Love Thy Neighbor? – Carpooling, Relational Costs, and the Production of Social Capital*

Kerwin Kofi Charles
University of Michigan
kcharles@umich.edu

Patrick Kline
University of Michigan
klinep@umich.edu

November, 2001

Abstract

This paper argues that individuals are more likely to have social capital the greater the incidence of people in their neighborhood who share certain traits which affect the ease and nature of social interaction. We argue that race and language are examples of such relational traits. The paper tests this prediction using an indicator of social capital never previously studied: whether someone uses a carpool to get to work. This measure retains nearly all of the strengths of previously used measures, and is free of most of their weaknesses. Analysis is conducted on a merged data set, with individual level data drawn from the 1990 IPUMS Census extract, and information on neighborhoods (PUMAs) derived from the 1990 Census STF3 tables. The model's predictions are confirmed for both race and language.

* We thank Robert Axelrod, Rebecca Blank, John Bound, Charles Brown, Mary Corcoran, John Dinardo, Jeff Dominitz, Glenn Loury, Gary Solon, Melvin Stephens Jr., and David Thatcher for comments and useful conversations. Correspondence to Charles at 408 Lorch Hall, 611 Tappan Street, Ann Arbor MI, 48109.

1. Introduction

The idea that many sociological and economic outcomes are determined not only by market forces, but also by factors related to the nature and quality of people's social, non-market interactions underlies the very active research program on "social capital". Sociologists and political scientists have long stressed social capital's possible importance.[1] In economics, theoretical work suggests that social capital facilitates cooperation, helping agents to avoid free-rider problems in repeated game interaction, and expensive legal and monitoring systems in their market activities. And, a number of studies by empirical economists document associations between social capital and positive economic outcomes across different communities and countries, and over time.[2]

A criticism of research on social capital has been that the phenomenon is usually vaguely and imprecisely defined in most studies, if it is defined at all. As a result, "social capital" runs the risk of being interpreted simply as the set of things the researcher cannot explain. In the context of empirical research, it is often not obvious how or whether the measure of social capital used in various studies correspond to the theoretical notion writers have in mind. Another criticism is that, despite much research on what social capital might or might not do, relatively little is known about social capital's "production function" - whether and why various factors determine if social capital exists.[3] The two criticisms are not unrelated; empirical analysis of determinants of social capital is impossible unless observable measures can be related to social capital in some precise way.

This paper attempts to address both questions raised by these criticisms. It studies the production of social capital, focusing on the role played by differences between individual and community characteristics in the creation of individual level social capital. To empirically assess these effects, it examines an individual level behavior which, almost by definition, varies with possession of the type of social capital the paper explicitly defines in an *a priori* obvious and necessary fashion.

---

[1] See Putnam (1993) and Coleman (1990).

[2] Theoretical work on social capital in economics probably begins with the seminal work of Loury (1977). See Grief, (1993), Abreu, (1998), Fudenburg and Masken, (1986), and Kreps et al. (1982) for results from the theory on repeated games. Arrow (1972) discusses how cooperation can lower transactions costs of economic activity. Important recent empirical pieces in economics include Knack and Keefer, (1997), La Porta et al., (1997), Putnam, (1993), (1995), (2000).

[3] See Durlaff (1999) and Portes and Landolt (1996) for a discussion of these and other criticisms. Glaeser et al (2000) discuss the relative sparseness of the literature on the production of social capital.

The focus on individual-level social capital separates our paper from most previous work which tends to focus on social capital measured at an aggregate level, such as the state or country.[4] We study individual social capital - an emphasis Glaeser et al (2000) call an "economic approach to social capital" - for two reasons. First, it is natural for economists to focus on the determinants of individual level behavior since rational decisions can only be made by individuals. Second, we believe that social capital can only exist for individuals; aggregate social capital is merely formed out of the different levels of social capital possessed by individuals. However, the way we model individual social capital emphasizes the phenomenon's fundamentally interactive nature. Our framework emphasizes that an individual's social capital operates and exists only *in relation* to other people. Thus, unlike human capital, for which an individual's investment decisions are affected only by his *own* characteristics, social capital investment is affected by the characteristics of the other people in a person's given sphere. The importance of the interaction between own and community characteristics is absent from work of Glaeser et al (2000) and others who model social capital as another type of human capital, determined only by individual characteristics such as wages, or age.

The paper uses an indicator of social capital never previously studied in the literature: the probability that a working man carpools to work. We believe that this measure retains all of the attractive features of previously used indicators, but is also free of most of their weaknesses.

The two most commonly used empirical measures of social capital in the previous literature are "trust" and "organizational membership". The "trust" variables used by many previous authors are derived from survey questions in which people are asked how much they trust others.[5] These questions measures latent sentiments, and economists have historically eschewed empirical strategies that rely on reports of latent emotions, or beliefs, preferring instead to focus on measurable behaviors.[6]

---

[4] See Putnam (1993), (1995), (2000); Knack and Keefer (1997); La Porta et al., (1997); Guiso et al., (2000); and Hall et al., (1999) for example.

[5] Fukuyama (1995), Guiso, Sapienza, and Zingales (2000), Knack and Keefer (1997), La Porta et al. (1997) and Putnam (1993; 2000) all trust measures in their papers. An example of the type of question from which this information is derived is Knack and Keefer (1997), whose measure of trust is from the question: "Generally speaking would you say that most people can be trusted, or that you can't be too careful in dealing with people?"

[6] Economists are not unique in this regard. Putnam (1995) says of trust that its centrality to social capital theory makes it ''.. desirable to have strong behavioral indicators of trends in social trust or misanthropy. I have discovered no such behavioral measures.'' Partially confirming the traditional concern, there is some evidence that reports of trust do not translate into trusting behavior. Glaeser et al. (2000) find that attitudinal surveys do not predict trusting behavior particularly well; survey questions on trust are only moderately correlated with an individual's trustworthiness. Interestingly, they also find that an individual's trust and trustworthiness vary with respect to the characteristics of the people with whom the person interacts. For instance, trust falls when individuals of different races or nationalities interact.

Given these problems with "trust" measures, some researchers have turned to another indicator – a survey based measure of the different organizations to which people belong.[7] Belonging to an organization or a club is an action, and it is often a *social* action, in that clubs bring people into contact with others. It may also be true that, as Putnam (1995) argues, organizational membership is related to trust since "people who join are people who trust." But there are reasons to be concerned about this measure as well, though these have rarely been emphasized in the literature. For one thing, the organizational membership questions in U.S. data often measure only the different *types* of organizations to which a person belongs. [8] Thus, membership in twelve benevolent societies on the one hand versus membership in a single social club are coded as the same thing: membership in a single type of club.

Even when there is information on both the number and types of clubs to which people belong, belonging to a club does not always foster social interaction.[9] And, interaction in a club need not occur among individuals in the particular sphere implicitly being studied.[10] Finally, the presumption that association between organizational membership and high social capital may simply be false. This is so because people may join organizations because the social capital *they already have* is low. People who join dating clubs are probably brought into contact with other people. But on the other hand, such people likely do not have a large circle of friends and acquaintances. If they did, meeting people to date using their stock of social capital would not be at all difficult, and there would be no need to rely on the benefits of a formal club.

The relative attractiveness of carpooling as an indicator of the social capital an individual possesses is clear. Carpooling is an action and not a report of a latent sentiment. Because carpoolers travel regularly to work with at least one other person, we can be sure that they know at least one person well. People likely carpool with those they already know and trust, for who would form this type of agreement with someone who might or might not show up on the day that it was his turn to drive or whose driving was careless? A carpool is a type of organization, but it is the type whose members *must* spend time together; anonymously paying dues or attending

[7] See DiPasquale and Glaeser, (1999), Maluccio, Haddad, and May (2000), Putnam (1993), (2000), Alesina and LaFerrara (2000).

[8] Most papers on social capital by economists in the U.S. use data from the same data source – the General Social Survey, or G.S.S.

[9] For example, Neighborhood Association Club Association clubs often require that a person wishing to join so indicate by filling out an application and paying dues. In return that person will get a local phonebook and possibly access to the local recreational center. Many people in a given community could belong to this club, and yet still remain quite socially isolated.

[10] Social capital at the level of the neighborhood, for example, is likely very poorly proxied for by information about membership in college Alumni Associations, given that the other members of this club will be scattered all over the country.

meeting sporadically does not suffice. Also, since the entire point of being in a carpool is to lower the various costs associated with travel, people in a carpool likely live in the same neighborhood[11] so it is possible to be quite explicit about the geographic sphere over which we expect carpoolers' social capital to operate.[12] Finally, carpooling is an activity of independent interest. Many communities have instituted carpool lanes and offered other inducements for residents to engage in this activity because of concerns about pollution and traffic congestion and commuter time increase. A better understanding of the determinants of this behavior is of substantial public policy interest.

The paper presents a simple theoretical framework in which an individual's stock of social capital is formed out of the different investments made in his separate pair-wise connections with other people. Investment in a particular pair-wise connection is easier if both persons *share* particular characteristics which affect the ease, frequency and nature of social interaction. We call such characteristics relational traits, and we focus on a person's race, and the language he speaks.

A simple prediction follows from model: people are more likely to have social capital with those who live close to them, the greater the incidence among their neighbors of people who share their relational traits. Thus, if there is a neighborhood A, identical to another neighborhood B, except that the incidence of persons of racial group 1 is larger in neighborhood A, while the incidence of persons of racial group 2 is larger in B, the social capital among type 1 persons should be larger in neighborhood A than in neighborhood B, and the social capital of type 2 persons in B should be larger than in A. Individuals' social capital is not observed, but we see individual carpooling behavior which should vary positively with social capital. An empirical test of the model is thus whether the difference in carpooling among people of racial group 1 between neighborhoods 1 and 2 is larger than the difference in carpooling among persons of racial group 2 between neighborhoods A and B.

We conduct the double difference tests described above for all possible pairs of racial and language groups using individual level data from the 1990 Census IPUMS, matched with community information from all 1726 PUMAs in the U.S. Our results control for latent state effects, and we employ a simple method to account for the problems which residential racial and language segregation pose for the estimation technique. Overall, the empirical results strongly support our model's predictions for both race and language.

[11] A survey of commuting behavior in the California Bay Area, which indicates that carpoolers spend an average of only 4.8 minutes picking up other passengers, confirms this.(DOT (1996)).

[12] Researchers have long been interested in examining neighborhood level social capital, starting with Jacobs (1961) and stretching to modern urban economists such as Glaeser and Sacerdote (2000).

The empirical approach employed here is similar in spirit to Borjas' (1995) study of ethnic capital, and to Erzo's (2001) attempt to assess interpersonal utility by studying support for welfare benefits. While clearly related to the literature which relates individual behavior to a summary index of community heterogeneity or diversity (Alesina and LaFerrara (2000), Costa, 2001), this paper is quite different in that it attempts to separately identify the effect that any particular shift in aggregate relational traits will have upon an individual with a particular trait. This provides us with detailed information about the magnitude of relational costs between each pair-wise grouping of relational traits and avoids the dangerous pitfall of conflating all forms of homogeneity together. In addition, we examine potential nonlinearities in the effects of these relational shifts that could not be dealt with by a single heterogeneity index. Nonetheless, the present paper obviously sheds light on the previous results.

The next section presents a theoretical overview that introduces key concepts and sets the stage for the empirical analysis. Section 3 discusses the basic empirical approach is greater detail. Section 4 discusses the data. Section 5 presents the initial results. Section 6 offers a discussion of the problems caused by racial and language segregation, and presents the modified results once these problems have been accounted for. Robustness tests are presented throughout. Section 7 concludes.

2. Theoretical Framework

2.1 The Production of Social Capital

We define social capital as the commodity which individuals use in non-market, social interactions to extract valuable and useful resources from each other. Let s_{ij}, $s_{ij} \geq 0$, be the amount of this commodity which an individual i possesses for exclusive use in social interaction with some different person j. The size of s_{ij} describes how much i can get from j in social interaction, and will in general differ from what he can get from a different individual.[13] We argue that all forms of social capital ultimately derive from these pair-wise connections.

Rather than the innumerable pair-wise social capital connection than an individual i possesses, we could focus instead his social capital *stock*, as measured against a particular universe, U. One individual stock measure is, $S_i^U = \sum_j s_{ij}, \quad j \in U$. Another is ϕ_i^U, which is a

[13] In principle, these pair-wise connections could be negative as well, as would occur if people sought to do each other ill when they interacted. We ignore this possibility in the paper. Also, we assume that the level of a pair-wise social capital connection is symmetric, so that $s_{ij} = s_{ji}$.

binary variable which equals 1 if the person has at least one non-zero pair-wise social capital connection with another person in the universe *U*. If *U* is "the neighborhood", the both S_i^U and ϕ_i^U are measures of a person's "neighborhood social capital stock".[14]

This framework highlights that a person with a very low social capital stock when assessed against a given universe, may yet have a high stock when measured against another. This distinction may be quite important, because different types of social capital stocks are probably of differentially import in different circumstances. An individual's global social capital may be important when he desires advice about where to send his son to college, while his neighborhood social capital is probably more important when he wishes that someone keep an eye on his house while he is vacationing. Given this possible distinction about different forms of social capital, empirical work should be explicit about the sphere in which the form of social capital under examination operates, so should focus on outcomes for which the particular type of social capital is important. In this paper we are interested in the individuals' neighborhood social capital stock, as summarized by the measure ϕ_i .

Understanding how social capital comes about, or what causes it to vary, necessarily requires some understanding of the determinants of the pair-wise connections, s_{ij}. We assume that the pair-wise social capital connection between two individuals i and j is determined by investments q_{ij} and q_{ji}, respectively, made by each of them. If, for a moment, we think of a pair-wise social capital as a "friendship", experience suggests that the formation of between two people, requires that expend effort in doing things like spending time with or getting to know the other person. The way in which these different investments combine to form a given social capital connection will likely depend on the particular form of social capital under discussion. For some, no pair-wise connection would arise unless both q_{ij} and q_{ji} were strictly positive. For others, the pair-wise connection may be the sum two investment levels q_{ij} and q_{ji}. In general, it is sufficient for our purposes to assume that

$$s_{ij} = f(q_{ij}, q_{ji}), \tag{1}$$

where $f_1 > 0$ and $f_2 > 0$.

An individual's investment in a pair-wise social capital connection will, like all human capital investments, depend on particular benefits and costs, with the level of investment rising in

[14] Most of the previous literature focuses on social capital measured at the aggregate, or community, level. This too is readily captured in our framework. For a given region or community, aggregate global social capital is just the aggregation of the all the individual stocks of global social capital of the people who live in that community.

benefits and falling in costs. Given the necessarily interactive nature of social capital, these benefits and costs for an individual may be sorted into two types. There are first what we call *autonomous* benefits and costs. These are factors which affect an individual's incentive to make social capital pair-wise investments, irrespective of the other person with whom the connection is being made. For example, a person who is in the last year of his life is unlikely to make any kind of human capital investments, including investment in social capital connections, because of the small number of years over which any benefits can be recouped. Similarly, someone whose value of time is low, such as someone with a low wage rate, should be more likely to make all forms of human capital connections, including all pair-wise social capital investment.

The second type of benefits and costs which affect investment in a particular pair-wise social capital connection derive from the fundamentally interactive nature of social capital. These are *relational* benefits and costs, in their size depends on whether the two parties share particular characteristics. For example, holding constant factors like his age and his wages, an individual is more likely to make an investment in a pair-wise connection with a next-door neighbor than he is to invest in someone living at the opposite side of town. The smaller distance which separates the person from his neighbor makes it more likely that social capital connection between them will be used; a next door neighbor is best able to look at someone's house while he is away on vacation. Also, living close together means that the mechanics of making the pair-wise investment are probably easier. To invest in a relationship with a neighbor, one need merely lean across the back fence; a connection with the another person in the neighborhood requires a walk or a drive of some distance. Race and language are other relational traits which, if not shared by people, makes their social interaction rare, difficult to enact, or strained.

Individual social capital investment can be written

$$q_{ij} = q_{ij}\left(C_i, RC_{ij}\right) \tag{2}$$

where C_i represents net autonomous social capital investment costs, and RC_{ij} denotes net relational costs which affect person i's social interaction with person j. The "net" in both definitions summarizes the difference between costs and benefits of a given type. Obviously, investment in assumed to be strictly falling in both sets of net costs. Since these two types of costs are functions of observed characteristics,

$$q_{ij} = q_{ij}\left(X_i, R_i, \Delta R_{ij}\right) \tag{3}$$

and

$$q_{ji} = q_{ji}\left(X_j, R_j, \Delta R_{ij}\right) \tag{4}$$

In (3) and (4), the vectors X_i and X_j are, like R_i and R_j, vectors of observed individual level characteristics. We distinguish between an individual's X and R characteristics to highlight the

fact that the latter are relational. The vector ΔR_{ij} is a *k*-dimensional vector of 0's and 1's, with elements Δr_{ij}^k. Each element Δr_{ij}^k indicates there is a *difference* between persons i and j in a particular relational characteristic, R^k. Note, a given trait may affect investment through both an autonomous and a relational aspect.

It is easy to see that, *ceteris paribus*, the pair-wise connection between any individuals i and j is smaller when $\Delta r^k = 1$,

$$s_{ij}\left(.,\Delta r_{ij}^k = 1\right) < s_{ij}\left(.,\Delta r_{ij}^k = 0\right). \tag{5}$$

It follows naturally that an individual's stock of social capital defined against a particular universe U such as his neighborhood is smaller the share of fraction of people in his neighborhood who share his particular vector of relational traits.

This last point can be stated more precisely. Consider a relational trait, R, with K distinct categories $r_1,..,r_K$. Let the overall relational distribution of these traits in a person's neighborhood be $\Pi = \left(\pi_{iN}^{r_1},..,\pi_{iN}^{r_K}\right)$, where $\pi_{iN}^{r_k}$ be the share of person i*'s* neighbors who are of type r_k, and $\sum_k \pi_{iN}^{r_k} = 1$. Let $\phi_i^{r_j}$, $j = 1,..,k$, be the probability that a person of type r_j has at least one positive social capital connection with someone in his neighborhood. We assume

$$\phi_i^{r_j} = \phi_i^{r_j}\left(\pi_{iN}^{r_1},..,\pi_{iN}^{r_K}\right). \tag{6}$$

Because the distribution of relational traits sums to 1, a marginal increase in the incidence of a given relational trait is necessarily accompanied by a reduction in the incidence of some other trait. Because we wish to be specific about which group is being lowered so that one can be raised, we adopt the notation $\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}$ to represent these simultaneous partial changes: a marginal change in the overall neighborhood distribution caused by a marginal increase in share of π^{r_k} and a simultaneous decrease in π^{r_j}. We call changes such as $\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}$ marginal distribution shifts.

Our results about interactions and relational cost imply that

$$\frac{\delta\phi_i^{r_j}}{\Delta\pi_{iN}^{r_k} \rightarrow \Delta\pi_{iN}^{r_j}} > 0,\ j \neq k\ . \tag{7}$$

A distribution shift which raises the incidence of a person's own type in his neighborhood raises the likelihood that the person has at least one close social capital connection in the neighborhood. It seems reasonable to assume that this effect is concave, but strictly speaking the sparse framework we have presented does not necessarily imply this. More importantly, our framework

says nothing about how marginal distribution shifts which do not affect the incidence of a person's own trait in his neighborhood affects his social capital. Thus,

$$\frac{\Delta \phi_i^{r_j}}{\Delta \pi_{iN}^{r_k} \rightarrow \Delta \pi_{iN}^{r_m}} = ?, \quad j \neq k, m \neq k \tag{8}$$

The probability $\phi_i^{r_j}$ is not directly observed, but the individual level outcome carpooling is. We have argued that an individual's carpooling behavior should depend on whether he has a non-zero social capital connection with someone in his neighborhood. Of course, individual carpooling will depend on many things, quite apart from social capital connections. And, the distribution of relational traits among a person's neighbors likely affects his carpooling behavior for reasons having nothing to do with social capital. In general, then, whether an individual of relational type r_j carpools to work may be written,

$$CP_i^{r_j} = CP_i^{r_j}\left(X, \phi_i^{r_j}, \nu(\Pi_{iN})\right) \tag{9}$$

In (9), X is a vector of factors like distance to work, wages, occupation, and family structure, measured at both the individual and neighborhood level. The function ν summarizes the ways in which the relational distribution in a person's neighborhood affect his carpooling, independent of its effect on his social capital.

The assumption that the probability that a person of relational type r_j carpools is an increasing function of the probability that he shares a non-zero social capital connection with at least one person in his neighborhood implies

$$\frac{\partial\left(CP_i^{r_j}\right)}{\partial \phi_i^{r_j}} = \beta > 0 \qquad \forall\, j \tag{10}$$

The fact that we unambiguously sign the effect of social capital on carpooling behavior distinguishes our paper from other studies which use other observable outcomes with which social capital is allegedly correlated. We have argued above that even if there is a correlation in some of these other cases, the sign of the relationship is not *ex ante* obvious for some of the most commonly used measures in the literature.

Carpooling is assumed to varied in a completely arbitrary fashion with changes in the function ν. Thus, the effect of any given distribution shift on carpooling, is ambiguous for all levels of overall neighborhood distribution Π. Thus, we assume

$$\left.\frac{\partial CP_i^{r_j}}{\partial \nu(.)}\right|_{\Pi} = ? \tag{11}$$

The effect of a given distribution shift affects ν and therefore carpooling in a completely unknown way, but in a way that is same for all persons.

We can now inquire about the effect on observed carpooling of different types of distribution shifts. If the overall distribution in the neighborhood is initially $\overline{\Pi}$, then by (7)-(11), the effect of the distribution shift $\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}$ on carpooling for an individual of type r_a is

$$\left.\frac{\Delta CP_i^{r_a}}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}} = \beta \left.\frac{\Delta\phi_i^{r_a}}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}} + \left.\frac{\partial CP_i^{r_a}}{\partial\nu}\right|_{\overline{\Pi}} \left.\frac{\Delta\nu}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}}. \qquad (12)$$

Because the last two terms in (12) are ambiguous, the entire expression is of ambiguous sign for all individuals, irrespective of their relational type. However, notice that the *difference* between expression (12) for $a = k$ and $a = j$ is

$$\left.\frac{\Delta CP_i^{r_k}}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}} - \left.\frac{\Delta CP_i^{r_j}}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}} = \beta\left[\left.\frac{\Delta\phi_i^{r_k}}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}} - \left.\frac{\Delta\phi_i^{r_j}}{\Delta\pi^{r_j} \rightarrow \Delta\pi^{r_k}}\right|_{\overline{\Pi}}\right] > 0 \qquad (13)$$

We know that the difference measured in expression (13) is positive, since by, the first term in the squared bracket is positive, while the second negative. This result, which is the fulcrum of all of the work which follows, can be explained intuitively.

There are two effects on individual carpooling when there is the marginal change in the distribution of different relational types within a person's neighborhood. One effect, of indeterminate sign, and having nothing to do with social capital, changes carpooling probability equally for all persons. The second effect changes carpooling probability by changing the likelihood of having a non-zero social capital connection in the neighborhood. If the distribution shift lowers the representation of a given type in the neighborhood, social capital among people of that same type is lowered. Those with whom they "get along" best are now less well represented among their neighbors. If the distribution shift raises the representation of some other type, then by the same argument, people of that other type should be unambiguously more likely to know someone well in their neighborhood. If we therefore compare the *change* in carpooling for people who see the representation of their own type in the neighborhood rise, to the *change* in carpooling for those who see the representation of people of their own type in the neighborhood fall, the difference between the first and second change should be positive. Importantly, this prediction applies *only* for people for whom the distribution shift changes the neighborhood representation of persons of their own type. We re-state expression (13) as the proposition *P*:

> *P*: If there is a marginal distribution shift within a neighborhood, with an initial overall racial distribution of a neighborhood, Π, so that there is a slight reduction in the fraction

of people who are of type r_1, and a simultaneous small increase in the fraction of those of type r_k, the *change* in carpooling probability for people of type r_k should be greater than the *change* in carpooling probability for people of type r_1.

The remainder of the paper is devoted to testing this prediction of the relational cost model. The next section discusses our empirical approach more formally. Subsequent sections discuss the data used to test the empirical models, a discussion of a potential shortcoming of the basic empirical approach, a simple method to deal with this problem, and results.

3. Empirical Set-Up

The empirical work in the paper attempts to estimate the various differences in (13) for different types of distribution shifts. Of course, we cannot actually conduct the various experiments that would yield prediction *P*. That is, we cannot create distribution shifts within neighborhoods and then observe how carpooling behavior changes (relatively) for the different people who comprise those neighborhoods. Instead, our empirical strategy relates differences in the actual racial distributions across different neighborhoods, to observed relative differences in carpooling behavior, by race and language.

We illustrate our approach using race. Assume that there are three races - either White (*W*), Black (*B*) and Hispanic (*H*). There are six possible three different distribution shifts: $\pi^B_{iNs} \rightarrow \pi^W_{iNs}$, $\pi^H_{iNs} \rightarrow \pi^W_{iNs}$, $\pi^B_{iNs} \rightarrow \pi^H_{iNs}$; and three shifts going in the other direction of the ones indicated. But since the effect of the latter three shifts is simply the opposite of the three shifts shown, we need only focus on this three.

Consider the empirical equation,

$$\begin{aligned} CP_{iNs} = \Gamma X + \delta_s s + \alpha_W W_i + \alpha_B B_i + \lambda_W \pi^W_{iN} + \lambda_B \pi^B_{iN} + \\ W_i\left(\gamma_{WW}\pi^W_{iN} + \gamma_{WB}\pi^B_{iN}\right) + B_i\left(\gamma_{BW}\pi^W_{iN} + \gamma_{BB}\pi^B_{iN}\right) + \varepsilon_{iNs} \end{aligned} \tag{14}$$

in which CP_{iNs} is an indicator variable which measures whether individual i in neighborhood N in state s goes to work by carpool, and ε_i is a random error term. The vector X is a set of individual and community level observable determinants of carpooling, s is a state effect and ε_{ips} is a random error. The binary variables W_i and B_i indicate whether the person is White or Black. Of course, the expressions π^W_{iN} and π^B_{iN} measure, respectively, the fraction of people in person i*'s* neighborhood who are White, and the fraction who are Black.

The assumption implicit in specification (14) is that two neighborhoods within a state which differ only with respect the relative prevalence of two of the three races, may be thought of as the before and after distributions of a neighborhood which has undergone a distribution shift. Thus, to test the prediction *P* for the distribution shift $\pi_{iNs}^{B} \rightarrow \pi_{iNs}^{W}$, our approach compares the extent to which there are carpooling differences between Blacks and Whites in neighborhoods with different Black and White relative populations, *within the same state*. From (14), the difference between the carpooling behavior among Whites who live in the neighborhoods with relatively more Whites, compared to carpooling among Whites who live neighborhoods with relatively fewer Whites but more Blacks, is

$$(\lambda_W + \gamma_{WW}) - (\lambda_B + \gamma_{WB}). \tag{15}$$

And, the carpooling difference between Blacks in the same neighborhoods is

$$(\lambda_W + \gamma_{BW}) - (\lambda_B + \gamma_{BB}). \tag{16}$$

The quantity $\lambda_W - \lambda_B$ is common to both (15) and (16), and corresponds to the ambiguously signed effect $\left.\frac{\partial CP_i^{r_a}}{\partial \nu}\right|_{\bar{\Pi}} \left.\frac{\Delta \nu}{\Delta \pi^{r_j} \rightarrow \Delta \pi^{r_k}}\right|_{\bar{\Pi}}$ from the theoretical discussion. The implication of prediction *P* is that the difference between (15) and (16) is positive. By similar reasoning, it is easy to show that prediction *P* also implies that, for the distribution shift $\pi^H \rightarrow \pi^W$ the coefficient $\gamma_{WW} > 0$, while for the distribution shift $\pi^H \rightarrow \pi^B$, the coefficient $\gamma_{BB} > 0$. These results suggest that to test the relational cost model in this simple three race example, one need simply estimate (14) on a random sample of persons in the United States, drawn ideally from all "neighborhoods" in the country. The estimated coefficients from this regression can be used to test whether:

$$\begin{gathered} (\gamma_{WW} - \gamma_{WB}) - (\gamma_{BW} - \gamma_{BB}) > 0 \\ \gamma_{WW} > 0 \\ \gamma_{BB} > 0 \end{gathered} \tag{17}$$

The three difference-in-difference test the prediction *P* in the case when the relation trait being studied has three distinct categories. In general, if the relational traits has n categories, estimation of an appropriately modified version of (14) will produce $\binom{n}{2}$ double difference estimates such as those in (17).

We estimate versions of (14) and then conduct the tests on the double difference estimates such as those in (17) on a sample described in detail in the next section. The models are run as

linear probability, so the error term ε is heteroskedastic. In addition, the error term almost surely does not satisfy the classical assumption that is it demonstrates no systematic correlation across different individual observations. For one thing, there will be unmeasured factors that are shared by many individuals in a given neighborhood. Also as Moulton (1990) has shown for individual-aggregate regressions of the form (14), because of correlation in the regressors for people from the same neighborhood, failure to correct for this will result in a biased estimate of the standard errors. Indeed, since people from the same neighborhood all have *exactly* the same distribution of neighborhood characteristics in the data, failure to deal with this problem would lead to severe bias in our study, especially for the key variables of interest. To control for these problems, the standard errors presented in the paper allow for clustering (arbitrary correlation of the errors) within each neighborhood, and are also corrected for heteroscedaticity.[15]

4. Data

The paper studies two relational traits – race and language. Why language is a relational trait is clear; it is mechanically difficult for people who speak one language to socially interact with persons who speak another. The argument for race as a relational trait is not as mechanically obvious, but a massive literature in the social sciences takes it as axiomatic that, in the United States, a person's race seriously circumscribes his social interactions. The data requirements for implementation of the empirical strategy outlined above are severe. The ideal data set would have individual level carpooling, race, language, and determinants of carpooling. It would also have information about the neighborhoods in which particular individuals live – most notably the racial and language composition of those neighborhoods, and neighborhood level determinants of individual carpooling. Observations on individuals from multiple neighborhoods within the same state would also be ideal. Finally, the ideal data source should be large, with enough observation on the smaller race and language groups to permit statistically meaningful comparisons involving these groups.

The individual level data in the paper are drawn from the 1% IPUMS Unweighted Sample of the 1990 United States Census. We restrict attention to working men aged 18-64. The Journey to Work portion of the 1990 Census asked working persons age 16 and above whether they usually traveled to work by car, truck, or van. If so, they were then asked how many people usually drove to work in the car, truck, or van with them.[16] Carpoolers in our study are defined as

[15] The regressions are estimated using the Stata "cluster" subroutine.

[16] If the person was driven to work by someone who then drove back home or to a non-work destination they were instructed to report "drove alone."

those men who usually went to work by car with at least one other person. The IPUMS data provides detailed information about wages, occupation and industry, time to work, and the number of cars available in the household– all likely important determinants of individual level carpooling, which we control for in all of the regressions.[17] In addition, we use information on family structure to control for a man's marital status and family size in the regressions. At first blush, this last set of controls might appear quite important, because what we classify as carpooling may simply be people going to work with a spouse. Failure to control for marital status and family size, to the extent that there are differences in these outcomes across different races or language groups, could lead to biases in our estimates. Detailed controls for these variables are included in our base specification. Further, we conduct robustness tests and present other evidence below that shows that our results are not driven by any systematic mis-measurement of carpooling associated with going to work with a spouse.

Of course, neither the IPUMS data nor any other data source provides a completely satisfactory description of a man's "neighborhood." The IPUMS data provides three pieces of information about respondents' spatial location – the man's state of residence; his metropolitan area (MA), and a geographic region called a Public Use Microdata Area (PUMA) in which the man resides. We eschew the MA in favor of the PUMA as the definition of a man's neighborhood in this paper for three reasons. First, PUMAs are much smaller than MAs, and therefore much more closely connected to conventional notions of what a neighborhood likely is. The median size of an MA is 229,290 people (2,932,707 acres) while the median size of a PUMA is only 123,936 people (667,440 acres). Second, not every IPUMS respondent is attached to an MA, whereas every person is matched to a PUMA.[18] Third, unlike MAs, PUMAs (from the state sample) do not cross state boundaries, so it is possible to account for unobserved state fixed effects. Fourth, there are more PUMAs than MAs providing us with more aggregate variation.

Data on the aggregate characteristics of PUMA's is constructed from an additional data source. The Census collects aggregate information about more than 200,000 geographic units, called "block groups", out of which most other levels of aggregate census geography are constructed. These data for 1990 are reported in the 1990 Census ST3F tables. By and large, block groups do not cross PUMAs boundaries, so we construct aggregate level PUMA

[17] Additional details about these variables may be found in the data Appendix.

[18] The Census defines an MA as a group of adjacent communities with a large population nucleus that have a high degree of economic and social integration. Each MA must contain either a Census designated "place" (i.e. city) with a Minimum population of 50,000 or a Census designated Urbanized Area with a population of at least 100,000. Because many areas do not meet these requirements there are only 342 MAs nationwide. These MAs hold about 77% of the total population of the United States but only about 16.5% of the total land area.

characteristics from the means reported in the ST3F tables of block groups within that PUMA.[19] When data from the IPUMS is merged with the PUMA data, we have a sample consisting of an observation for each working man in the IPUMS sample, and aggregate information – including the racial distribution and language distribution – of the PUMA in which that man resides. Our primary data set has observations on more than half a million working men between 18 and 64 drawn from 1726 PUMAs, drawn every state and the District of Columbia.

Even though the PUMA is the smallest geographic location to which an individual in the Census can be traced, there may still be the criticism that a PUMA is certainly larger than the areas that people view as their neighborhood. Yet, even if there were available information at some smaller geographic level, there would still be a strong argument for using PUMA data. The reason is concern about Tiebout sorting.[20] If people choose their neighborhoods carefully, with an eye to the ease with which they can get along with them, then the effect of variables summarizing the characteristics of those immediate neighbors on individual behavior would be endogenous. One way to deal with this problem would be to instrument for the neighborhood characteristics. A very nice set of instruments would be the characteristics of the geographic area a few levels larger than the small neighborhood. As such, PUMA characteristics are ideal candidates.

Our analysis focuses on measuring the effect of shifting the neighborhood distributions for various races and languages. Race is divided into five groups: White, Black, Asian-Pacific, Hispanic, and Other. White, Black, and Asian were defined according to the usual census criteria; the definition of Hispanic and Other, however, require some extra discussion. The census does not officially define Hispanic as a race. Individuals who fill out the census are asked to choose from five categories: White, Black, Asian, Native American, and Other. In a separate question they are asked whether they are of Hispanic Origin or not, and if so which national origin they are from. In order to avoid confusing race with ancestry (or ethnicity), we decided to classify Hispanics as those individuals who indicated that they were not White, Black, Asian, or Native American, but were of Hispanic origin. We then lumped together Native Americans and Non-Hispanic Others because of their small sizes into one category that will simply be referred to as Other henceforth.

Language is divided into 7 categories: English, Spanish, French, Italian, German, Chinese, and Other. Other is simply defined as any language other than English-Chinese. Language refers to what individuals report as the dominant language spoken at home. Because

[19] Census block group data was matched with PUMA identifier using CIESIN's online Master Area Block Level Equivalency engine at http://plue.sedac.ciesin.org/plue/geocorr/.
[20] See Erzo (2001) for an excellent discussion of similar concerns.

there are undoubtedly many people who may speak English in addition to the other language that they speak at home, the sample probably contains many bilingual people who are assigned to single non-English language. In this sense, language does not provide for the neat, distinct classifications that are possible for race. However, the presence of bilinguals should bias our results *against* finding a role for the incidence of neighbors who speak one's language, when that language is something other than English. In effect, the relevant "own" relational group for bilinguals mis-measured; we focus only other who speak that language, missing the fact that bilinguals should be able to interact with people who speak English just as well. If we find significant effects even in the face of the resulting attenuation bias, we can be reasonably convinced that there is a true effect.

In addition to using aggregate census data to construct race and language composition variables, we also compute race and language segregation variables. The intuition being that we are interested in the effect of changes in PUMA composition controlling for intra-PUMA segregation. Using the well-known dissimilarity index[21], which measures the percent of the population that would have to move in order to obtain an even racial distribution, we calculate segregation levels for each PUMA using the variation among each PUMAs block groups.

Table 1 lists the means and standard deviations of the key variables from the matched IMPUS sample. Note that under our definition of race Whites comprise 83.6% of the individual observations and PUMA's are, on average, 81.3%, white. Hispanics only comprise 3.9% of the individual observations and the mean percent Hispanic of the PUMAs is 3.7%. This distribution is somewhat different than what we would see under other definitions of Hispanic and White. This difference derives from our desire to distinguish between race and ancestry or ethnicity.[22]

The means and standard deviations of the other variables in this summary table, except for individual level carpooling, should be very familiar. The key point is the very detailed body of information for which our empirical analysis controls. The table shows, by our preferred definition, travelling to work with at least one other person, 13.4% of working men carpool to get to work. Under more restrictive definition, in which carpooling is said to occur when there are at least 2 other people in the car, the frequency of carpooling falls to 3.1%. Raising the cutoff to three or more other passengers in a car (a very high cutoff) lowers the incidence of carpooling to fall even further to 1.3%.

Table 2 summarizes carpooling for each racial and language group. The table shows that there is substantial variance in carpooling by both race and language. Hispanics carpool the most

[21] See Sakoda (1981) for more on dissimilarity indices.
[22] We attempted alternative definitions of race, and the results are essentially unchanged.

and Whites the least, irrespective of the definition of carpooling. Indeed, Hispanics tend to carpool about four times as much as Whites. For languages, Spanish speakers are shown to carpool the most, and Italian speakers carpool the least, with English speakers barely above the Italians. Since most Hispanics (under our definition) are likely to speak Spanish, the strong similarity in carpooling for the two groups is not surprising. The results in this table indicate that these different groups may have different propensities to carpool. Alternatively, this variance in carpooling across different groups may derive from differences in other observable factors. The next section analyzes carpooling more fully in a regression context, and presents the results of the tests described above.

5. Initial Results

Some Base Results

Before turning to main analysis conducted on the Census IPUMS sample, we present some results on carpooling from the 1995 wave of the National Personal Transportation Survey (*NPTS*). The *NPTS* is a nationally representative sample of the civilian, non-institutionalized population of the United States, consisting of about 42,000 households (95,000 people). The survey contains data on household access to public transportation and the usual driving patterns of household members. By the standards of the Census, the NPTS is tiny, with very few observations on smaller racial groups. It has no information on the language a respondent speaks, nor is there the rich neighborhood information contained in the merged Census data set. We cannot use this data set to test the relational cost model. On the other hand, since the *NPTS* is designed to permit very detailed analyses on transportation decisions, the quality of its information about carpooling and its determinants is unmatched. An analysis of these data therefore helps highlight the strengths and limitations of the Census data used in our study.

The NPTS contains a question on carpooling virtually identical to the one in the IPUMS. Respondents are asked “Do you usually drive to work alone or do you carpool?” According to the NPTS, 11.4% of employed men between the ages of 18-64 usually carpool to work. This corresponds to 13.4% from the Census who drive to work with at least one other person. The NPTS asks people who do not carpool why they do not, and 62.74% of these non-carpoolers indicate that one of their reasons for not doing so is that they don’t know anyone to carpool with. This offers very strong support for our proposition that social capital is a strong determinant of carpooling. Another section of the NPTS – the “trip” section – is a sample consisting of individual trips taken by respondents. The “trips” file means indicate that for 64.01% of the “trip”

labeled as "trips to work" in which more than one passenger traveled, the passengers belonged to different households. This number rises to 76.43% for "trips to work" with more than three people. Data in the trip file are collected at the level of the trip and not at the individual level so it is possible that one individual contributes more than one observation to this sample. Nonetheless, the means from this sample suggest that the people we call carpoolers are doing what we claim – riding regularly to work with someone from the neighborhood. Moreover, we directly control for family size and marital status in the regressions. We also find virtually the same results when we do robustness tests in which carpooling is said to occur only when the number of drivers in the pool is large.

Finally, there is a concern that many factors which affect carpooling are simply not inquired about in the Census. For example, the decision to carpool might be affected by the proximity of bus or subway stops. In principle, the absence of this information is not important in our analysis, unless these other putatively important determinants of carpooling are systematically related to PUMA racial and language distributions across neighborhoods. But a first order question is whether factors not available in the Census data affect carpooling behavior. We use the NPTS to analyze the effect on carpooling of variables not available in our primary data set. Appendix table 1 reports results for a linear probability model of carpooling behavior using the NPTS data, and the different variables available in that survey. Reassuringly, the results indicate that the availability of public transportation variables which are absent from the Census data do not appear to be a large determinant of carpooling. Among controls for the availability of public transportation, only street car service is statistically significant. By contrast, the variables which according to the *NPTS* have the most important effect on carpooling - the minutes to work, individual income, the number of cars in the family, and individual education – are all measures which are available in our Census data at an even greater level of detail.

Table 3 begins our formal analysis of carpooling in the Census sample. This initial table does not test the model. Rather, it presents the results of a base specification, in which carpooling probability is related to detailed individual and community level characteristics which might be related to the distribution race and language within neighborhood. The variables in this base specification are included in all of the subsequent regressions in which the effects of shifts in neighborhood distribution are studied.[23] The results are linear probability estimates, with standard errors corrected for heteroskedasticity and clustering by PUMA. The sample consists of observations on 496,280 working men, drawn from all 50 states and the District of Columbia, and

[23] See the Data Appendix for a detailed description of these variables and our reasons for including them.

from all 1,726 PUMAs. Alternative definitions of carpooling, in which the behavior is said to occur only when more than a certain number of people are in the pool, are used as a robustness test.

Several interesting patterns are evident in the table. Young people are more likely to carpool, as are those who live in large households and those who are married. Controlling for the latter pair of variables means that we can be assured that later results present the effect of carpooling of neighborhood shifts, above and beyond any within-family joint carpooling. Not surprisingly, the likelihood of carpooling varies inversely with the number of automobiles in the household.

The results suggest that carpooling is a middle class phenomenon. The quadratic in annual earnings reveals an effect which is initially rising and then decreasing. The control for homeownership may well be picking up this same class effect, with carpooling shown to be lower for wealthy individuals who own their home.[24] Oddly, the same class pattern is not found for education. Recipients of bachelor's degrees carpool less than those with just high school degrees, but receiving more education than a bachelor's makes one more likely to carpool. Note that the effect of income on carpooling vanishes when carpooling is defined as riding with 2 or more people besides the driver, while the effect of education on carpool persists. This may indicate that there is no income effect on neighborhood social capital formation but merely an effect on intra-household carpooling.

Since the Census has no direct measure of linear distance to work, we use whether an individual works in the same PUMA in which he lives and how long it takes to get to work as proxies for distance. As we would expect, travel time has a very strong positive effect on carpooling. The potential savings in effort and resources from carpooling increase with trip size. The strong effect of working in the same PUMA in which he lives is an additional estimate of ths distance effect.

The base specification includes a rich vector of geographic controls to account for the fact that social interaction and commuting behavior might be qualitatively different in urban areas than in other places. In particular, it seems reasonable to suppose that having lots of people nearby makes interaction less costly.[25] Furthermore traffic patterns, as well as available public transportation services likely differ between cities and suburbs and rural areas. If certain populations such as Blacks and Hispanics are more urbanized on average than Whites, failure to

[24] This effect may indicate that the wealth effect is very strong, since we might also expect homeowners to be more socially connected (DiPasquale and Glaeser, 1999) and consequently more likely to carpool.
[25] Empirical studies by Festinger et al. (1950) and Glaeser and Sacerdote (2000) seem to support this notion.

control for these effects could lead to endogeneity problems for the variables of main interest in the subsequent regressions. There is no single, ideal measure of urbanity, so we use a variety of possible geographic controls.

The results show that PUMA Population Density has a negative effect on carpooling. This may be because people in denser areas are more likely to use public transportation. However, the dummy for living in an urbanized area is positively related to carpooling (especially in the more restrictive definitions of carpooling). Since this dummy is a weaker test of urbanization (see Data Appendix) and is really just a contrast to being rural, this may just indicate that carpooling is most prevalent in the suburbs. This certainly conforms to popular stereotypes and also makes sense because of the spatial organization of most major metropolitan areas which contain jobs in an inner core and large portions of the workforce in the suburbs. If suburban workers have longer commutes, as shown above, the returns to carpooling should increase. Suburban residents also face more direct incentives to carpool to urban cores in the form of federal highways and High Occupancy Vehicle (HOV) lanes that have minimum passenger requirements.

A particularly noteworthy set of controls in the base specification, given the tests we later perform, are the controls for individuals' industry and occupation, and for industry and occupation affiliation of workers in the neighborhood. To the extent that the distribution of occupation and industries among the workers in a neighborhood are related to the racial distribution in that neighborhood, a failure to control for both own and community level industry and occupation may lead us to incorrectly attribute any effects found for neighborhood racial and language distributions to social capital, rather than to the fact that people of the same race are simply going to the same place when they go to work and are thus more likely to carpool. The large number of industry and occupation effects, at both the individual and community level, makes it difficult to summarize the effect of these variables on carpooling. We simply note that, consistent with our concern, most of the estimated effects are strongly statistically significant. Their inclusion in all of the regressions raises our confidence that any effect we find for race and language composition of communities, above and beyond the occupation distribution in those communities, is truly a measure of a social capital effect, rather than the unmeasured propensity of people from the same race to be more likely to be going to the same workplace. We also control for the average time to work among workers in the community as an additional guard against this concern.

Initial Estimates of Relational Cost Effects

Having examined the base determinants of carpooling, we turn to a test of the paper's central hypothesis that a shift in the neighborhood distribution of race and language affect an individual's carpooling propensity differently based upon an individual's own race and language. We run the same base regression discussed above, but add the various neighborhood racial distribution variables, and the interaction terms as shown in (14). We have shown how functions of these interaction terms yield double-difference estimates which test the prediction *P* for different pairwise distribution shifts in neighborhood composition. Table 4 presents the estimated difference-in-difference effect, and tests for the significance of these effects. The tests are straightforward *t*-tests, since the functions being tested come from a single equation model. The first column of the table shows the shift in the relative neighborhood distribution to which the particular difference in difference estimate corresponds.

The complete regression results from which the coefficients are drawn in order to test for the various conditions of the model's main prediction are presented in the Appendix. We remind the reader that these regressions control for state fixed effects, and that the standard errors account for clustering at the level of the PUMA.

Overall, the difference-in-difference estimates support the model's prediction, so long as the groups being considered are racial minorities. For example, the table indicates that the effect of a distribution shift whereby a neighborhood is made marginally more Hispanic as a result of the lowering of the share of Blacks causes carpooling among Hispanics to rise relative to the change for Blacks. However, the estimated effect is only statistically when carpooling is defined as riding with three or more passengers. The results for the Black-Asian, and Asian-Hispanic distribution shifts are more encouraging. Each estimated effect is positive and statistically significant.

Unfortunately, the results are not at all encouraging for distribution shifts involving Whites. Every difference-in-difference estimate involving Whites, irrespective of the definition of carpooling, points the wrong direction. If they are to be believed, the estimated coefficients for the White-Black distribution shift, for example, indicates that when the share of Whites in a neighborhood rises because of a small reduction in the share of Blacks in that neighborhood, carpooling among Whites rises by *less* than it does for Blacks. This is exactly the opposite of what the model predicts. The results would seem to seem to suggest two possible explanations – neither of which is very plausible: that people from every minority group prefer interacting with whites relative to its own group, or that whites prefer interacting with minorities relative to other whites. Received anecdotal wisdom suggests that both of these hypotheses are highly dubious.

Table 5 presents the results for different pair-wise distribution shifts in the neighborhood incidence of different languages. The results in this table are also mixed. On the one hand, the model's predictions are very nicely confirmed for the overwhelming majority of pair-wise distribution shifts. Thus, a slight increase in the share of neighborhood which speaks Spanish at the expense of any smaller language group raises the incidence of carpooling more among Spanish speakers than it does for that other group. Indeed, except for the Spanish-English distribution shift, the results suggest that Spanish-speakers seem particularly sensitive to the relational concern, with their relative carpooling rates moving in statistically significant ways which confirm the predictions. The only pair-wise comparison involving a smaller language with a perverse estimated sign is for the French-Chinese distribution shift. For all pair-wise involving English speakers the estimated coefficients is either of the wrong sign, or else is statistically insignificant.

On the whole, the results from (14) confirm the paper's essential argument about relational cost as summarized in proposition *P*. It is quite disturbing, however, that the results do not hold up for distribution shifts involving Whites and English speakers – the two groups which are the majority racial and language groups in the country. In fact, we speculate that the large majority status of these groups in the United States, combined with the fact that there is a large amount of residential segregation in the U.S. might explain why the estimated difference-in-difference estimates may be of the wrong sign. The next section describes the basic problem, and a straightforward approach for dealing with it. Modified results are also presented in this section.

6. Segregation and Non-Linearities

Basic Problem

The results presented in the previous section which test prediction *P* are from regression which estimate linear approximations of the effect of different pair-wise distribution shifts. The technique is clearly appropriate if the true effect of different neighborhood shifts on carpooling probabilities is linear. However, if the effect of distribution shifts on carpooling is non-linear, the linear approximation estimated by the regressions we run only test the proposition *P* under very specific conditions.

To illustrate the potential problem, the graphs in Figure 1 focus on only two racial groups – Blacks and Whites. On the *x*-axis of both figures, the share of a person's neighborhood who are White is shown moving from left to right, and the share Black is measured from right to left. Any particular point on the *x*-axis is the overall racial distribution $\left(\pi_i^B, \pi_i^W\right)$ in a person's

neighborhood and a distribution shift is represented by a movement along the *x*-axis in any direction.

There are three functions in the upper graph. One shows how social capital for Whites varies as π_i^W increases. The probability that a White person has at least one non-zero social capital connection in his neighborhood, ϕ^W is shown to be upward sloping and concave in π_i^W. The probability that a Black person has at least one non-zero social capital connection in his neighborhood, ϕ^B, is upward sloping and concave in π_i^B (or falling at an increasing rate ϕ^W). Carpooling is an increasing function the particular race's social capital function and of the function ν. Importantly, the function ν can take *any* shape. In the upper graph, the ν function is drawn as an inverted U, but this is completely arbitrary. The true ν function likely has a very different non-linear shape, and might even be discontinuous over certain ranges.

The second panel of the figures depicts the White and Black carpooling functions, given the assumptions about the two ϕ's and ν in the upper panel. Notice that, *at any given overall racial distribution measured on the x-axis*, the slope of the White carpool function exceeds that of the Black carpool function. This difference in slopes at every point is what is implies by the relational cost model, and is what we would hope that regression performed on model would show. However, because the exact nonlinear form of the two carpool functions is unknown, combined with the fact that the overall racial distribution of neighborhoods in which Whites and Blacks reside may be very different, regression performed on a representative sample of Blacks and Whites need not provide the comparison we wish, even if the true relationship is consistent with the model, as is true in the case shown here.

Suppose, for example, that the vast majority of Whites live in neighborhood with overall racial distributions in the range *AA*, and that most Blacks live in neighborhoods with overall racial distributions in the range *BB*. As has been shown by White (1980), when regression analysis estimates a linear approximation to some unknown non-linear relationship, the attendant specification error is largest in those ranges where the explanatory variables are most thinly distributed. Put differently, this result implies that, if the Whites tend to be disproportionately in the range *AA* regression analysis estimates the slope of the White carpooling relationship most accurately in the range *AA*. This is the slope of the line L_1 if an linear specification is assumed. By similar reasoning, the slope of the Black carpooling function, estimated on representative sample of the Black population, estimates the slope most accurately in the range *BB*- that is, the slope of the line L_2.

But the test of proposition *P* is a comparison of the slopes of the two carpooling functions, *at the same point* of the overall racial distribution. The consequence of White's (1980) argument is that we can only truly know the slope of the White function in the range *AA*, and only know the slope of the Black function in the range *BB*. Extrapolation of these known slopes - the two slopes L_1 and L_2 - to ranges where there is relatively sparse representation of the race under study likely will dramatically misstate the slope of the race's carpooling function at that point in the overall racial distribution. Estimation of the regressions of the form presented above on representative samples of Blacks and Whites must implicitly make such extrapolations. In the example illustrated in the figure, these extrapolations yield results quite different from the truth, which is that in both the range *AA* and in the range *BB*, the true White carpooling function has a greater slope than does the Black carpooling function in the same range.

The problems associated with estimating linear approximations to an unknown non-linear relationship has been discussed by Yitzhaki (1996) and has been formally addressed by Barsky et al (2001) in their study of racial differences in wealth. A simple solution, which is a variant of the non-parametric method suggested by Barsky et al, would be to restrict the analysis to relatively small ranges of the overall distribution of the *x*-variable(s).[26] In the context of the illustrated example, an appropriate test of the model would be to conduct the analysis *only* for blacks and whites who live in neighborhoods in the range *AA*, or an analysis *only* for Blacks and Whites who live in neighborhoods like those in *BB*. Intuitively, restricting Blacks and Whites to the same range of the *x*-variable makes it less likely that the linear approximations of the carpooling relationships for Blacks and Whites are estimated at very different points on the *x*-axis, thereby making comparison of the slopes inappropriate. Of course, unless it is possible to estimate the slope for both Blacks and Whites in the restricted range of the *x*-variable, then a comparison of the slopes cannot be done. In other words, the restricted range of the *x*-variable should not only be ideally narrow, but should also contain observation for the different groups for which the comparison is being done.

[26] Barsky et al attempt to isolate how much of the difference between blacks and whites can be explained by income differences. The fact that the underlying conditional wealth functions is non-linear of unknown functional form, combined with the fact that the incomes for blacks and whites incomes are distributed very differently means that linear approximations to the conditional wealth function are inappropriate for the reasons described above. They employ a non-parametric estimation method, in which the income of whites is re-weighted so that the distribution of "synthetic" white function approximates the true income distribution for blacks. Their weighting scheme drops whites whose incomes exceed that of the black order statistic in income. Their method is thus a more complicated version of the simple restriction we recommend here.

Actual Segregation and Restricted Samples

Figures 2 present the actual neighborhood racial distribution of the neighborhoods in which the working men in the pooled IPUMS sample live. The figure shows the effect of racial segregation in the U.S. The vast majority of Whites live in PUMAs which are more than 80 percent White. By contrast, non-Whites tend to live in neighborhoods which are substantially less White. Similarly, dramatic differences in the patterns are evident for all of the races. The median percent PUMA Black for Blacks in the sample is larger than the median percent PUMA Black for non-Blacks. Figure 3 shows the actual neighborhood language distribution of our sample. The same pattern of residential segregation is clearly evident for language as well.

These two figures dramatically illustrate why the estimated effects for distribution shifts involving Whites and English speakers may be of the wrong sign in the full sample. Given the patterns of residential segregation in the data, the slope of the carpooling functions for Whites and non-Whites (and for English and non-English speakers), are estimated at very different points in the overall distribution of racial and language distributions of neighborhoods. The comparisons of the slopes implicit in the regression framework are therefore inappropriate.

The distribution of Whites and racial minorities makes the choice of a restriction clear. Figure 2 indicates that a natural restriction is to restrict the percent PUMA variable to greater than some cut point around eighty percent, preserving most of the White distribution and limiting the domain of the percent PUMA White variable to around 0.2. Obviously, this also implicitly restricts all individuals to neighborhoods whose population is less than 20 percent non-White. Luckily, such neighborhoods (neighborhoods where the majority group is a majority and the minority groups compose less than 20 percent of the population) are where most people live anyway. If this restriction is imposed, it should be the case that the gap between the median neighborhood distributions for different races falls substantially. Moreover, there should be enough observations in this restricted sample to make pair-wise comparisons possible. Since English Speakers are more of a majority than Whites, a natural restriction for language is that the neighborhood be more than approximately 85% English, slightly above the restriction for Whites.

Table 6a shows the effect of imposing these restriction on the pooled IPUMS sample. Before the restriction, the difference in the median percent PUMA White for Whites and Non-Whites is 0.2 – fully one-fifth the range of the variable. Similar differences are evident for all of the racial groups, and for “percent English” and “percent Spanish” in the language categories as well. When we restrict the sample to observations for which the percent of the neighborhood White is at least 0.8, this single restriction is enough to drop the difference in the medians of all

of the "percent of neighborhood" variables between the own and other group by more than 200% in every case, and by 400% for the percent PUMA White variable. Restricting the sample to neighborhoods greater than 85% English-speaking has an equally dramatic effect on the difference in neighborhood language medians. Spanish speakers in particular are helped by the restriction which reduces the gap between medians by a factor of 10.

Table 6b shows that, overall, imposing the restriction that the PUMA be greater than 80% White causes us to drop 35% of the individual observations from the original sample. We retain 73% of Whites, 24% of the Blacks, 37% of the Asians, and 26% of the Hispanics. With the restriction that the neighborhood be greater than eighty-five percent English speaking, 77% of the English speakers in the original sample are retained, as are 21% of the Spanish speakers, and more than 50% each of the original German, Italian and French speakers. Overall, imposing the language restriction causes 37% of the individual level observations to be dropped.

The effect of imposing the restrictions can also be expressed in terms of the percent of the original 1726 PUMAs lost. Imposing the race restriction retains 63% of the original PUMAs, while imposing the language restriction allows 71% of the original PUMAs to be retained. Figures 4 and 5 show where the dropped PUMAs are from. Figure 4 shows that most of the PUMAs dropped because of the race restriction are in the Mid-Atlantic to South corridor. This is probably because of the large number of segregated black neighborhoods in the south. Figure 5 indicates that PUMAs dropped because of the language restriction tend to be in the Southwest. This is similarly attributable to the prominence of segregated Spanish speaking neighborhoods in that region.

Restricted Sample Results

Tables 7 and 8 presents difference in difference estimates of proposition *P*, using the results from regressions of the form of (14), but performed on the restricted samples described above.[27] As before, the results control for state effects, with standard errors clustered by PUMA and corrected for heteroskedasticity.

The results in Table 7 strongly support the predictions of the relational cost model. None of the estimated effects is statistically significant with the wrong sign, and many of them go the right way significantly. This suggests that the wrong-signed results presented earlier for Whites must have been caused by the problem caused by the segregation and non-linearity problems

[27] We attempted restrictions other than those shown here. Specifically, we varied the minimum values for the percent PUMA White and percent PUMA English-speaking from 0.7 to 0.9. The estimated results are relatively stable across these different restrictions.

described above. When this problem is corrected by a suitable restriction, we find, as the model predicts, that a marginal increase in the percent PUMA White and an attendant reduction in some other race is predicted to cause carpooling for Whites to rise by more than it does for people of the group who see the incidence of people of their own type in their neighborhood fall. This result is particularly strong for interactions between Whites and Blacks, and especially for the definition of carpooling for which we are most confident that any pooling is occurring with people outside of the respondent's household.

Interestingly, the estimated effects indicate that for interactions among people from different racial minority groups, the negative effects of relational costs are strongest for the interaction with Asian. For example, the estimates predict that a slight increase in the fraction Black of a neighborhood at the expense of lowering the incidence of Asians would cause the incidence of carpooling among Blacks to rise dramatically relative to that for Asians. By contrast, the results predict that neighborhood shifts in percent Black and percent Hispanic produce no change in the relative carpooling probability of Blacks and Hispanics. Just as the interactions between Blacks and Hispanic do not seem to be dramatically affected by relational cost considerations, Asian-White distribution shifts are estimated to produce no statistically different effect on the carpooling probabilities of Whites and Asians.

The race results seem to indicate that while relational costs exist between people from different races, the magnitude of the negative effects differ depending on the particular races being studied. Thus, racial heterogeneity alone seems to be an inappropriate measure of relational costs within a neighborhood, since only certain racial relationships have any salience.

Table 8 presents tests of the conditions implied by the model for distribution shifts involving language on the sample restricted to PUMAs of more than 85% English speakers. The results in this table confirm the relational cost hypothesis even more dramatically than did the results for race. This is to be expected; language proficiency, unlike racial identity, is a mechanical and almost necessary impediment to social capital formation. Most of the estimated effects of neighborhood distribution shifts in language composition point the right way significantly. The results improve under the more restrictive definitions of carpooling, indicating (as we would expect) that neighborhood composition is most important for interactions occurring outside of the household.

As with race, we find that there is substantial variance in the magnitude of the language estimates depending upon which pair-wise relationships are being analyzed. Certain groups such as English Speakers and Italian Speakers seem to have little or no relational costs with each other, perhaps because most Italian speakers also speak English. By contrast, French speakers and

Chinese speakers, groups that are unlikely to understand each other's language, seem to have substantial barriers preventing them from interacting. Again, this indicates that any attempt to measure or predict social capital by summarizing the neighborhood distribution in terms of a single index of heterogeneity masks differences in the salience of various group relationships.

Overall, the results present us with striking confirmation of the existence of relational costs between members of groups possessing different relational traits. Furthermore, the difference between the restricted and unrestricted results suggest that the magnitude of relational costs varies with respect to the neighborhood distribution. In particular, the fact that in the unrestricted model the minority groups had more positive coefficients and the majority groups had less positive coefficients than in the restricted model suggests that social capital may be concave in a group's own share of the population.[28] The difference-in-difference estimates from the restricted sample reflect the effect of variation within a small range of potential population distributions. If we could estimate these effects over different ranges of equal small width, we would most probably obtain different answers.[29] Luckily, this range is the range that represents most of the real variation in the United States and is therefore the linear approximation of most greatest interest. It is important to keep in mind, however, that this range is most likely inappropriate for analysis in other countries with different distributions of relational traits.

The idea that social capital is concave in the neighborhood distribution is also important because it indicates that heterogeneity indices may be inappropriate as measures of or determinants of social capital. As mentioned before, certain relational traits are not salient with respect to certain groups. However, if social capital is concave in the one's own group share, some pair-wise shifts in the neighborhood distribution leading to more heterogeneity according to a single heterogeneity index (namely those shifts decreasing the majority group's share and increasing the shares of minority groups) can actually *increase* aggregate social capital.

7. Conclusion

Most of the previous literature on social capital analyzes the phenomenon at the aggregate level, such as the state, region, or country. This paper assesses how individual level social capital is determined, both because it is out of these individual stocks that aggregate social capital is formed, and because analysis of social capital's determinants is rendered virtually impossible unless the distinct decisions and actions of individuals are isolated and analyzed. It

[28] Revisit figure 1 for a reminder of why concavity would generate these results.
[29] Recall, we *cannot* conduct the analysis over alternative small ranges of the overall neighborhood distributions because residential segregation guarantees that cross-race comparisons in certain ranges of the distribution are impossible.

belongs to the small literature devoted to an "economic approach" to social capital (Glaeser (2000)).

This paper develops a simple framework which argues that an individual's stock of social capital should be negatively affected by the difference between his own traits and the traits of those with whom he comes into contact. These are individual level characteristics which affect the ease, frequency or nature of social interaction. We focus on the relational traits of race and language, and on the social relations between people in a neighborhood. Many previous authors have hinted that race and language may be important determinants of social interaction, but previous explicit tests of these ideas differ from the approach presented here for two main reasons.

First, we use an indicator of social capital never previously studied. Specifically, we study individual carpooling propensity as a measure of the social capital people have with others in their neighborhood. For a variety of reasons, we believe that carpooling is superior to previously used indicators of individual social capital. Second, our results do not merely focus on the effect of an aggregate measure of community diversity. Rather, we explicitly study the interaction between own and community characteristics for several distinct categories of the relational traits.

Using a merged dataset drawn from the 1990 1% Census IPUMS file, and the aggregate 1990 STF3 tables, we estimate a difference in difference model to test the main implications of the simple framework we present. Overall, the results for both race and language are strongly consistent with the relational cost hypothesis, especially after we account for the problem posed by racial segregation and the fact that carpooling likely varies in an unknown, non-linear way with the racial and language composition of neighborhoods.

The indicator of social capital introduced here holds great promise as a future empirical measure. Carpooling is likely to be useful in exploring a variety of outcomes which authors have speculated may be related to social capital, but for which the evidence has been, at best, shaky. Examples of issues on which future work might focus is the relationship between community level social capital, as measured by the incidence of carpooling, and outcomes such as crime and education which should be decisively related to neighborhood level social capital (Jacobs, 1961). Finally, carpooling itself is likely to be of increasing interest to policy-makers, dealing with the transportation problems of the United States. Our results indicate that social capital may serve as an important and overlooked determinant of this mode of transportation choice.

Bibliography

Abreu, D. ‘‘On the Theory of Infinitely Repeated Games with Discounting,’’ *Econometrica,* LVI (1988), 383–396.

Alesina, A., R. Baqir, and W. Easterly. 1999. ‘‘Public Goods and Ethnic Divisions.’’ *Quarterly Journal of Economics* 114:1243–1284.

Alesina, A. and E. LaFerrara. 2000. “Participation in Heterogeneous Communities.” *Quarterly Journal of Economics* 115:847-904.

Alesina, A., R. Baqir, and C. Hoxby, ‘‘Political Jurisdictions in Heterogeneous Communities,’’ unpublished, 1999.

Arrow, Kenneth. 1972. “Gifts and Exchanges.” *Philosophy and Public Affairs* 1:343-363

Barsky, Robert, Bound, John, Charles, Kerwin and Lupton, Joseph. “Accounting for the Black-White Wealth Gap: A Non-Parametric Approach”, NBER Working Paper 8466.

Berg, J., J. Dickhaut, and K. McCabe, ‘‘Trust, Reciprocity, and Social History,’’ *Games and Economic Behavior,* X (1995), 122–142.

Besley, Timothy. 1995. “Nonmarket Institutions for Credit and Risk Sharing in Low-Income Countries.” *The Journal of Economic Perspectives* 9:115-127.

Borjas, George J. 1992. “Ethnic Capital and Intergenerational Mobility.” *Quarterly Journal of Economics* 107:123-50.

_____. 1995. “Ethnicity, Neighborhoods, and Human Capital Externalities.” *American Economic Review* 85:365-390.

Coleman, J. 1988. “Social Capital in the Creation of Human Capital.” *American Journal of Sociology* 94:S95-S121.

_____. 1990. *The Foundations of Social Theory*. Cambridge: Harvard University Press.

Brock, W. and S. Durlauf. 1999. “Interaction Based Models.” working paper, University of Wisconsin at Madison and forthcoming, Handbook of Econometrics 5, J. Heckman and E. Leamer eds., Amsterdam: North Holland.

Collier, P. 1998. “Social Capital and Poverty.” Mimeo. Social Capital Initiative, The World Bank.

Costa, D. and M. Kahn. 2001. “Understanding The Decline in Social Capital, 1952-1998” NBER Working Paper #8295.

DiPasquale, D. and E. Glaeser. 1999. “Incentives and Social Capital: Are Homeowners Better Citizens?” *Journal of Urban Economics* 45:354-384.

DiIulio, John J. 1996. “Help Wanted: Economists, Crime, and Public Policy.” *Journal of Economic Perspectives* 10:3-24.

Durlauf, Steven N. 1999. “The Case Against Social Capital.” Unpublished.

Ferguson, Erik. 1997. “The Rise and Fall of the American Carpool: 1970-1990.” *Transportation* 24:349-376.

Fukuyama, F. 1995. *Trust*. New York: Free Press

Furstenberg, F. and M. Hughes. 1995. “Social Capital and Successful Development Among At-Risk Youth” *Journal of Marriage and the Family* 57:580-592.

Geolytics. 1998. *Census CD+ Maps 2.1*.

Glaeser, E., D. Laibson, and B. Sacerdote. 2000. “The Economic Approach to Social Capital.” NBER Working Paper #7728.

Glaeser, E., D. Laibson, J. Scheinkman, and C. Soutter. 2000. “Measuring Trust.” *Quarterly Journal of Economics* 115:811-846.

Glaeser, E. and B. Sacerdote, 2000. “The Social Consequences of Housing.” NBER Working Paper #8034

Goldin, C., and L. Katz, ‘‘Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910–1940,’’ *Journal of Interdisciplinary History*, XXIX (1999), 683–723.

Gonzalez, Arturo. 1998. “Mexican Enclaves and the Price of Culture.” *Journal of Urban Economics* 43:273-291.

Guiso, L., P. Sapienza, and L. Zingales. 2000. “The Role of Social Capital in Financial Development.” *NBER* Working Paper #7563.

Hall, Robert E. and C. Jones. 1999. “Why Do Some Countries Produce So Much More Output Per Worker Than Others?” *Quarterly Journal of Economics* 114:83-116

Helliwell, J. and R. Putnam. 1999. “Education and Social Capital.” NBER Working Paper #7121.

Huang, H., H. Yang, and M. Bell. 2000. “The Models and Economics of Carpools.” *Annals of Regional Science* 34:55-68.

Jacobs, J., *The Death and Life of Great American Cities* (New York: Vintage, 1961).

Knack, S. and P. Keefer. 1997. “Does Social Capital Have an Economy Payoff? A Cross-Country Investigation,” *Quarterly Journal of Economics* 112:1251-1288.

La Porta, R., F. Lopez-de-Salanes, A Schleifer, and R. Vishny. 1997. “Trust in Large Organizations,” *American Economic Review Papers and Proceedings* 87:333-338.

Laumann, E. and R. Sandefur. 1998. “A Paradigm for Social Capital.” *Rationality and Society*. 10:481-495.

Lazear, Edward P. 1999. “Culture and Language.” *Journal of Political Economy* 107:S95-126. Part 2.

Loury, G., ‘‘A Dynamic Theory of Racial Income Differences,’’ in *Women, Minorities and Employment Discrimination*, P. Wallace and A. LeMund, eds. (Lexington, MA: Lexington Books, 1977).

Massey, D. 1996. “The Age of Extremes: Concentrated Poverty and Affluence in the Twenty First Century.” *Demography* 33:395-412.

Moulton, Brent R. 1990. “An Illustration of a Pitfall in Estimating The Effects of Aggregate Variables in Micro Units,” *Review of Economics and Statistics*. Vol 72. 334-338.

Park, B. and M. Rothbart. 1982. “Perception of Out-Group Homogeneity and Levels of Social Categorization: Memory for the Subordinate Attributes of In-Group and Out-Group Members.” *Journal of Personality and Social Psychology* 42:1051-1068.

Pettigrew, T. 1998. “Intergroup Contact Theory.” *Annual Review of Psychology* 49:65-85.

Portes, A. 1998. “Social Capital: Its Origins and Application in Modern Sociology.” *Annual Review of Sociology* 1-14.

Portres, A. and P. Landolt. 1996. “The Downside of Social Capital.” *The American Prospect* 26:18-22.

Putnam, R. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Putnam, R. 1995. “Tuning in, tuning out: The strange disappearance of social capital in America” PS, Political Science & Politics; Washington; Dec 1995.

Putnam, R. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Sakoda, J. 1981. “A Generalized Index of Dissimilarity.” *Demography* 18:245-250.

Tajfel, H. 1981. *Human Groups and Social Categories* Cambridge: Cambridge University Press.

Temple, Jonathan and Paul A. Johnson. 1998. “Social Capability and Economic Growth” *Quarterly Journal of Economics* 113:965-990.

White, H 1980. “Using Least Squares to Approximate Unknown Regression Functions,” *International Economic Review* 21(1), 149-169.

Yitzhaki, Shiomo 1996. “On Using Linear Regressions in Welfare Economics,” *Journal of Business and Economic Statistics*, 14(4), 478-486.

Data Appendix

We included a number of controls that might be correlated with social capital, carpooling, and the neighborhood racial distribution. Since our goal is not to fully explain the variance in carpooling, we selected controls that we thought might obfuscate the relationship between the racial distribution and carpooling. The controls can be divided into 3 categories: individual, geographic, and aggregate.

Geographic Controls:

Population Density (STF3): Measures the number of people per acre in an individual's PUMA.

Urban (IPUMS): Dummy indicating whether the individual lives in a census designated urbanized area. Since a PUMA can contain many neighborhoods that are part of urbanized areas and many that are not this gives us an approximation of the density of the individual's general town area. In many cases this town area is actually larger than a PUMA. If an individual lives in a metropolitan area, that whole area may be one Urbanized Area. Thus, one should think of the urban dummy as primarily serving as a contrast to rural status.

Small Lot (IPUMS): Dummy indicating whether an individual lives on a parcel of land less than an acre. This variable gives us an approximation of the density of the individual's immediate neighborhood.

City (IPUMS): Dummy indicating whether an individual lives in an incorporated city. Incorporated cities have population densities substantially higher than their surrounding urbanized areas. This is another approximation of "town" density.

Individual Controls:

All individual data comes from the 1990 IPUMS.

Number of Children in Household: Series of dummies for the number of the person's own children living in the household with him.

Household Size: Series of dummies for household size (in persons).

Married: Dummy for whether an individual is married.

Work in Same Puma: This variable is a dummy variable indicating whether an individual's PUMA matches the individual's PUMA of work.

Travel Time: Gives the total amount of time in minutes that it usually took the respondent to get from home to work last week, including any stops the worker usually made on the way to work.

Age: Series of age dummies.

Income: We measure Income as an individual's pre-tax wage and salary income. Income is specified in our regressions as a quadratic.

Education: Education is specified as a series of mutually exclusive dummies: high school or less, bachelor's or less, grad school or more.

Homeowner: We include a dummy for homeownership in order to account for unobserved differences in wealth and community involvement.

Not Citizen: The 1990 Census asks citizenship status of all foreign-born respondents. We include a dummy for those foreign born respondents who indicate that they are not U.S. citizens.

Vehicles: Vehicles measures the number of vehicles in the individual's household. We break this variable into a series of dummies in the regressions.

Occupation & Industry: We created a series of individual dummies for occupation and industry based upon the 1990 Census Occupation & Industry Schemes.[30]

Aggregate Controls:

All aggregate data was constructed by matching block groups to PUMAs and then using block group level 1990 STF3 tables to estimate PUMA averages.

Education: We include variables for the percent of the PUMAs population over the age of 18 that has received a high school degree or less, bachelor's degree or less, and graduate degree or more.

Mean Travel Time: Mean (PUMA) Travel time to work represents the total number of minutes it usually took the person to get to work during the reference week. The elapsed time includes time spent waiting for public transportation, picking up passengers in carpools, and time spent in other activities related to getting to work.

Industry & Occupation: We calculated the percent of each PUMAs working population that belonged to each industry and occupation type. These groups were made so as to match the individual groups.

Race and Language Group Dissimilarity: In order to calculate segregation levels differently for each group, we separated the entire population into members of that group and non-members. We then used a Stata plug-in ado file called "seg" to calculate the dissimilarity by PUMA between block groups in the composition of group members and non-group members. If there was no variation in the composition of group members and non-group members by block group than a PUMA was assigned a score of zero. If no group and non-group members lived in the same block group the PUMA was assigned a score of one, indicating complete segregation. See Sakoda (1981) for more on dissimilarity.

[30] For more on census occupation & industry codes see http://www.ipums.umn.edu/usa/volii/99occup.html and http://www.ipums.umn.edu/usa/volii/99indus.html.

Table 1: Summary Statistics

Individual Characteristics	Mean	Std. Dev.	Neighborhood (PUMA) Characteristics	Mean	Std. Dev.
Carpools (Riders >1)	0.134	0.341	Percent High School Grad or Less	0.553	0.132
Carpools (Riders >2)	0.031	0.174	Percent More Than Bachelors	0.061	0.040
Carpools (Riders >3)	0.013	0.113	Mean Travel Time (Minutes)	21.67	5.04
White	0.836	0.371	Population Density (Persons/Acre)	1.333	3.323
Black	0.090	0.286	*Racial Composition Variables*		
Asian-Pacific	0.028	0.166	Percent White	0.802	0.203
Hispanic	0.039	0.193	Percent Black	0.122	0.177
Other Race	0.007	0.085	Percent Asian	0.029	0.062
English Language	0.865	0.341	Percent Hispanic	0.039	0.073
Spanish Language	0.079	0.269	Percent Other	0.009	0.027
French Language	0.007	0.084	White Dissimilarity	0.470	0.135
Italian Language	0.004	0.064	Black Dissimilarity	0.600	0.141
German Language	0.006	0.079	Asian Dissimilarity	0.563	0.167
Chinese Language	0.006	0.075	Hispanic Dissimilarity	0.644	0.188
Other Language	0.033	0.179	Other Dissimilarity	0.873	0.151
Age	37.1	11.5	*Language Composition Variables*		
Married	0.642	0.479	Percent English	0.859	0.156
Size of Household	3.298	1.585	Percent German	0.007	0.005
In School	0.103	0.304	Percent Italian	0.006	0.012
High School Grad or Less	0.463	0.499	Percent French	0.009	0.019
More than Bachelors	0.086	0.281	Percent Spanish	0.077	0.134
Earnings	28490	23759	Percent Chinese	0.006	0.019
Not Citizen	0.060	0.238	Percent Other Language	0.036	0.046
Homeowner	0.665	0.472	English Dissimilarity	0.296	0.075
Urban	0.765	0.424	German Dissimilarity	0.546	0.129
Small Lot	0.595	0.491	Italian Dissimilarity	0.718	0.189
City	0.180	0.384	French Dissimilarity	0.563	0.124
Work In Same Puma	0.629	0.483	Spanish Dissimilarity	0.414	0.084
Travel Time (Minutes)	24.53	18.32	Chinese Dissimilarity	0.799	0.195
Number Of Vehicles In Household	2.14	1.08	Other Language Dissimilarity	0.447	0.140
Number of Individual Observations	496280		**Number of PUMAs**	1726	

Sample includes working men age 18-64

Aggregate data compiled from STF3 block group tables matched with PUMAs

Table 2: Mean Carpooling Among Different Racial and Language Groups, Under Alternative Definitions of Carpooling

	Carpools (Riders >1)	Carpools (Riders >2)	Carpools (Riders >3)
Full Sample	.134	.031	.013
Race			
White	.123	.027	.010
Black	.176	.046	.021
Asian	.153	.039	.017
Hispanic	.242	.086	.041
Other Race	.189	.051	.020
Language			
English Language	.126	.027	.011
Spanish Language	.219	.075	.036
French Language	.141	.036	.018
Italian Language	.105	.018	.008
German Language	.143	.038	.022
Chinese Language	.147	.046	.025
Other Language	.144	.036	.014

Table 3: Linear Probability Estimate of Carpooling Determinants Among Working Men Age 18-64 From Merged IPUMS-STF3 Data

	(1) Carpools: Riders >1	(2) Carpools: Riders >2	(3) Carpools: Riders >3
Age 18-22	0.0471	0.0049	-0.0009
	(0.0029)	(0.0015)	(0.0010)
Age 23-30	0.0291	0.0034	-0.0009
	(0.0020)	(0.0010)	(0.0007)
Age 31-45	0.0135	-0.0013	-0.0017
	(0.0019)	(0.0009)	(0.0006)
Age 46-55	0.0163	0.0020	0.0004
	(0.0019)	(0.0009)	(0.0006)
Married	0.0148	0.0016	0.0011
	(0.0016)	(0.0008)	(0.0006)
In School	-0.0254	-0.0073	-0.0032
	(0.0017)	(0.0009)	(0.0006)
Bachelor's Degree	-0.0257	-0.0071	-0.0026
	(0.0012)	(0.0006)	(0.0004)
Grad School +	-0.0180	-0.0038	-0.0012
	(0.0020)	(0.0010)	(0.0007)
Log Earnings	0.0365	-0.0003	-0.0032
	(0.0059)	(0.0033)	(0.0024)
Log Earnings Squared	-0.0030	-0.0002	0.0001
	(0.0003)	(0.0002)	(0.0001)
Homeowner	-0.0183	-0.0056	-0.0019
	(0.0014)	(0.0008)	(0.0005)
Urban	0.0014	0.0046	0.0031
	(0.0017)	(0.0009)	(0.0006)
Small Lot	0.0007	0.0002	-0.0001
	(0.0013)	(0.0007)	(0.0005)
City	-0.0032	-0.0007	-0.0000
	(0.0024)	(0.0012)	(0.0009)
Work In Same Puma	-0.0164	-0.0116	-0.0056
	(0.0015)	(0.0009)	(0.0006)
Log Travel Time	0.0405	0.0188	0.0099
	(0.0009)	(0.0006)	(0.0004)
1 Car	0.0302	-0.0133	-0.0105
	(0.0038)	(0.0024)	(0.0018)
2 Cars	-0.0470	-0.0307	-0.0182
	(0.0042)	(0.0025)	(0.0018)
3 Cars	-0.0591	-0.0337	-0.0199
	(0.0043)	(0.0026)	(0.0019)
4+ Cars	-0.0722	-0.0409	-0.0242
	(0.0046)	(0.0027)	(0.0021)
Log Population Density	-0.0027	-0.0010	-0.0003
	(0.0009)	(0.0005)	(0.0003)
Percent Bachelor's Degree	0.0451	0.0057	0.0026
	(0.0218)	(0.0117)	(0.0075)
Percent Grad School +	0.2706	0.0615	0.0121
	(0.0764)	(0.0427)	(0.0286)
Log Mean Travel Time	-0.0228	-0.0017	0.0029
	(0.0063)	(0.0033)	(0.0022)
Child Dummies	Yes	Yes	Yes
Household Size Dummies	Yes	Yes	Yes
Occupational Controls	Yes	Yes	Yes
Industry Controls	Yes	Yes	Yes
Observations	496280	496280	496280
PUMAs	1726	1726	1726
R-squared	0.0609	0.0386	0.0234

Data drawn from merged IPUMS Census Sample. Controls not shown included in Appendix Table 2.

All regressions include controls for state fixed effects.

Standard errors adjusted for clustering by PUMA

Table 4: Difference-in-Difference Estimates of Effects of Racial Distribution Shifts, from Linear Probability Model on Full Sample

Neighborhood Distribution Shift	Difference in Difference Estimate	(1) Carpools: Riders >1	(2) Carpools: Riders >2	(3) Carpools: Riders >3
White → Black	$(g_{BB} - g_{BW}) - (g_{WB} - g_{WW})$	-.028	-.006	-.001
		(.011)	(.006)	(.004)
White → Hispanic	$(g_{HH} - g_{HW}) - (g_{WH} - g_{WW})$	-.080	-.043	-.008
		(.034)	(.027)	(.021)
White → Asian	$(g_{AA} - g_{AW}) - (g_{WA} - g_{WW})$	-.007	.006	.016
		(.029)	(.017)	(.010)
Black → Hispanic	$(g_{HH} - g_{HB}) - (g_{BH} - g_{BB})$	.073	.055	.045
		(.051)	(.036)	(.028)
Black → Asian	$(g_{AA} - g_{AB}) - (g_{BA} - g_{BB})$	.191	.095	.066
		(.055)	(.029)	(.020)
Asian → Hispanic	$(g_{HH} - g_{HA}) - (g_{AH} - g_{AA})$	.291	.225	.160
		(.075)	(.051)	(.040)
Individual Obs		496280	496280	496280
PUMAs		1726	1726	1726
R^2		0.0622	0.0398	0.0243

Data drawn from merged IPUMS Census Sample.

All regressions contain controls for the variables listed in Table 3 and Appendix Table 2.

All regressions include controls for state fixed effects and PUMA wide dissimilarity.

Standard errors adjusted for clustering by PUMA

Table 5: Difference-in-Difference Estimates of Effects of Language Distribution Shifts, from Linear Probability Model on Full Sample

Neighborhood Distribution Shift	Difference in Difference Estimate	(1) Carpools: Riders >1	(2) Carpools: Riders >2	(3) Carpools: Riders >3
English → Spanish	$(g_{SS} - g_{SE}) - (g_{ES} - g_{EE})$	-.050	-.023	-.013
		(.016)	(.011)	(.007)
English → French	$(g_{FF} - g_{FE}) - (g_{EF} - g_{EE})$	.111	.032	-.009
		(.079)	(.071)	(.053)
English → German	$(g_{GG} - g_{GE}) - (g_{EG} - g_{EE})$	3.081	3.372	3.235
		(1.340)	(1.682)	(1.486)
English → Italian	$(g_{II} - g_{IE}) - (g_{EI} - g_{EE})$	.295	-.054	-.093
		(.172)	(.069)	(.046)
English → Chinese	$(g_{CC} - g_{CE}) - (g_{EC} - g_{EE})$	-.328	-.036	.053
		(.075)	(.047)	(.036)
Spanish → French	$(g_{FF} - g_{FS}) - (g_{SF} - g_{SS})$	.900	.498	.251
		(.193)	(.111)	(.087)
Spanish → German	$(g_{GG} - g_{GS}) - (g_{SG} - g_{SS})$	3.645	3.829	3.877
		(1.539)	(1.760)	(1.500)
Spanish → Italian	$(g_{II} - g_{IS}) - (g_{SI} - g_{SS})$	.913	.167	-.024
		(.262)	(.134)	(.097)
Spanish → Chinese	$(g_{CC} - g_{CS}) - (g_{SC} - g_{SS})$	.171	.202	.211
		(.140)	(.078)	(.053)
French → German	$(g_{GG} - g_{GF}) - (g_{FG} - g_{FF})$	2.174	4.003	3.592
		(2.145)	(1.742)	(1.471)
French → Italian	$(g_{II} - g_{IF}) - (g_{FI} - g_{FF})$	1.352	2.936	4.046
		(1.880)	(1.996)	(1.721)
French → Chinese	$(g_{CC} - g_{CF}) - (g_{FC} - g_{FF})$	-.804	-1.989	-2.119
		(.964)	(.831)	(.812)
German → Italian	$(g_{II} - g_{IG}) - (g_{GI} - g_{GG})$	1.352	2.936	4.046
		(1.880)	(1.996)	(1.721)
German → Chinese	$(g_{CC} - g_{CG}) - (g_{GC} - g_{GG})$	5.422	5.075	5.339
		(1.964)	(2.399)	(1.831)
Italian → Chinese	$(g_{CC} - g_{CI}) - (g_{IC} - g_{II})$	.232	-.132	-.039
		(.436)	(.244)	(.116)
Individual Obs		496280	496280	496280
PUMAs		1726	1726	1726
R^2		0.0621	0.0405	0.0253

Data drawn from merged IPUMS Census Sample.

All regressions contain controls for the variables listed in Table 3 and Appendix Table 2.

All regressions include controls for state fixed effects and PUMA wide dissimilarity.

Standard errors adjusted for clustering by PUMA

Table 6a: Median Neighborhood Distributions of Race and Language Groups Among Individuals of Same and Other Types in Unrestricted and Restricted IPUMS Sample

	Unrestricted			Restricted			
Percent of Neighborhood:	Among Persons of Same Type	Among Persons of Different Type	Difference	Among Persons of Same Time	Among Persons of Different Type	Difference	Ratio of Differences
Race							
White	0.90	0.68	0.22	0.93	0.88	0.05	4.19
Black	0.27	0.04	0.23	0.08	0.02	0.05	4.33
Asian	0.08	0.01	0.07	0.03	0.01	0.02	2.95
Hispanic	0.13	0.01	0.13	0.04	0.00	0.03	3.65
Language							
English	0.93	0.75	0.18	0.95	0.92	0.03	6.81
Spanish	0.21	0.02	0.20	0.03	0.01	0.02	10.67
French	0.01	0.00	0.01	0.01	0.00	0.00	3.17
Italian	0.02	0.00	0.02	0.01	0.00	0.01	1.44
German	0.01	0.01	0.00	0.01	0.01	0.00	1.01

Table 6b: Reduction in Sample Due to Restrictions

	Observations			PUMAs		
Observations of Type:	Unrestricted	Restricted	Ratio	Unrestricted	Restricted	Ratio
Race	496,280	324,145	0.65			
White	414,713	301,114	0.73	1,726	1,096	0.63
Black	44,701	10,706	0.24	1,726	1,096	0.63
Asian	14,138	5,215	0.37	1,726	1,096	0.63
Hispanic	19,158	5,070	0.26	1,726	1,096	0.63
Language	496,280	351,819	0.71			
English	429,482	331,660	0.77	1,726	1,219	0.71
Spanish	38,979	8,156	0.21	1,726	1,219	0.71
French	3,511	1,784	0.51	1,726	1,219	0.71
Italian	2,037	1,057	0.52	1,726	1,219	0.71
German	3,102	2,216	0.71	1,726	1,219	0.71

Table 7: Difference-in-Difference Estimates of Effects of Racial Distribution Shifts, from Linear Probability Model on Sample of Neighborhoods > 80% White

Neighborhood Distribution Shift	Difference in Difference Estimate	(1) Carpools: Riders >1	(2) Carpools: Riders >2	(3) Carpools: Riders >3
White → Black	$(g_{BB} - g_{BW}) - (g_{WB} - g_{WW})$	.164	.098	.064
		(.086)	(.047)	(.030)
White → Hispanic	$(g_{HH} - g_{HW}) - (g_{WH} - g_{WW})$	.148	.119	.039
		(.170)	(.144)	(.117)
White → Asian	$(g_{AA} - g_{AW}) - (g_{WA} - g_{WW})$	-.120	-.009	.044
		(.142)	(.083)	(.056)
Black → Hispanic	$(g_{HH} - g_{HB}) - (g_{BH} - g_{BB})$	.168	-.074	.025
		(.332)	(.232)	(.170)
Black → Asian	$(g_{AA} - g_{AB}) - (g_{BA} - g_{BB})$	.446	.275	.247
		(.263)	(.155)	(.098)
Asian → Hispanic	$(g_{HH} - g_{HA}) - (g_{AH} - g_{AA})$	.777	.721	.498
		(.427)	(.265)	(.192)
Individual Obs		324145	324145	324145
PUMAs		1096	1096	1096
R^2		0.0579	0.0332	0.0193

Data drawn from merged IPUMS Census Sample.

All regressions contain controls for the variables listed in Table 3 and Appendix Table 2.

All regressions include controls for state fixed effects and PUMA wide dissimilarity.

Standard errors adjusted for clustering by PUMA

Table 8: Difference-in-Difference Estimates of Effects of Language Distribution Shifts, from Linear Probability Model on Sample of Neighborhoods > 85% English

Neighborhood Distribution Shift	Difference in Difference Estimate	(1) Carpools: Riders >1	(2) Carpools: Riders >2	(3) Carpools: Riders >3
English → Spanish	$(g_{SS} - g_{SE}) - (g_{ES} - g_{EE})$	.597	.366	.253
		(.231)	(.196)	(.158)
English → French	$(g_{FF} - g_{FE}) - (g_{EF} - g_{EE})$	.640	.188	.167
		(.281)	(.163)	(.157)
English → German	$(g_{GG} - g_{GE}) - (g_{EG} - g_{EE})$	2.992	4.133	3.804
		(1.507)	(1.832)	(1.648)
English → Italian	$(g_{II} - g_{IE}) - (g_{EI} - g_{EE})$	.949	-.442	-.145
		(.791)	(.320)	(.233)
English → Chinese	$(g_{CC} - g_{CE}) - (g_{EC} - g_{EE})$	2.978	5.189	7.006
		(3.354)	(3.526)	(3.224)
Spanish → French	$(g_{FF} - g_{FS}) - (g_{SF} - g_{SS})$	.882	.513	.824
		(.930)	(.711)	(.447)
Spanish → German	$(g_{GG} - g_{GS}) - (g_{SG} - g_{SS})$	4.283	4.162	5.082
		(2.005)	(2.079)	(1.840)
Spanish → Italian	$(g_{II} - g_{IS}) - (g_{SI} - g_{SS})$	1.758	.199	-.299
		(1.204)	(.771)	(.637)
Spanish → Chinese	$(g_{CC} - g_{CS}) - (g_{SC} - g_{SS})$	5.762	7.873	8.397
		(2.775)	(3.202)	(2.858)
French → German	$(g_{GG} - g_{GF}) - (g_{FG} - g_{FF})$	3.480	3.319	3.626
		(2.602)	(1.924)	(1.596)
French → Italian	$(g_{II} - g_{IF}) - (g_{FI} - g_{FF})$	5.517	2.433	4.206
		(3.004)	(2.460)	(1.757)
French → Chinese	$(g_{CC} - g_{CF}) - (g_{FC} - g_{FF})$	-4.904	3.746	8.611
		(6.131)	(4.536)	(3.259)
German → Italian	$(g_{II} - g_{IG}) - (g_{GI} - g_{GG})$	5.517	2.433	4.206
		(3.004)	(2.460)	(1.757)
German → Chinese	$(g_{CC} - g_{CG}) - (g_{GC} - g_{GG})$	8.016	16.278	15.700
		(5.364)	(4.881)	(4.300)
Italian → Chinese	$(g_{CC} - g_{CI}) - (g_{IC} - g_{II})$	1.261	3.489	6.194
		(4.512)	(3.433)	(2.695)
Individual Obs		351819	351819	351819
PUMAs		1219	1219	1219
R^2		0.0623	0.0366	0.0221

Data drawn from merged IPUMS Census Sample.

All regressions contain controls for the variables listed in Table 3 and Appendix Table 2.

All regressions include controls for state fixed effects and PUMA wide dissimilarity.

Standard errors adjusted for clustering by PUMA

Appendix Table 1: Linear Probability Estimate of Carpooling Determinants Among Working Men Age 18-64 in NPTS

	(1) Usually Carpools to Work
Highschool	-0.0597
	(0.0120)
Bachelors	-0.0308
	(0.0052)
Grad School +	0.0168
	(0.0068)
Age	-0.0086
	(0.0015)
Age Squared	0.0001
	(0.0000)
Urban	-0.0123
	(0.0140)
Town	-0.0038
	(0.0090)
Suburb	-0.0129
	(0.0100)
MSA Size	0.0038
	(0.0020)
Housing Unit Density (Units/Square Mile), BG	0.0000
	(0.0000)
Pop Density, Block Group	-0.0000
	(0.0000)
# of Vehicles in Household	-0.0222
	(0.0027)
# of People in Household	0.0098
	(0.0019)
White	-0.0220
	(0.0132)
Black	-0.0004
	(0.0168)
Asian	-0.0207
	(0.0200)
Other	0.0000
	(0.0000)
Hispanic	0.0147
	(0.0140)
Income 0-30,000	-0.0103
	(0.0085)
Income 30,000-50,000	-0.0199
	(0.0074)
Income 50,000-80,000	-0.0062
	(0.0074)
Income 80,000+	-0.0166
	(0.0081)
Minutes From Home To Work	0.0010
	(0.0002)
Miles To Work	0.0001
	(0.0002)
Bus Service Available	0.0079
	(0.0059)
Streetcar Service Available	-0.0280
	(0.0146)
Subway Service Available	-0.0119
	(0.0132)
Commuter Train Service Available	-0.0005
	(0.0124)
Other Public Transit Available	0.0024
	(0.0135)
Constant	0.2887
	(0.0442)
State Fixed Effects	Yes
Observations	17454
R^2	0.0291

Appendix Table 2: Extra Controls for Base Regressions

	(1)	(2)	(3)
	Carpools: Riders >1	Carpools: Riders >2	Carpools: Riders >3
1 Child	-0.0101	-0.0040	-0.0042
	(0.0022)	(0.0011)	(0.0008)
2 Children	-0.0310	-0.0077	-0.0047
	(0.0027)	(0.0015)	(0.0012)
3 Children	-0.0455	-0.0147	-0.0060
	(0.0035)	(0.0021)	(0.0016)
4 Children	-0.0701	-0.0284	-0.0151
	(0.0050)	(0.0033)	(0.0025)
1 Person Household	-0.1818	-0.0610	-0.0293
	(0.0041)	(0.0027)	(0.0020)
2 Person Household	-0.0890	-0.0506	-0.0250
	(0.0036)	(0.0025)	(0.0020)
3 Person Household	-0.0776	-0.0371	-0.0203
	(0.0036)	(0.0025)	(0.0018)
4 Person Household	-0.0605	-0.0294	-0.0149
	(0.0035)	(0.0023)	(0.0017)
5 Person Household	-0.0393	-0.0193	-0.0113
	(0.0037)	(0.0025)	(0.0018)
Not Citizen	0.0387	0.0281	0.0149
	(0.0032)	(0.0021)	(0.0015)
Managerial or Professional Occupation	-0.0736	-0.0464	-0.0295
	(0.0062)	(0.0045)	(0.0036)
Technical, Sales, or Administrative Support Occupation	-0.0800	-0.0501	-0.0314
	(0.0062)	(0.0045)	(0.0036)
Service Occupation	-0.0845	-0.0545	-0.0338
	(0.0063)	(0.0046)	(0.0037)
Farming, Forestry, or Fishing Occupatoin	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Precision, Production, Craft, or Repair Occupation	-0.0517	-0.0428	-0.0289
	(0.0063)	(0.0046)	(0.0037)
Operator, Fabricator, or Repair Occupation	-0.0578	-0.0442	-0.0295
	(0.0063)	(0.0046)	(0.0037)
Military Occupation	-0.0877	-0.0449	-0.0300
	(0.0098)	(0.0060)	(0.0044)
Agriculture, Forestry, or Fishing Industry	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Mining Industry	0.0421	0.0271	0.0167
	(0.0095)	(0.0072)	(0.0053)
Construction Industry	0.0499	0.0005	-0.0060
	(0.0067)	(0.0048)	(0.0037)
Nondurable Manufacturing Industry	-0.0117	-0.0241	-0.0150
	(0.0066)	(0.0047)	(0.0036)
Durable Manufacturing Industry	0.0090	-0.0143	-0.0081
	(0.0065)	(0.0047)	(0.0036)
Transportation, Communications, or Other Public Utility Industry	-0.0354	-0.0285	-0.0148
	(0.0064)	(0.0046)	(0.0035)
Wholesale Trade Industry	-0.0316	-0.0290	-0.0166
	(0.0065)	(0.0047)	(0.0036)
Retail Trade Industry	-0.0476	-0.0352	-0.0177
	(0.0064)	(0.0046)	(0.0036)
Finance, Insurance, or Real Estate Industry	-0.0215	-0.0273	-0.0150
	(0.0065)	(0.0046)	(0.0036)
Business or Repair Services Industry	-0.0298	-0.0311	-0.0165
	(0.0066)	(0.0047)	(0.0036)
Personal Services Industry	-0.0321	-0.0332	-0.0191
	(0.0073)	(0.0049)	(0.0037)

(continued below)

Appendix Table 2: Extra Controls for Base Regressions (continued)

	(1) Riders >1	(2) Riders >2	(3) Riders >3
Entertainment or Recreation Services Industry	-0.0253	-0.0309	-0.0165
	(0.0074)	(0.0050)	(0.0038)
Professional or Related Services Industry	0.0004	-0.0202	-0.0115
	(0.0065)	(0.0046)	(0.0036)
Public Administration Industry	0.0074	-0.0125	-0.0066
	(0.0067)	(0.0047)	(0.0036)
Military Industry	-0.0227	-0.0264	-0.0129
	(0.0080)	(0.0056)	(0.0041)
Percent Agriculture, Forestry, and Fishing Industry	0.2126	0.2329	0.1749
	(0.0537)	(0.0326)	(0.0237)
Percent Mining Industry	1.0268	0.7824	0.3917
	(0.2886)	(0.1839)	(0.1573)
Percent Construction Industry	1.0712	0.7420	0.3596
	(0.2913)	(0.1904)	(0.1626)
Percent Nondurables Manufacturing Industry	0.8279	0.6765	0.3510
	(0.2853)	(0.1807)	(0.1540)
Percent Durables Manufacturing Industry	0.8471	0.6328	0.3167
	(0.2852)	(0.1795)	(0.1536)
Percent Transportation Industry	0.7230	0.6026	0.2812
	(0.2935)	(0.1860)	(0.1586)
Percent Communications Industry	1.3499	0.7666	0.3477
	(0.3027)	(0.1874)	(0.1590)
Percent Wholesale Trade Industry	0.8520	0.5820	0.2663
	(0.3041)	(0.1967)	(0.1677)
Percent Retail Trade Industry	0.9231	0.7041	0.3439
	(0.2868)	(0.1813)	(0.1550)
Percent Finance, Insurance, and Real Estate Industry	0.9045	0.6936	0.3236
	(0.2933)	(0.1889)	(0.1598)
Percent Business & Repair Services Industry	0.8545	0.7080	0.3379
	(0.3028)	(0.1860)	(0.1541)
Percent Personal Services Industry	0.9385	0.5658	0.2802
	(0.2988)	(0.1841)	(0.1572)
Percent Entertainment & Recreation Services Industry	1.0287	0.6763	0.3174
	(0.2970)	(0.1844)	(0.1558)
Percent Health Services Industry	0.9982	0.7013	0.3288
	(0.2869)	(0.1784)	(0.1513)
Percent Educational Services Industry	0.9274	0.6729	0.3204
	(0.2894)	(0.1814)	(0.1548)
Percent Other Professional & Related Specialties Industry	0.6931	0.5778	0.2961
	(0.3020)	(0.1852)	(0.1583)
Percent Public Administration Industry	1.1921	0.7892	0.3889
	(0.2866)	(0.1826)	(0.1559)
Percent Executive, Administrative, and Managerial Occupation	-0.9652	-0.6502	-0.3261
	(0.2984)	(0.1909)	(0.1587)
Percent Professional Specialty Occupation	-1.0408	-0.6765	-0.3175
	(0.2934)	(0.1800)	(0.1540)
Percent Technicians & Related Support Occupation	-0.7169	-0.6065	-0.2750
	(0.3094)	(0.1869)	(0.1576)
Percent Sales Occupation	-0.8809	-0.6873	-0.2966
	(0.2929)	(0.1871)	(0.1595)
Percent Administrative Support Occupation	-1.0054	-0.7489	-0.3448
	(0.2900)	(0.1854)	(0.1580)
Percent Private Services Occupation	-1.2490	-0.8297	-0.3605
	(0.4052)	(0.2350)	(0.1714)
Percent Protective Services Occupation	-0.4779	-0.5220	-0.2786
	(0.3079)	(0.1878)	(0.1538)

(continued below)

Appendix Table 2: Extra Controls for Base Regressions (continued)

	(1)	(2)	(3)
	Riders >1	Riders >2	Riders >3
Percent Other Services Occupation	-1.0448	-0.6759	-0.3254
	(0.2941)	(0.1806)	(0.1522)
Percent Farming, Forestry, and Fishing Occupation	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Percent Precision Production, Craft, & Repair Occupation	-0.8071	-0.6736	-0.3370
	(0.2837)	(0.1799)	(0.1524)
Percent Machine Operators, Assemblers, & Inspectors Occupation	-0.6588	-0.6119	-0.3203
	(0.2887)	(0.1800)	(0.1525)
Percent Transportation & Material Moving Occupation	-1.0788	-0.8041	-0.3874
	(0.3235)	(0.2023)	(0.1704)
Percent Handlers, Equipment Cleaners, Helpers & Laborers Occupation	-0.4471	-0.3264	-0.1255
	(0.3222)	(0.2045)	(0.1696)
Observations	496280	496280	496280
PUMAs	1726	1726	1726
R-squared	0.0609	0.0386	0.0234

All regressions include controls for state fixed effects.

Standard errors adjusted for clustering by PUMA

Appendix Table 3: Coefficients and Standard Erros From Linear Probability Model Used to Construct Pairwise Racial Results

	(1) Carpools: Riders >1	(2) Carpools: Riders >2	(3) Carpools: Riders >3
Percent White	0.1057	0.0021	-0.0118
	(0.0823)	(0.0541)	(0.0283)
Percent Black	0.0852	0.0876	0.0265
	(0.0867)	(0.0608)	(0.0343)
Percent Asian	0.0924	0.0245	0.0197
	(0.1559)	(0.0965)	(0.0679)
Percent Hispanic	0.1676	0.0168	-0.0157
	(0.1388)	(0.0815)	(0.0516)
White X Percent White	-0.1237	-0.0191	-0.0071
	(0.1012)	(0.0639)	(0.0346)
White X Percent Black	-0.1061	-0.0999	-0.0429
	(0.1001)	(0.0677)	(0.0382)
White X Percent Asian	-0.0775	-0.0185	-0.0265
	(0.1669)	(0.1030)	(0.0701)
White X Percent Hispanic	-0.2109	-0.0186	0.0039
	(0.1505)	(0.0876)	(0.0554)
Black X Percent White	0.0970	0.0701	0.0175
	(0.1355)	(0.1062)	(0.0783)
Black X Percent Black	0.0866	-0.0166	-0.0197
	(0.1685)	(0.1235)	(0.0914)
Black X Percent Asian	-0.0989	-0.0317	-0.0572
	(0.1916)	(0.1281)	(0.0978)
Black X Percent Hispanic	-0.1565	-0.0062	-0.0045
	(0.1659)	(0.1185)	(0.0866)
Asian X Percent White	-0.0729	0.1471	0.0440
	(0.3326)	(0.0860)	(0.0536)
Asian X Percent Black	-0.0389	0.0737	0.0120
	(0.3335)	(0.0915)	(0.0546)
Asian X Percent Asian	-0.0334	0.1540	0.0408
	(0.3570)	(0.1128)	(0.0781)
Asian X Percent Hispanic	-0.2187	0.1068	0.0309
	(0.3489)	(0.1088)	(0.0732)
Hispanic X Percent White	-0.0299	-0.0183	0.0146
	(0.1470)	(0.1438)	(0.0830)
Hispanic X Percent Black	-0.0268	-0.1258	-0.0420
	(0.1548)	(0.1464)	(0.0860)
Hispanic X Percent Asian	-0.3020	-0.2383	-0.1323
	(0.2040)	(0.1637)	(0.1032)
Hispanic X Percent Hispanic	-0.1966	-0.0606	0.0179
	(0.1804)	(0.1573)	(0.0945)
Observations	496280	496280	496280
PUMAs	1726	1726	1726
R^2	0.0622	0.0398	0.0243

Data drawn from merged IPUMS Census Sample.

All regressions contain controls for the variables listed in Table 3 and Appendix Table 2.

All regressions include controls for state fixed effects and PUMA wide dissimilarity.

Standard errors adjusted for clustering by PUMA

Appendix Table 4: Coefficients Used to Construct Pairwise Language Results

	(1)	(2)	(3)
	Carpools: Riders >1	Carpools: Riders >2	Carpools: Riders >3
Percent English	-0.0022	-0.0715	-0.0639
	(0.0470)	(0.0297)	(0.0191)
Percent Spanish	-0.0703	-0.0596	-0.0673
	(0.0612)	(0.0376)	(0.0221)
Percent French	0.4997	0.5307	0.0594
	(0.3018)	(0.2854)	(0.1021)
Percent Italian	-0.5276	-0.2294	-0.1425
	(0.1552)	(0.0948)	(0.0595)
Percent German	-0.2805	0.1914	0.2299
	(0.5865)	(0.3958)	(0.2895)
Percent Chinese	-0.2153	-0.1446	-0.0981
	(0.1183)	(0.0901)	(0.0579)
English X Percent English	-0.0098	0.0454	0.0490
	(0.0501)	(0.0308)	(0.0192)
English X Percent Spanish	0.0038	0.0164	0.0496
	(0.0647)	(0.0388)	(0.0228)
English X Percent French	-0.6111	-0.5740	-0.0482
	(0.2962)	(0.2863)	(0.1020)
English X Percent Italian	0.3020	0.1990	0.1288
	(0.1590)	(0.0936)	(0.0609)
English X Percent German	0.2004	-0.1816	-0.2015
	(0.5778)	(0.3894)	(0.2971)
English X Percent Chinese	0.2086	0.1450	0.1144
	(0.1210)	(0.0918)	(0.0576)
Spanish X Percent English	0.0672	0.1507	0.1034
	(0.0704)	(0.0406)	(0.0277)
Spanish X Percent Spanish	0.0305	0.0983	0.0911
	(0.0810)	(0.0472)	(0.0306)
Spanish X Percent French	-1.3278	-0.9348	-0.2521
	(0.3621)	(0.3052)	(0.1212)
Spanish X Percent Italian	-0.2982	0.0175	0.0921
	(0.2624)	(0.1559)	(0.0972)
Spanish X Percent German	-0.2602	-0.5004	-0.7697
	(0.9396)	(0.6596)	(0.3842)
Spanish X Percent Chinese	-0.2198	0.0090	0.0131
	(0.1555)	(0.1067)	(0.0688)
French X Percent English	-0.1373	0.0451	0.0006
	(0.1451)	(0.0697)	(0.0541)
French X Percent Spanish	-0.1688	-0.0073	-0.0140
	(0.1752)	(0.0890)	(0.0619)
French X Percent French	-0.6274	-0.5423	-0.1057
	(0.3539)	(0.2864)	(0.1101)
French X Percent Italian	-0.2120	-0.2769	-0.1289
	(0.4991)	(0.2069)	(0.1568)
French X Percent German	1.0288	-0.5992	-0.5223
	(1.5939)	(0.7591)	(0.5230)
French X Percent Chinese	-0.2966	0.0399	-0.0235
	(0.3805)	(0.1379)	(0.1050)

(continued below)

Appendix Table 4: Coefficients Used to Construct Pairwise Language Results (continued)

	(1)	(2)	(3)
	Carpools: Riders >1	Carpools: Riders >2	Carpools: Riders >3
Italian X Percent English	0.2358	0.1455	0.1109
	(0.1362)	(0.0484)	(0.0319)
Italian X Percent Spanish	0.2587	0.1584	0.1206
	(0.1706)	(0.0785)	(0.0549)
Italian X Percent French	-1.1078	-0.6390	-0.0779
	(0.6340)	(0.3610)	(0.2142)
Italian X Percent Italian	0.8425	0.2447	0.0978
	(0.2690)	(0.1056)	(0.0749)
Italian X Percent German	1.7259	0.9502	-0.3845
	(1.1530)	(0.6362)	(0.4117)
Italian X Percent Chinese	0.3068	0.1853	0.1233
	(0.2439)	(0.1088)	(0.0687)
German X Percent English	-0.0914	-0.1062	0.0272
	(0.1620)	(0.1177)	(0.0856)
German X Percent Spanish	-0.1544	-0.1917	-0.0051
	(0.1836)	(0.1293)	(0.0989)
German X Percent French	-0.6302	-0.9070	-0.1647
	(0.6265)	(0.3268)	(0.1582)
German X Percent Italian	0.9649	-0.6027	-0.5520
	(0.5632)	(0.5599)	(0.4170)
German X Percent German	3.2000	3.0389	3.0112
	(1.3660)	(1.5676)	(1.3435)
German X Percent Chinese	-0.7958	-0.6715	-0.2531
	(0.5280)	(0.3160)	(0.2133)
Chinese X Percent English	-0.1333	0.1281	0.1168
	(0.1359)	(0.0692)	(0.0483)
Chinese X Percent Spanish	-0.1641	0.0789	0.1022
	(0.1630)	(0.0824)	(0.0530)
Chinese X Percent French	0.2297	1.5980	2.2715
	(0.8707)	(0.7995)	(0.7969)
Chinese X Percent Italian	0.0600	0.3827	0.2487
	(0.3559)	(0.2561)	(0.1328)
Chinese X Percent German	-1.6698	-1.1727	-1.8399
	(1.5404)	(1.7020)	(1.0455)
Chinese X Percent Chinese	-0.2434	0.1916	0.2352
	(0.1941)	(0.1059)	(0.0704)
Observations	496280	496280	496280
PUMAs	1726	1726	1726
R^2	0.0621	0.0405	0.0253

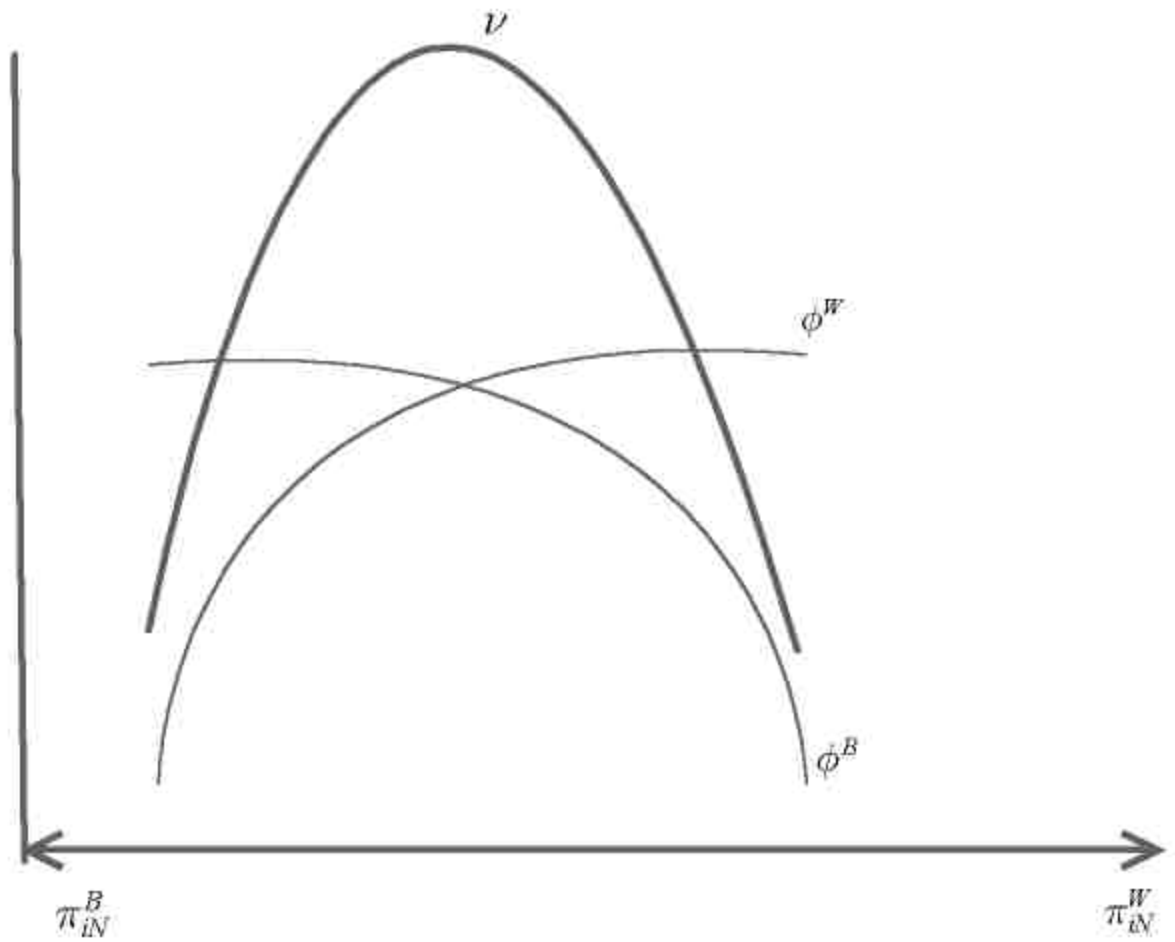
Data drawn from merged IPUMS Census Sample.

All regressions contain controls for the variables listed in Table 3 and Appendix Table 2.

All regressions include controls for state fixed effects and PUMA wide dissimilarity.

Standard errors adjusted for clustering by PUMA

Figure 1.



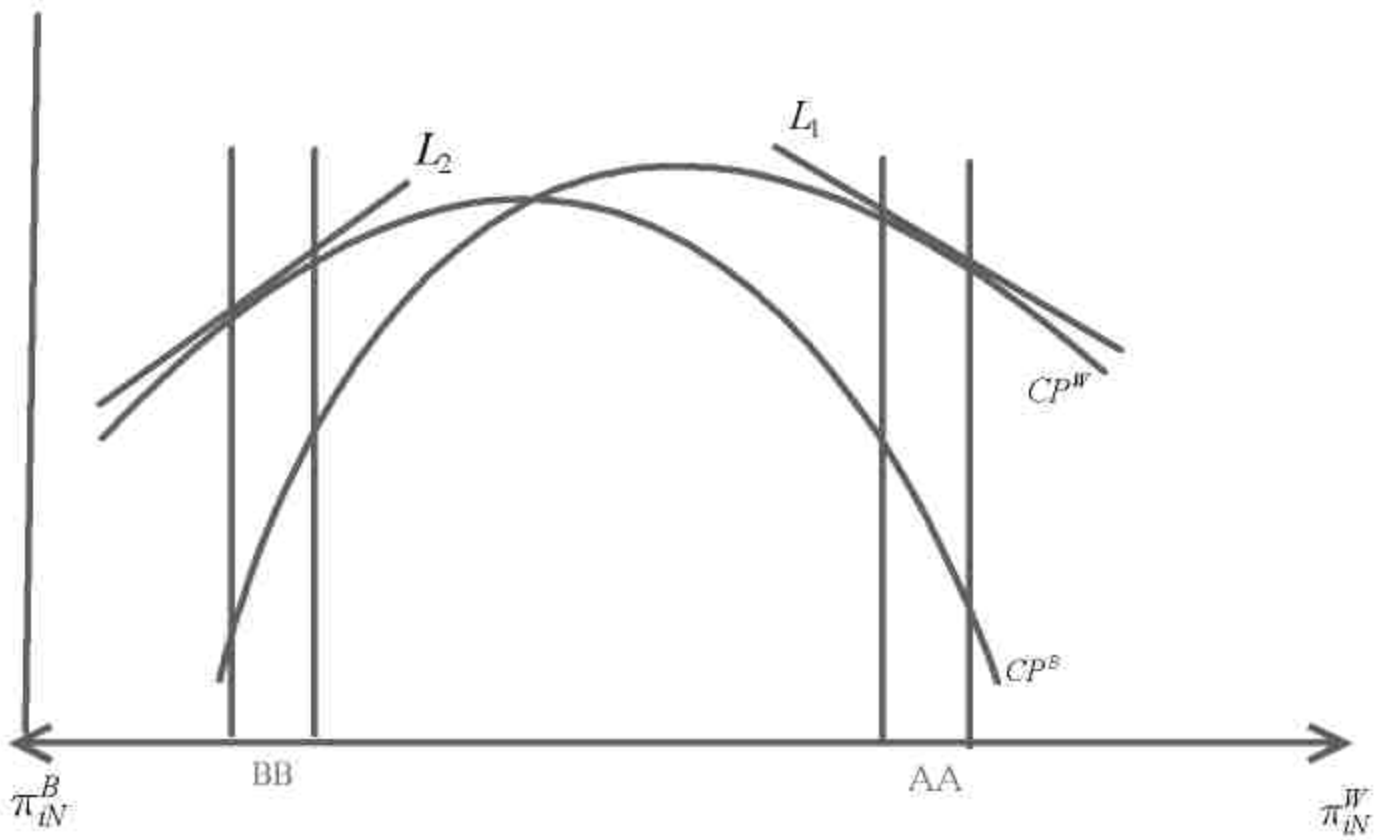
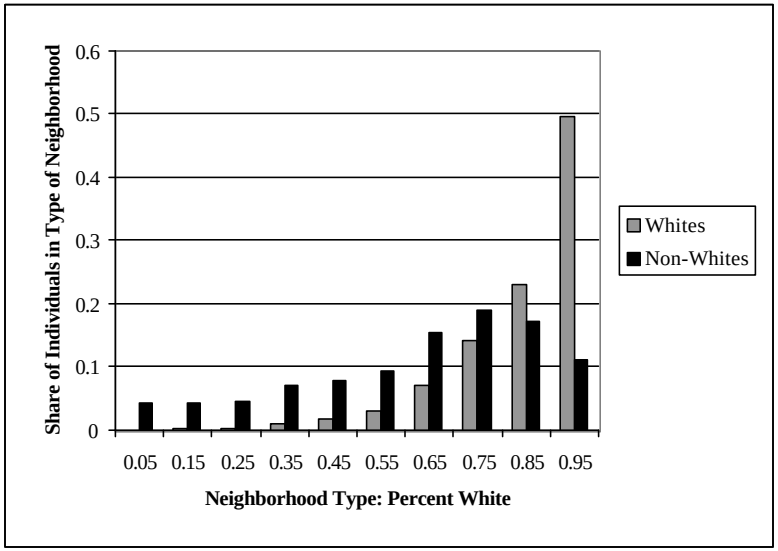
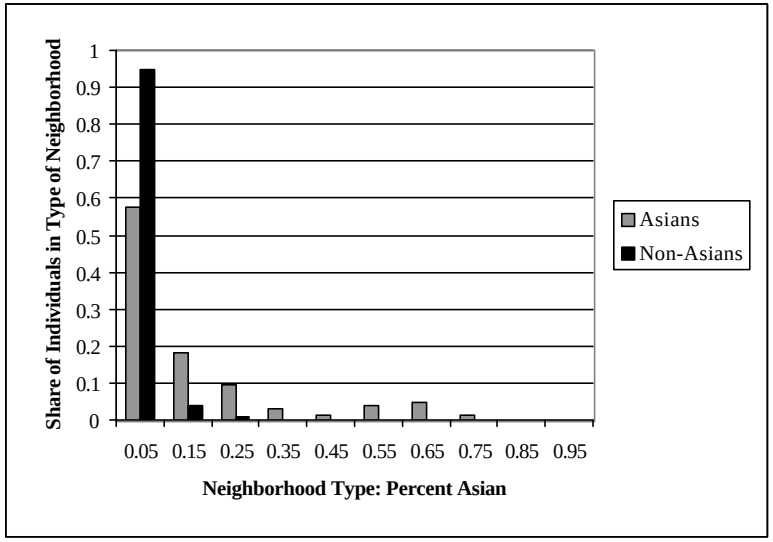
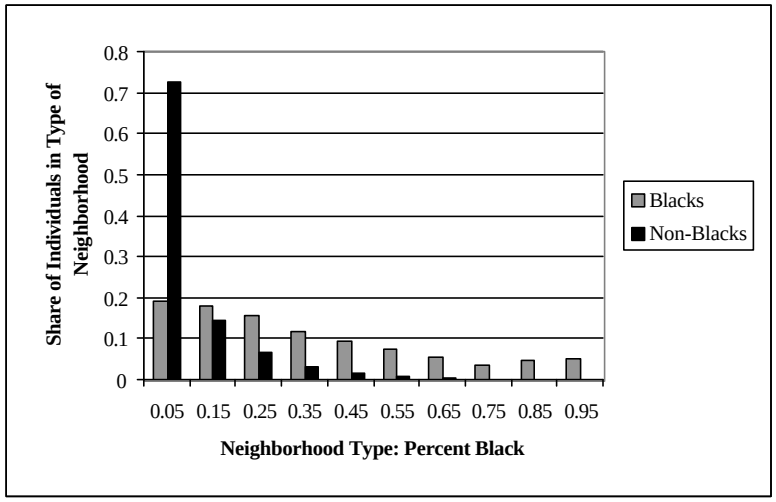


Figure 2: Distribution of Individuals Across Types of Neighborhoods







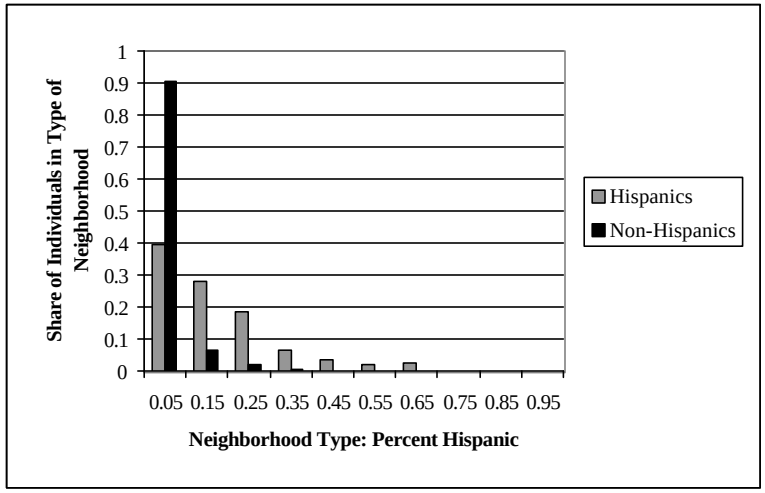
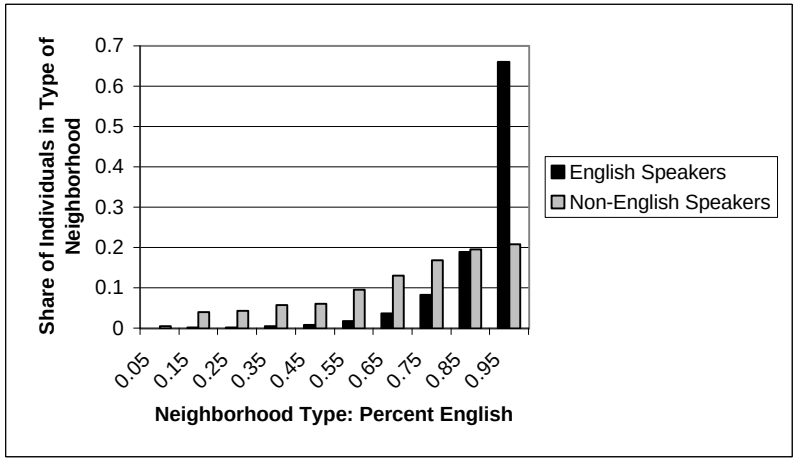
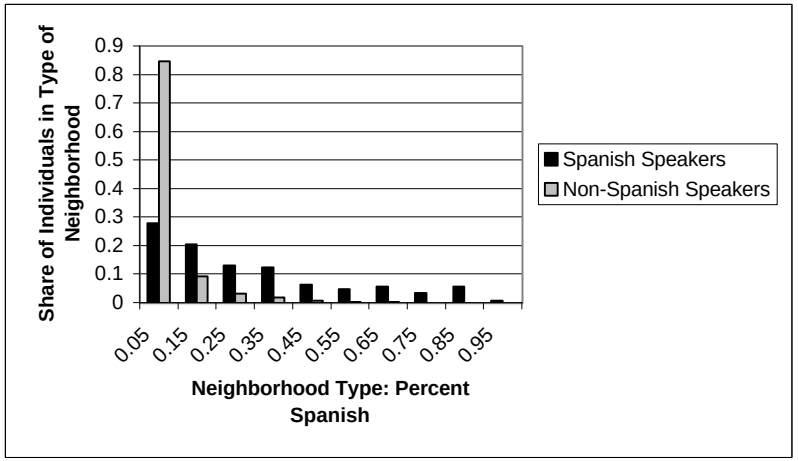
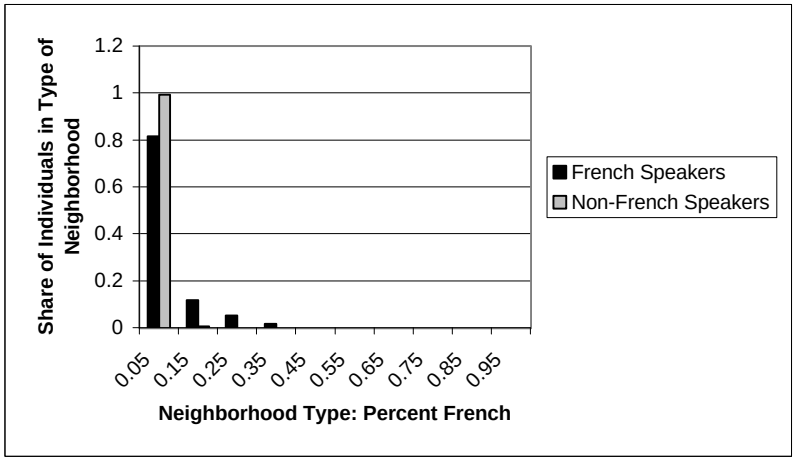
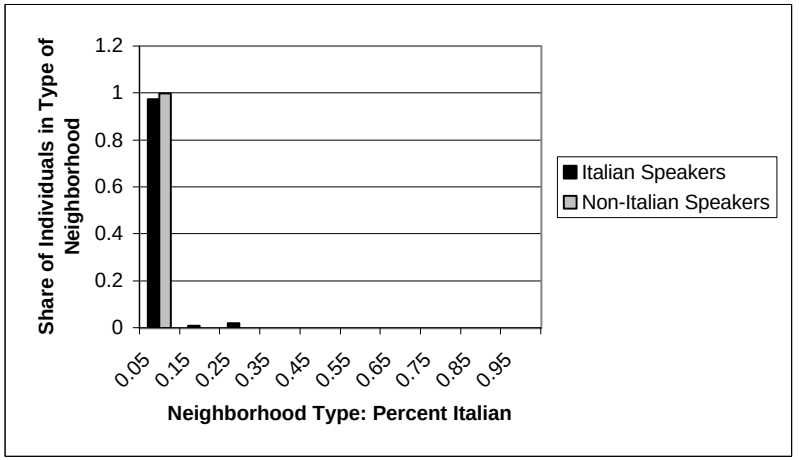


Figure 3: Distribution of Individuals Across Neighborhood Types









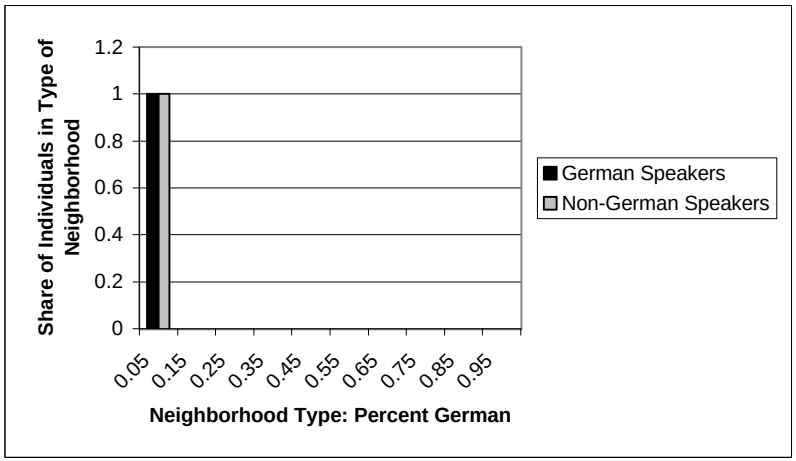
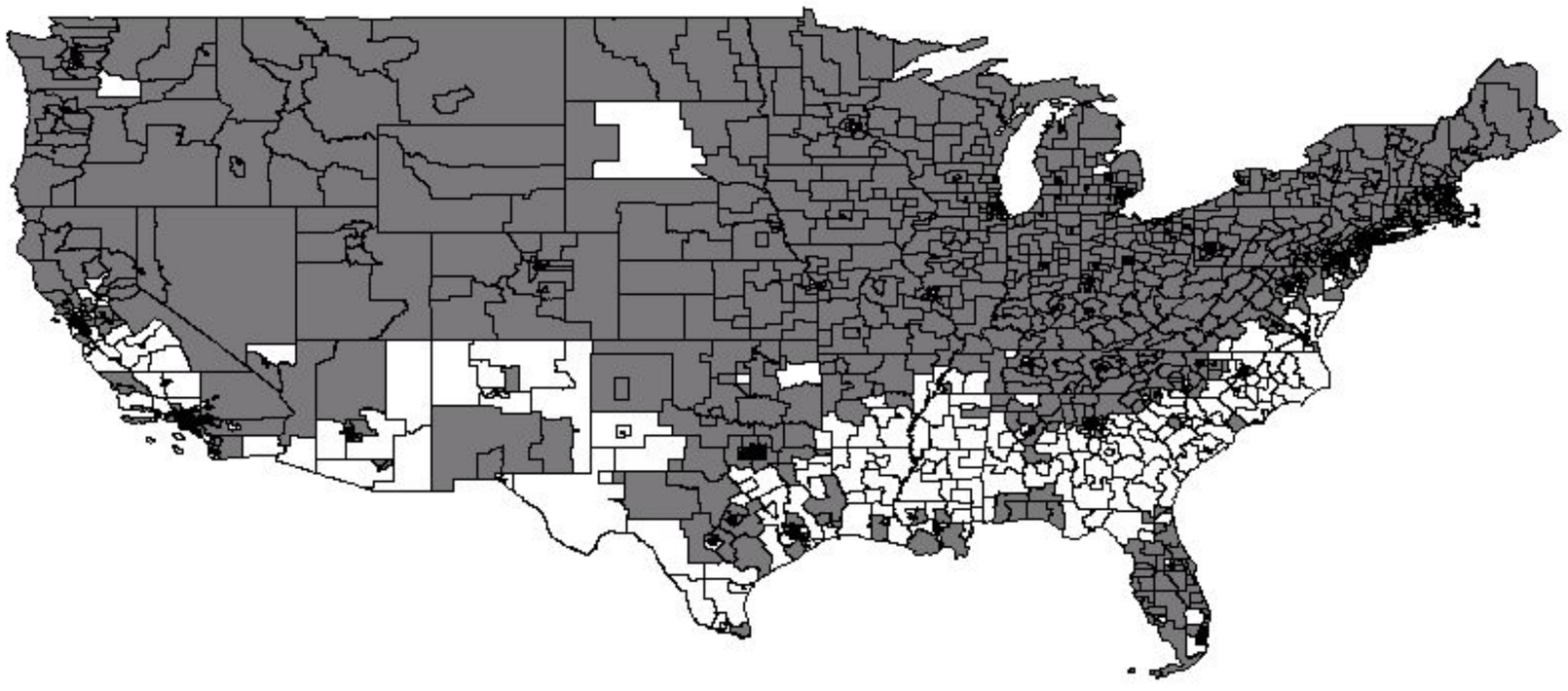


Figure 4: Map of Continental United States Showing PUMAs More Than 80% White

(note: Hawaii and Alaska not pictured here)



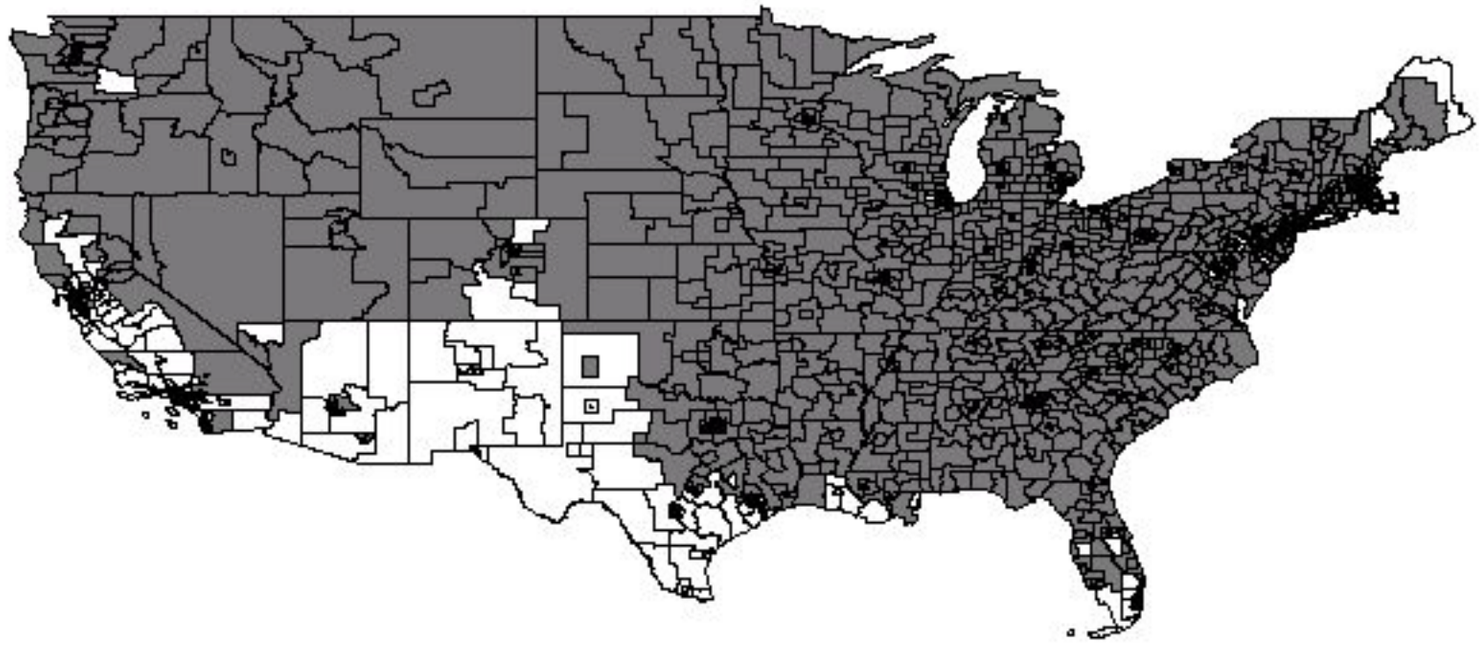
Percent in PUMA White > 80%
Percent in PUMA White <= 80%

N
W
E
S

1000 0 1000 2000 Miles

Figure 5: Map of Continental United States Showing PUMAs More Than 85% English

(note: Hawaii and Alaska not pictured here)



Percent in PUMA English > 85%
Percent in PUMA English <= 85%

N
W
E
S

1000 0 1000 2000 Miles