48	33.26	0.69	2.25	4.71	0.69	4.16
Thai	784	35.96	0.54	18.37	0.67	2.64	7.65	0.67	12.50
Japanese	1,752	37.29	0.45	32.02	0.71	2.01	3.37	0.75	1.54
Mon-Khmer	515	34.62	0.69	3.69	0.63	3.58	15.73	0.37	40.39
Vietnamese	2,932	34.46	0.55	13.10	0.62	2.87	9.96	0.62	15.08
Tagalog	7,097	37.08	0.53	50.57	0.69	2.39	7.02	0.97	1.23
English	184,557	36.03	0.42	27.06	0.56	2.23	10.59	1.00	5.21

Note: The table is generated from the 1990 PUMS 5% dataset for California for each major language group. I drop the institutionalized sample, and choose only the language groups with both participation and non-participation observations in all the eight major MSAs. Furthermore, I restrict attention to language groups with at least 1000 total observations of both genders between 20 and 55 years old, and at least 500 observations of females between 20 and 55 years old. The language-group characteristics are calculated at the individual level of females between 20 and 55 years old.

Table 3: Summary Statistics for Welfare Participants and Non-participants

Welfare	Freq.	Child Present	College %	Married	No. Kid(>0)	Single Mom %	English Fluency	Immigrant %
No	76,311	0.53	15.73	0.64	2.67	9.06	0.66	78.12
Yes	4,574	0.73	2.47	0.33	3.34	48.19	0.56	69.00

Note: The table is generated from the 1990 PUMS 5% dataset for California for each major language group. I drop the institutionalized sample and the English speaking sample, and choose only the language groups with both participation and non-participation observations in all the eight major MSAs. Furthermore, I restrict attention to language groups with at least 1000 total observations of both genders between 20 and 55 years old, and at least 500 observations of females between 15 and 55 years old.

Table 4: Benchmark Estimation on Welfare Participation

	CA	CA with Excluded	US
	b/se	b/se	b/se
Network	0.045***	0.070***	0.032***
	(0.017)	(0.015)	(0.005)
Age/100	0.221***	0.207***	0.120***
	(0.073)	(0.068)	(0.034)
$Age^2/10000$	-0.414***	-0.377***	-0.283***
	(0.100)	(0.093)	(0.046)
College Graduate	-0.065***	-0.059***	-0.061***
	(0.005)	(0.004)	(0.002)
Married	-0.047***	-0.045***	-0.054***
	(0.003)	(0.002)	(0.001)
Child Present	0.008***	0.008***	0.000
	(0.003)	(0.003)	(0.001)
Single Mom	0.067***	0.065***	0.073***
	(0.003)	(0.003)	(0.001)
No. Kid	0.012***	0.011***	0.012***
	(0.000)	(0.000)	(0.000)
English Mastery	-0.005***	-0.009***	-0.013***
	(0.002)	(0.001)	(0.001)
N	80885	90534	389369
* p<0.10, ** p<0.05, *** p<0.01			

Note: Variable Age is defined as age/100, and Age^2 is defined as $Age^2/10000$. Column I shows the result from the logit estimation for the main sample observations in California. Column II adds groups that do not span the whole choice set. Column III extends the sample to inlcude every language group with more than 1000 total sample observations and more than 500 female sample observations between 20 and 55 years old.I drop MSAs with no observations in either welfare participation or no participation. In practice, I drop 861 (out of 390,230) observations and 15 (out of 276) MSAs.

Table 5: Estimation of Coefficient on Individual Preference over Locations

	Child Present	Single Mom	English Fluency
Poverty Level	-1.71***	-0.96	-8.48***
	(0.41)	(0.79)	(0.46)
Black	-0.00	-0.02	-0.01*
	(0.01)	(0.01)	(0.01)
Unemployment	-0.15***	0.27***	-0.14***
	(0.03)	(0.06)	(0.04)
Professional	-4.70***	10.46***	12.89***
	(1.37)	(2.66)	(1.52)
* p<0.10, ** p<0.05, *** p<0.01			

Note: The table is generated from the 1990 PUMS 5% dataset for California. I drop institutionized sample and keep only MSAs with population density greater than 4%. I furhter keep only female smaples between 20 years old and 55 years old speaking foreign language at home. I drop language groups that have less than 1000 person speaking the language in the sample, or have less than 500 female smaples between 20 years old and 55 years old speaking the language. I also exclude language groups that does not span the whole choice set. Variables in the first column are the MSA-level neighborhood characteristics, and variables in the first row are the individual characteristics. The optimizer package used is NLopt BOBYQA (http://ab-initio.mit.edu/wiki/index.php/NLopt)

Table 6: Estimation of Coefficient on Heterogeneous Effect of Geographic Characteristics

	Child Present	Single Mom	English Fluency
Poverty Level	1.95	-3.40*	5.17
	(1.91)	(1.85)	(2.27)
Black	0.04	-0.01	0.01
	(0.03)	(0.03)	(0.03)
Unemployment	-0.19	0.06	-0.18
	(0.15)	(0.14)	(0.19)
Professional	-7.44	6.41	-21.62**
	(5.85)	(5.59)	(7.94)
* p<0.10, ** p<0.05, *** p<0.01			

Note: The table is generated from the 1990 PUMS 5% dataset for California. I drop institutionized sample and keep only MSAs with population density greater than 4%. I furhter keep only female smaples between 20 years old and 55 years old speaking foreign language at home. I drop language groups that have less than 1000 person speaking the language in the sample, or have less than 500 female smaples between 20 years old and 55 years old speaking the language. I also exclude language groups that does not span the whole choice set. Variables in the first column are the MSA-level neighborhood characteristics, and variables in the first row are the individual characteristics. The optimizer package used is NLopt BOBYQA (http://ab-initio.mit.edu/wiki/index.php/NLopt)

Table 7: Estimates of the Coefficent on Individual Effect on Welfare Participation

Age/100	$Age^2/10000$	College	Married
4.62*	-8.88***	-1.44***	-1.05***
(2.14)	(2.94)	(0.10)	(0.06)

Child Present	Single Mom	Number of kids	English Fluency
1.68	1.40	0.26***	3.02
(1.34)	(1.27)	(0.01)	(1.66)

* p<0.10, ** p<0.05, *** p<0.01

Note: The dependent variable is individual's welfare participation choice. The inference is obtained through the bootstrap method. The optimizer package used is NLopt BOBYQA (http://ab-initio.mit.edu/wiki/index.php/NLopt)

Table 8: Estimation of Average Marginal Effect of Network

	(OLS)	(IV: CA only)	(IV: Network Measure)
Network	0.22***	0.16**	0.15**
	(0.03)	(0.07)	(0.07)
Aux. IV Est.		0.76***	0.81***
		(0.16)	(0.17)
P-value (%): Est.$\leq$0.045	0.00	5.31	7.56
F-stat of IV est.	N/A	22.3	21.9
* p<0.10, ** p<0.05, *** p<0.01			

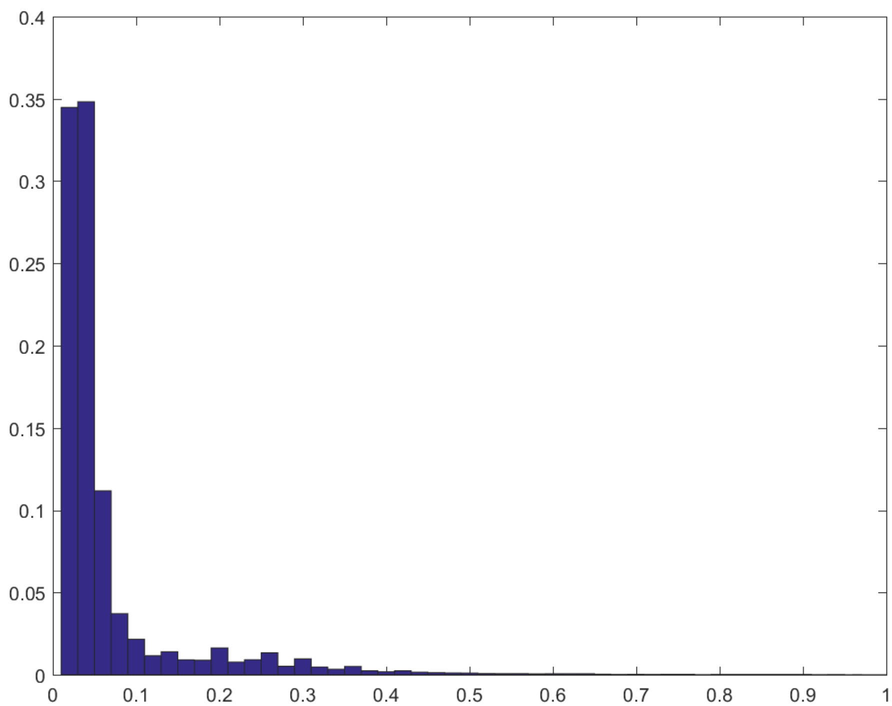
Note: The average marginal effect is calculated given individual's residence choice. Specification (I) assumes network measure is exogenous to language group specific unobserved heterogeneity. Specification (II) implements instrument on network strength. Specification (III) implements on both network strength and average participation rate of a language group. The second row of the table reports coefficient estimates from auxiliary linear regressions of networks on simulated networks. The third row reports the probability that the estimates are smaller than the estimate from the baseline model. The fourth row reports the F-stats for auxilliary regression.

Table 9: Benchmark Estimation on Welfare Participation Controlling for Network Strength

	CA	CA with Excluded	US
	b/se	b/se	b/se
Network	0.134***	0.145***	0.048***
	(0.028)	(0.025)	(0.006)
Network Strength	-0.013***	-0.011***	-0.004***
	(0.003)	(0.003)	(0.001)
Age/100	0.221***	0.206***	0.119***
	(0.073)	(0.068)	(0.034)
$\text{Age}^2/10000$	-0.415***	-0.377***	-0.282***
	(0.100)	(0.093)	(0.046)
College Graduate	-0.065***	-0.060***	-0.062***
	(0.005)	(0.004)	(0.002)
Married	-0.047***	-0.045***	-0.054***
	(0.003)	(0.002)	(0.001)
Child Present	0.008***	0.008***	0.000
	(0.003)	(0.003)	(0.001)
Single Mom	0.067***	0.065***	0.073***
	(0.003)	(0.003)	(0.001)
No. Kid	0.012***	0.011***	0.012***
	(0.000)	(0.000)	(0.000)
English Mastery	-0.005***	-0.009***	-0.013***
	(0.002)	(0.001)	(0.001)
N	80885	90534	389369
* p<0.10, ** p<0.05, *** p<0.01			

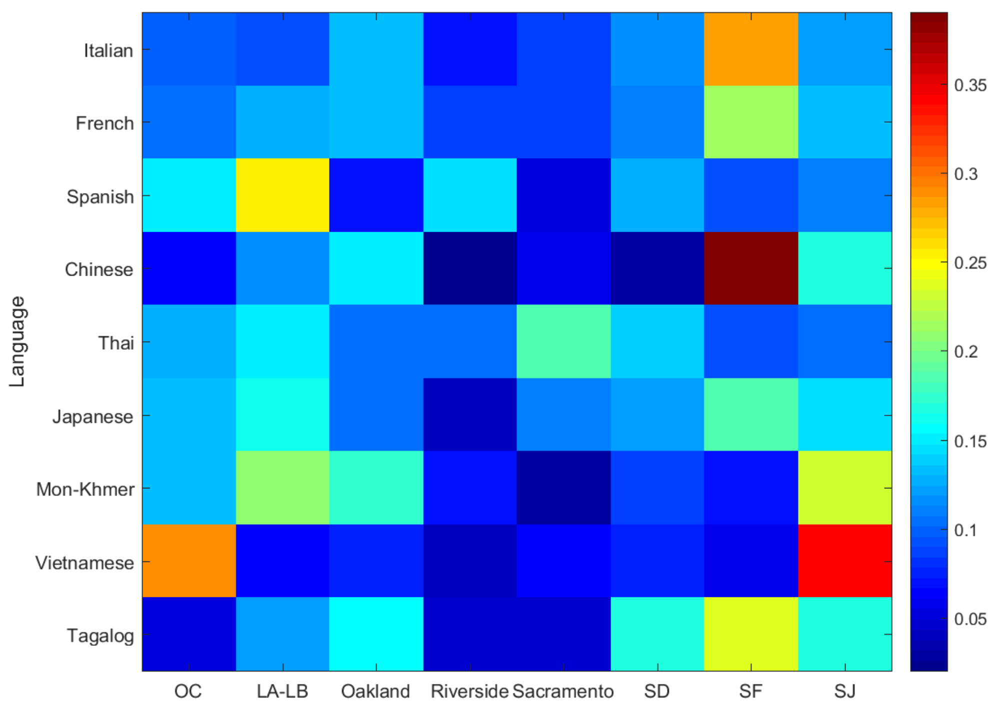
Note: Variable Age is defined as age/100, and Age^2 is defined as $\text{Age}^2/10000$. Column I shows the result from the logit estimation for the main sample observations in California. Column II adds groups that do not span the whole choice set. Column III extends the sample to inlcude every language group with more than 1000 total sample observations and more than 500 female sample observations between 20 and 55 years old.I drop MSAs with no observations in either welfare participation or no participation. In practice, I drop 861 (out of 390,230) observations and 15 (out of 276) MSAs.

Figure 1: Histogram of Women's Welfare Participation Probability



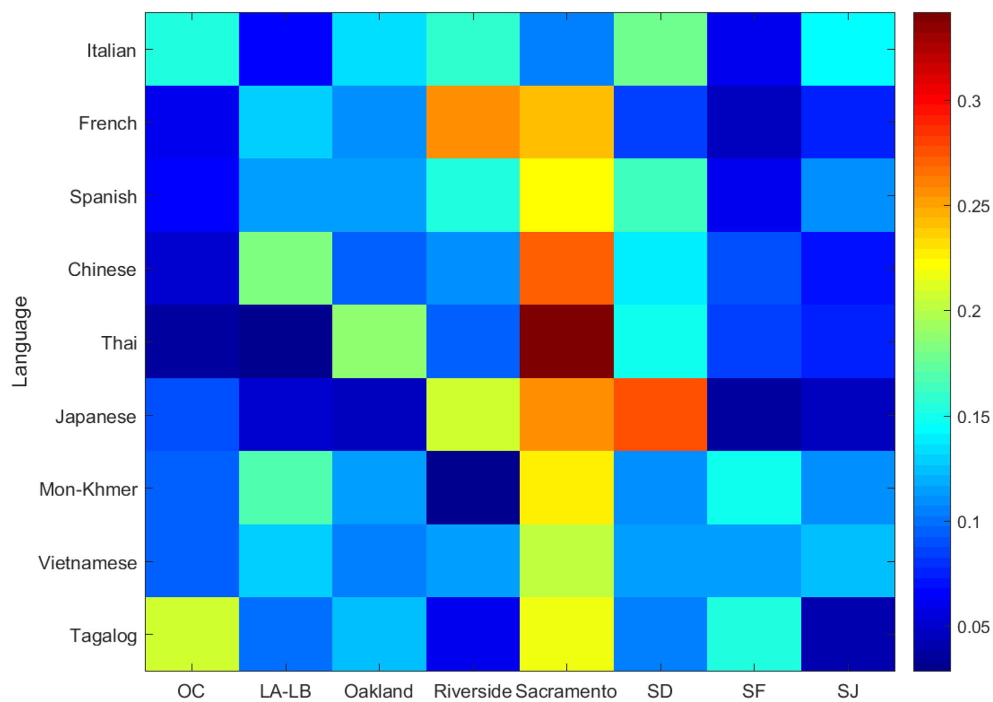
Note: Values on the x-axis is a woman's probability of participating in welfare programs. About 13.7% of all samples have participation probability greater than 10%.

Figure 2: Network Measure across Locations/Language Groups



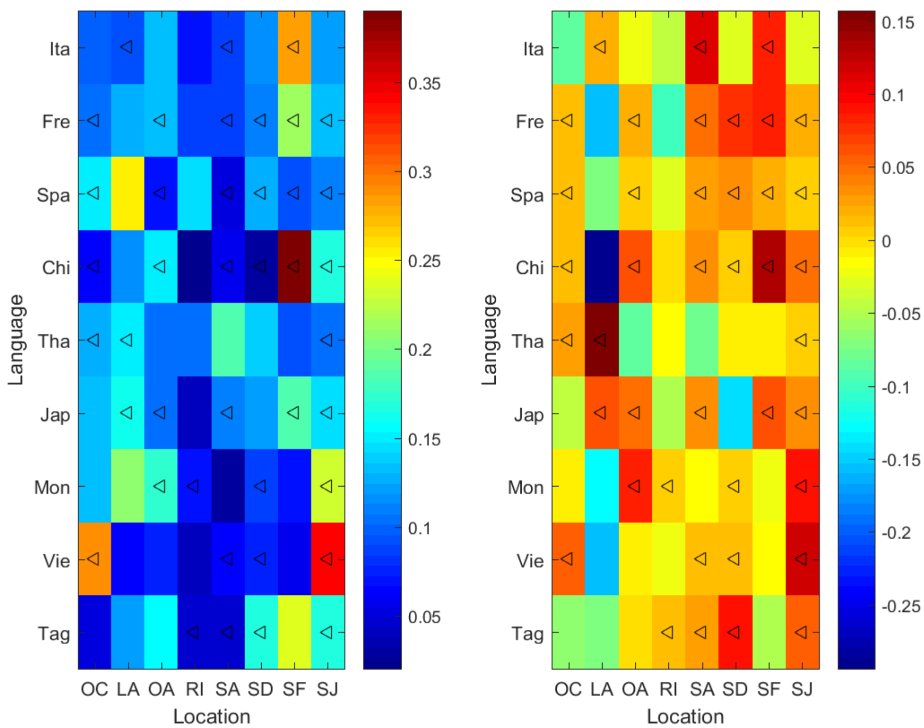
Note: The color in each cell corresponds to the relative density of each language group in each MSA, normalized as the percentage of the sum of the relative density of each language group. The redder color indicates stronger networks, whereas bluer color indicates weaker networks.

Figure 3: Percentage of Participants across Locations/Language Groups



Note: The color in each cell corresponds to the normalized rate of welfare participation for each language group in each MSA. The relative rate of a neighborhood is calculated as its welfare participation rate divided by the sum of welfare participation rates of all MSAs.

Figure 4: Counter-factual of Migration and Welfare Take-up Pattern



Note: The figure on the left is the same figure as Figure (2) with triangle indicating increase in the share of welfare participants if the heterogeneous welfare facility of neighborhoods is removed. The figure on the right shows the magnitude of the change.