ECONOMICS SEMINAR

Presenter:
Morris Davis
Rutgers University

“Neighborhood Choices, Neighborhood Effects and Housing Vouchers”

Host: Marie Rekkas

Tuesday, November 7, 2017
3:30 - 5:00 p.m.
WMC 3611
Neighborhood Choices, Neighborhood Effects and Housing Vouchers*

Morris A. Davis
Rutgers University
morris.a.davis@rutgers.edu

Jesse Gregory
University of Wisconsin - Madison
jmgregory@ssc.wisc.edu

Daniel A. Hartley
Federal Reserve Bank of Chicago
Daniel.A.Hartley@chi.frb.org

Kegon T. K. Tan
University of Rochester
ttan8@ur.rochester.edu

July 20, 2017

Abstract

We study how households choose neighborhoods, how neighborhoods affect child ability, and how housing vouchers influence neighborhood choices and child outcomes. We use two new panel data sets with tract-level detail for Los Angeles county to estimate a dynamic model of optimal location choice for renting households and, separately, the impact of living in a given tract on child test scores, a proxy for child ability. We simulate optimal location choices and the resulting changes in child ability of the poorest households in our sample under various housing-voucher policies that incentivize households to relocate to tracts that beneficially impact child ability. When vouchers are restricted such that they can only be applied to units in the top 5% of tracts based on tract impact on child ability, we compute an “optimal” voucher amount of $300 per month where the benefits to child ability net of voucher costs are maximized. We also compute a “break-even” voucher amount of $700 per month in which benefits are equal to costs.

JEL Classification Numbers: I240, I31, I38, J13, R23, R38
Keywords: Neighborhood Choice, Neighborhood Effects, Housing Vouchers

*We thank Dionissi Aliprantis, Patrick Bayer, Edward Coulson, Steven Durlauf, Edward Glaeser, Nathaniel Hendren, John Kennan, Kyle Magnum, Christopher Palmer, Stephen Ross, Amy Ellen Schwartz, Holger Sieg, Chris Taber, Stijn Van Nieuwerburgh, Jim Walker, Maisy Wong and numerous seminar participants for helpful comments. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.
1 Introduction

In this paper we investigate how households optimally choose a neighborhood in which to live, how neighborhoods affect the ability of children, and how housing vouchers affect neighborhood choices and child ability. These topics have been studied individually before, but our approach is different and our data are new. We show that some neighborhoods can significantly improve child cognitive ability, but parents differ in their willingness to rent in these neighborhoods. We use our framework to simulate the impact of various housing-voucher policies on child ability. The policies we consider have the feature that voucher amounts vary by neighborhood, and larger vouchers are assigned to neighborhoods more likely to positively impact child ability. We conclude by discussing the costs and benefits of a targeted housing voucher that can only be applied in a small set of neighborhoods that we find substantially improve child ability. For Los Angeles, the area of our study, we compute a “surplus-maximizing” voucher amount for this policy of $300 per month. At this amount, the gap between the benefits of the voucher to child ability and the costs is maximized. We also compute a “break-even” voucher amount of $700 per month in which benefits are equal to costs.

Our paper has three main sections, and the first two reflect contributions to distinct literatures. In our first section, we specify and estimate a dynamic model of optimal location choice using detailed micro panel data, in the spirit of Kennan and Walker (2011) and Bayer, McMillan, Murphy, and Timmins (2015). We estimate the model using panel data from the Federal Reserve Bank of New York (FRBNY) Consumer Credit Panel / Equifax. This is a 5% random sample of U.S. adults with an active credit file and any individuals residing in the same household. To our knowledge we are the first to use these data to estimate a location choice model. We restrict our sample to renters residing in Los Angeles County. We study renters to mitigate the influence of availability of credit on location choice, and we focus on Los Angeles County to match our results with estimates of the impact of neighborhoods on child ability, discussed next.

Our estimation sample from the FRBNY Consumer Credit Panel / Equifax data consists of more than 1.75 million person-year observations. This huge sample allows us to estimate a full vector of model parameters for many discrete “types” of people. Our use of many types in estimation captures permanent heterogeneity in preferences for neighborhoods, as compared to an estimation framework with fewer types which would necessarily attribute more systematic variation in neighborhood choices across households to period-by-period unobservable shocks. As a corollary, our use of many types in estimation helps us better identify how households adjust neighborhood choices in response to policy changes. We find
that for many types of households, utility varies greatly across Census tracts; and, for many Census tracts, the utility of living in the tract varies widely across types.

In our second section, we estimate the impact of neighborhoods, in our case specific Census tracts in Los Angeles county, on the cognitive ability of children. There is a large literature in the social sciences studying these “neighborhood effects” on child ability, adolescent behavior, health, labor earnings, and other individual level outcomes. Empirical studies using observational data often find strong associations between neighborhood quality, broadly defined, and positive individual-level outcomes: See Leventhal and Brooks-Gunn (2000), Durlauf (2004) and Ross (2011) for recent surveys. While these studies typically attempt to account for selection issues, the fact that individuals endogenously sort into neighborhoods leaves open the possibility of non-causal explanations for these patterns.

We make two contributions to this literature. First, we use a new longitudinal dataset in estimation, the Los Angeles Family and Neighborhood Survey (LA FANS). The LA FANS data allow for a substantially richer set of controls than are typically available in observational studies of neighborhood effects. Second, we estimate the impact of neighborhoods on child ability using a “value-added approach,” in which changes in student ability over time, as measured by changes in math test scores, are regressed on neighborhood fixed effects and a set of individual-level controls including, most importantly, lagged child test scores. The value-added approach has been applied widely in assessing teacher quality, for example Kane and Staiger (2008) and Chetty, Friedman, and Rockoff (2014), but has not yet been used in the neighborhood effects literature.

The key advantage of the value-added approach for our application is that the method recovers estimates of the effect of specific Census tracts on child ability, as compared to the average effect of neighborhoods associated with particular observable characteristics such as average income level and racial composition, the typical approach in the neighborhood-effects literature. We estimate economically important variation in neighborhood value-added across Census tracts in Los Angeles County: Our findings imply that 13 years of exposure to a Census tract providing value-added one standard deviation above the mean tract, on average, boosts the level of a child’s ability in the cross-section by one-half of one standard deviation. In support of a causal, as opposed to selection-driven, interpretation of our neighborhood value-added estimates, we show that after we have controlled for children’s

---

1 For example, Cutler and Glaeser (1997) study the impact of segregation on outcomes of African-Americans using topographical features of cities as instruments for location choice and Aaronson (1998) measures neighborhood effects by studying outcomes of siblings at least three years apart in age after a move.

2 See Aaronson (1998) for examples of instruments used by other researchers in this field and their potential limitations.
lagged test scores and demographics, controlling additionally for variables such as parental ability, parental demographics, and household income and assets, which are strongly predictive of child ability in simple cross-sectional regressions, add very little in explanatory power for changes over time in child test scores.

In the final sections of the paper, we overlay the results of the previous two sections to study how various housing-voucher policies affect optimal location choices of households and the ability of children. We begin by demonstrating that our model can replicate the results of the MTO experiment without using any MTO data in estimation; this is akin to an out-of-sample forecasting test of the model. When we implement a voucher program in the model much like the voucher program of the MTO experiment, simulated test scores fail to rise, just like the actual MTO results. Test scores failed to increase because the set of voucher-eligible neighborhoods in low-poverty areas included relatively low-rent, low-value-added neighborhoods and many households receiving a voucher chose to move to one of these neighborhoods.

In the final section of the paper, we analyze the impact of housing voucher policies that vary in the dollar amount of the voucher and/or the extent to which voucher amounts target specific sets of high-value-added tracts. The results of this section highlight the benefits of our structural-estimation approach. Once we understand households’ preferences for neighborhoods and disutility from rents, and given a mapping of neighborhoods to child ability and earnings, we can use simulations of the model to quantitatively evaluate the impact of any housing voucher policy on child ability and outcomes. Section 1 of our paper uses available data on household migration patterns, rents and characteristics of the housing stock to inform us as to the desirability of various neighborhoods and the sensitivity of household choices to rents and section 2 tells policy makers which neighborhoods should be targeted by vouchers. We show that vouchers that more directly target or aggressively subsidize high-value-added tracts yield larger improvements to average child value-added and adult wages. We conclude the section by considering a voucher policy in which vouchers can only be used in the 5% of tracts with the highest child value-added. As mentioned earlier, the surplus-maximizing voucher amount of this policy is $300 per month and the break-even voucher amount is $700 per month.

Throughout our paper, we use the words “child test scores” and “child ability” interchangeably, but we note there may be many ways in which neighborhoods affect child ability and adult earnings that we do not currently capture. In this sense, our paper is close to that of Galiani, Murphy, and Pantano (2015) who estimate a structural model of location choice using MTO data and run counterfactual experiments with their estimated model. Our paper is different in that we study the impact of MTO and other public policies on child well-being.

An important example is the work of Chetty, Hendren, and Katz (2015) who show that neighborhoods
we can generate tract-level benefits and optimal voucher amounts using data on hand, even if imperfect. In the event policymakers and researchers wish to specify an alternative set of tract-level benefits, we can use our framework to determine surplus-maximizing and break-even housing vouchers.

2 Location Choice Model and Estimates

2.1 Model

We consider the decision problem of a household head deciding where his or her family should live. As in Kennan and Walker (2011) and Bayer, McMillan, Murphy, and Timmins (2015), we model location choices in a dynamic discrete choice setting. For purposes of exposition, we write down the model describing the optimal decision problem of a single family which enables us to keep notation relatively clean. When we estimate the parameters of this model, we will allow for the existence of many different “types” of people in the data. Each type of person will face the same decision problem, but the vector of parameters that determines payoffs and choice probabilities will be allowed to vary across types of people.

The family can choose to live in one of \( J \) locations. Denote \( j \) as the family’s current location. We write the value to the family of moving to location \( \ell \) given a current location of \( j \) and current value of a shock \( \epsilon_\ell \) (to be explained later) as

\[
V(\ell | j, \epsilon_\ell) = u(\ell | j, \epsilon_\ell) + \beta EV(\ell)
\]

In the above equation \( EV(\ell) \) is the expected future value of having chosen to live in \( \ell \) today and \( \beta \) is the factor by which future utility is discounted. We assume the household problem does not change over time, explaining the lack of time subscripts.

\( u \) is the flow utility the agent receives today from choosing to live in \( \ell \) given a current location of \( j \) and a value for \( \epsilon_\ell \). We assume \( u \) is the simple function

\[
u(\ell | j, \epsilon_\ell) = \delta_\ell - \kappa \cdot 1_{\ell \neq j} + \epsilon_\ell
\]

\( \delta_\ell \) is the flow utility the household receives this period from living in neighborhood \( \ell \), net of rents and other costs; \( \kappa = \kappa_0 + \kappa_1 \cdot D_{\ell j} \) are all costs (utility and financial) a household must pay when it moves to a different neighborhood i.e. when \( \ell \neq j \), which we specify as the sum of a fixed cost \( \kappa_0 \) and a cost that increases at rate \( \kappa_1 \) with distance in miles between the centroid affect college attainment and future adult wages of children even if those neighborhoods do not affect child test scores.
of tracts $\ell$ and $j$ denoted $D_{\ell j}$; $1_{\ell \neq j}$ is an indicator function that is equal to 1 if location $\ell \neq j$ and 0 otherwise; and $\epsilon_\ell$ is a random shock that is known at the time of the location choice. $\epsilon_\ell$ is assumed to be iid across locations, time and people. The parameters $\delta_\ell$, $\kappa_0$ and $\kappa_1$ may vary across households, but for any given household these parameters are assumed fixed over time. $\epsilon_\ell$ induces otherwise identical households living at the same location to optimally choose different future locations. Note that $\delta_\ell$ is the type-specific indirect utility of living in neighborhood $\ell$, and this utility may depend on attributes such as amenities, crime, school quality, pollution, access to public transportation and possibly child value-added, a point to which we return later.

Denote $\epsilon_1$ as the shock associated with location 1, $\epsilon_2$ as the shock with location 2, and so on. In each period after the vector of $\epsilon$ are revealed (one for each location), households choose the location that yields the maximal value

$$V (j | \epsilon_1, \epsilon_2, \ldots, \epsilon_J) = \max_{\ell \in 1, \ldots, J} V (\ell | j, \epsilon_\ell) \quad (1)$$

$EV (j)$ is the expected value of (1), where the expectation is taken with respect to the vector of $\epsilon$.

While this model looks simplistic, it is the workhorse model used to study location choice. Differences in models reflect specific areas of study and availability of data. For example, in their study of migration across states, Kennan and Walker (2011) replace $\delta$ with wages after adjusting for cost of living. Bishop and Murphy (2011) and Bayer, McMillan, Murphy, and Timmins (2015) specify $\delta$ as a linear function of spatially-varying amenities with the aim of recovering individuals’ willingness to pay for those amenities. We allow the $\delta$’s to vary flexibly across neighborhoods, with the aim of realistically forecasting the substitution patterns that are likely to occur in response to government policies that change the relative prices of neighborhoods.

When the $\epsilon$ are assumed to be drawn i.i.d. from the Type 1 Extreme Value Distribution, the expected value function $EV (j)$ has the functional form

$$EV (j) = \log \left\{ \sum_{\ell=1}^J \exp \tilde{V} (\ell | j) \right\} + \zeta \quad (2)$$

where $\zeta$ is equal to Euler’s constant and

$$\tilde{V} (\ell | j) = \delta_\ell - \kappa \cdot 1_{\ell \neq j} + \beta EV (\ell) \quad (3)$$

That is, the tilde symbol signifies that the shock $\epsilon_\ell$ has been omitted. Additionally, it can
be shown that the log of the probability that location $\ell$ is chosen given a current location of $j$, call it $p(\ell \mid j)$, has the solution

$$p(\ell \mid j) = \tilde{V}(\ell \mid j) - \log \left\{ \sum_{\ell' = 1}^{J} \exp \left[ \tilde{V}(\ell' \mid j) \right] \right\}$$  \hspace{1cm} (4)

Subtract and add $\tilde{V}(k \mid j)$ to the right-hand side of the above to derive

$$p(\ell \mid j) = \tilde{V}(\ell \mid j) - \tilde{V}(k \mid j) - \log \left\{ \sum_{\ell' = 1}^{J} \exp \left[ \tilde{V}(\ell' \mid j) - \tilde{V}(k \mid j) \right] \right\}$$ \hspace{1cm} (5)

One approach to estimating model parameters such as Rust (1987) is to solve for the value functions at a given set of parameters, apply equation (5) directly to generate a likelihood over the observed choice probabilities, and then search for the set of parameters that maximizes the likelihood. This approach is computationally intensive because it requires solving for the value functions at each step of the likelihood, which involves backwards recursions using equations (2) and (3). In cases such as ours, involving many parameters to be estimated, this approach is computationally infeasible.

Instead, we use the approach of Hotz and Miller (1993) and employed by Bishop (2012) in similar work. This approach does not require that we solve for the value functions. Note that equation (3) implies

$$\tilde{V}(\ell \mid j) - \tilde{V}(k \mid j) = \delta_\ell - \delta_k - \kappa [1_{\ell \neq j} - 1_{k \neq j}] + \beta [EV(\ell) - EV(k)]$$ \hspace{1cm} (6)

But from equation (2),

$$EV(\ell) - EV(k) = \log \left\{ \sum_{\ell' = 1}^{J} \exp \tilde{V}(\ell' \mid l) \right\} - \log \left\{ \sum_{\ell' = 1}^{J} \exp \tilde{V}(\ell' \mid k) \right\}$$

Now note that equation (4) implies

$$p(k \mid \ell) = \tilde{V}(k \mid \ell) - \log \left\{ \sum_{\ell' = 1}^{J} \exp \tilde{V}(\ell' \mid \ell) \right\}$$

$$p(k \mid k) = \tilde{V}(k \mid k) - \log \left\{ \sum_{\ell' = 1}^{K} \exp \tilde{V}(\ell' \mid k) \right\}$$
and thus

\[ \log \left\{ \sum_{\ell' = 1}^{J} \exp \left[ \tilde{V} (\ell' \mid \ell) \right] \right\} - \log \left\{ \sum_{\ell' = 1}^{K} \exp \left[ \tilde{V} (\ell' \mid k) \right] \right\} \]

is equal to

\[ \tilde{V} (k \mid \ell) - \tilde{V} (k \mid k) - [p (k \mid \ell) - p (k \mid k)] - \kappa \cdot 1_{\ell \neq k} - [p (k \mid \ell) - p (k \mid k)] \]

The last line is quickly derived from equation (3). Therefore,

\[ EV (\ell) - EV (k) = - [p (k \mid \ell) - p (k \mid k)] + \kappa \cdot 1_{\ell \neq k} \]

and equation (6) has the expression

\[ \tilde{V} (\ell \mid j) - \tilde{V} (k \mid j) = \delta_{\ell} - \delta_{k} - \kappa [1_{\ell \neq j} - 1_{k \neq j}] - \beta [p (k \mid \ell) - p (k \mid k)] + \kappa \cdot 1_{\ell \neq k} \]

Combined, equations (5) and (7) show that the log probabilities that choices are observed are simple functions of model parameters \( \delta_1, \ldots, \delta_J, \kappa_0, \kappa_1 \) and \( \beta \) and of observed choice probabilities. In other words, a likelihood over choice probabilities observed in data can be generated without solving for value functions.

### 2.2 Data and Likelihood

We estimate the model using panel data from the FRBNY Consumer Credit Panel / Equifax. The panel is comprised of a 5% random sample of U.S. adults with a social security number, conditional on having an active credit file, and any individuals residing in the same household as an individual from that initial 5% sample.\(^5\) For years 1999 to the present, the database provides a quarterly record of variables related to debt: Mortgage and consumer loan balances, payments and delinquencies and some other variables we discuss later. The data does not contain information on race, education, or number of children and it does not contain information on income or assets although it does include the Equifax Risk Score™ which provides some information on the financial wherewithal of the household as demon-

\(^5\)The data include all individuals with 5 out of the 100 possible terminal 2-digit social security number (SSN) combinations. While the leading SSN digits are based on the birth year/location, the terminal SSN digits are essentially randomly assigned. A SSN is required to be included in the data and we do not capture the experiences of illegal immigrants. Note that a SSN is also required to receive a housing voucher.
strated in Board of Governors of the Federal Reserve System (2007). Most important for our application, the panel data includes in each period the current Census block of residence. To match the annual frequency of our location choice model, we use location data from the first quarter of each calendar year. Other authors have used the FRBNY Consumer Credit Panel / Equifax data to study the relationship of interest rates, house prices and credit (see Bhutta and Keys (2015) and Brown, Stein, and Zafar (2013)) and the impact of natural disasters on household finances (Gallagher and Hartley, 2014), but we are the first to use this data to estimate an optimal location-choice model.

We restrict our sample to individuals who, from 1999 through 2013, are never observed outside of Los Angeles county and who never hold a home mortgage, yielding 1,787,558 person-year observations. We study renters to mitigate any problems of changing credit conditions and availability of mortgages during the sample window; and we study Los Angeles in particular to link our estimates of utility to measures of neighborhood effects on child outcomes we estimate for each Census tract in Los Angeles (to be discussed later). We exclude from our estimation Census tracts with fewer than 150 rental units and tracts that are sparsely populated in the northern part of the county. The panel is not balanced, as some individuals’ credit records first become active after 1999.

An advantage of the size of our data is that we can estimate a full set of model parameters for many “types” of people, where we define a type of person based on observable demographic and economic characteristics. Previous studies of neighborhood choice such as Bayer, McMillan, Murphy, and Timmins (2015) have had access to much smaller data sets and as a result have had to restrict variation in model parameters across the population.

Table 1 compares sample statistics from the FRBNY Consumer Credit Panel / Equifax data to Census data for the tracts in Los Angeles County. This table includes data for both owners and renters. Column (2) shows the implied total population of adults ages 18-64 in the FRBNY Consumer Credit Panel / Equifax data, computed as twenty times the total number of primary individuals, and (3) shows the average population counts of adults from the 2000 and 2010 Census. The table shows that coverage in the low poverty tracts is very high, above 90%. Coverage remains high but falls for the higher-poverty tracts, either because many individuals lack credit history or do not have a social security number. Columns (5) and (6) compare the percentage of households with a mortgage in the two data sets. Not surprisingly, the percentages fall quite dramatically with the poverty rate, and generally speaking the percentages reported in the two data sets are close. The final row

---

6In the FRBNY Consumer Credit Panel / Equifax data, renters and homeowners without a mortgage are observationally equivalent. According to data from the 2000 Census, 85% percent of the units without a home mortgage are renter-occupied for the 1,748 Census tracts of our study.

7On average, each Census tract in Los Angeles has about 4,000 people.
Table 1: Comparison of Equifax and Census Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>610,336</td>
<td>654,004</td>
<td>93.3%</td>
<td>61.6%</td>
<td>62.6%</td>
</tr>
<tr>
<td>5-10</td>
<td>1,395,831</td>
<td>1,478,114</td>
<td>94.4%</td>
<td>50.0%</td>
<td>50.2%</td>
</tr>
<tr>
<td>10-15</td>
<td>1,033,076</td>
<td>1,135,194</td>
<td>91.0%</td>
<td>40.5%</td>
<td>39.2%</td>
</tr>
<tr>
<td>15-20</td>
<td>751,098</td>
<td>870,869</td>
<td>86.2%</td>
<td>37.3%</td>
<td>34.9%</td>
</tr>
<tr>
<td>20-25</td>
<td>630,830</td>
<td>761,841</td>
<td>82.8%</td>
<td>30.7%</td>
<td>26.9%</td>
</tr>
<tr>
<td>&gt;25</td>
<td>1,085,466</td>
<td>1,497,545</td>
<td>72.5%</td>
<td>23.9%</td>
<td>19.0%</td>
</tr>
<tr>
<td>Public Housing$^e$</td>
<td>34,988</td>
<td>42,431</td>
<td>82.5%</td>
<td>27.0%</td>
<td>23.9%</td>
</tr>
</tbody>
</table>

Notes:

$a$ Data are computed as 20 times the average (1999-2014) number of Equifax primary individuals ages 18-64.

$b$ Data shown are the average (2000 and 2010) of the Census tract population ages 18-64.

$c$ Data are the average share of households in Equifax with a mortgage, 2008-2012.

$d$ Data are the average share of households in the American Community Survey tract-level tabulations with a mortgage, 2008-2012.

$e$ Data shown are for 13 tracts with 250+ non-senior public housing units and above 10% poverty rate in 2000.

We stratify households into types using an 8-step stratification procedure. We begin with the full sample, and subdivide the sample into smaller “cells” based on (in this order): The racial plurality, as measured by the 2000 Census, of the 2000 Census block of residence (4...
5 age categories (cutoffs at 30, 45, 55, and 65), number of adults age 18 and older in the household (1, 2, 3, 4+), and then the presence of an auto loan, credit card, student loan and consumer finance loan. We do not subdivide cells in cases where doing so would result in at least one new smaller cell with fewer than 20,000 observations. In a final step applied to all bins, we split each bin into three equally-populated types based on within-bin credit-score terciles. After all the dust settles, this procedure yields 144 types of households.

The benefit of working with a data set like the FRBNY Consumer Credit Panel / Equifax data is that its size allows estimates of the substitutability of neighborhoods, i.e. the vector of $\delta_j$, to vary based on a rich set of observables, explaining why we use so many types. Much smaller panel data sets simply do not allow for this and the number of types in estimation is typically small: For example, Kennan and Walker (2011) use 2 types in estimation. The following figures from our data are instructive. Figure 1 shows the typical location choices made by type 133 in our sample: A 2-adult household with an Equifax Risk Score $^TM$ below 580 and first observed living in a Census block that is predominantly black. The light blue areas show all Census tracts with poverty rates less than 10% and the tan areas show all Census tracts with higher poverty rates. The areas in dark blue show the most chosen low-poverty Census tracts for this type and the areas in black show the most chosen high-poverty tracts. Figure 1 shows this type predominantly clusters its location choices in one crescent-shaped area in the south-central part of the county. Figure 2 shows the same set of location choices for type 20 in our sample, a 2-adult household with a 590-656 Equifax Risk Score $^TM$ first observed in a predominantly Hispanic Census block. Comparing figures 1 to 2, few of the most popular neighborhood choices overlap of these two types. If, counterfactually, we assumed that the vector of $\delta_j$ of the two types were the same, the model would attribute the systematic variation in optimal neighborhood choices entirely to differences in the i.i.d. utility shocks experienced.

Our sample is comprised of 1,748 Census tracts. Allowing a separate value of $\delta$ for each tract and for each type would require estimating more than 250,000 parameters. Conceptually, with a large enough sample we could separately estimate every $\delta$ by type. Currently, for each type of household in our sample, we have data on approximately 2,000 households.

---

10 We assign race based on the racial plurality of all persons in the Census block, owners and renters, when they are first observed, which in most cases is 1999. The mean number of households and residents at the Census-block level in our sample of 1,748 tracts is 41 and 118, respectively, and Census blocks are highly homogenous by race and by tenure choice. Of the Census blocks in our sample that are at least 5% Hispanic, 26% are 75% or more Hispanic. The equivalent statistic for whites is 27%, for African Americans 9%, for Asians 2%. Similarly, of the Census blocks in our sample tracts that are at least 5% renter occupied, 25% are 75% or more renter occupied.

11 Whenever we refer to a household “age” in the FRBNY Consumer Credit Panel / Equifax data, we are referring to the age of the person in the household in the initial random sample. We are not using the ages of any other people in the household.
Figure 1: Location Choices of 2-adult black households w/ <580 Equifax Risk Score$^{TM}$

Figure 2: Location Choices of 2-adult Hispanic households w/ 590-656 Equifax Risk Score$^{TM}$
followed over 10 years. Therefore, for parsimony, and to exploit the fact that geographically nearby tracts likely provide similar utility, for each type we specify that the utility of location \( j \), \( \delta_j \), is a function of latitude \( (\text{lat}_j) \) and longitude \( (\text{lon}_j) \) of that location according to the formula

\[
\delta_j = \sum_{k=1}^{K} a_k B_k (\text{lat}_j, \text{lon}_j)
\]

The \( B_k \) are parameter-less basis functions. For each type, we use \( K = 89 \) basis functions. Additionally, we allow the values of \( a_k \) to vary for tracts above and below 10% poverty threshold. Inclusive of the two moving cost parameters, we estimate \( 2 \times 89 + 2 = 180 \) parameters per type. With 144 types, we estimate a total of 25,920 parameters.

To define the log likelihood that we maximize we need to introduce some more notation. Let \( i \) denote a given household, \( t \) a given year in the sample, \( j_{it} \) as person \( i \)'s starting location in year \( t \) and \( \ell_{it} \) as person \( i \)'s observed choice of location in year \( t \). Denote \( \tau \) as type and the vector of parameters to be estimated for each type as \( \theta_\tau \). The log likelihood of the sample is

\[
\sum_{\tau} \sum_{i \in \tau} \sum_{t} p (\ell_{it} \mid j_{it}; \theta_\tau)
\]

(8)

\( p(.) \) is the model predicted log-probability of choosing \( \ell_{it} \) given \( j_{it} \). For each \( \tau \) we use the quasi-Newton BFGS procedure to find the vector \( \theta_\tau \) that maximizes the sample log likelihood.

### 2.3 Estimates and Model Fit

Our estimation procedure ultimately yields estimates of \( \delta_j \), \( \kappa_0 \) and \( \kappa_1 \) for each type to match model-predicted moving probabilities to those in the data.\(^{12}\) Figures 3 and 4 show the surface of indirect utilities across Los Angeles County that we estimate for types 133 and 20, respectively, such that the model can replicate as best as possible the location choices shown in figures 1 and 2. These figures illustrate the flexibility of our specification. These surfaces are quite different, reflecting the very different optimal location choices of these types.

Due to our large number of types and tracts, it is impossible to report all parameter estimates. Instead, we summarize the estimates by examining the model’s in-sample fit along a number of dimensions. Figure 5 compares actual and model-predicted annual migration rates in our sample. About 8.5 percent of our sample moves to a different tract in each year, and that percentage falls from just above 11 percent for those under 30 to just above 3 percent for those aged 65 and above. Figure 6 compares the distribution of distances moved

\(^{12}\)We fix \( \beta = 0.95 \).
Figure 3: Indirect Utility, Type 133

Figure 4: Indirect Utility, Type 20
Figure 5: Model Fit: Moving Probabilities

( measured as the straight line distance between tract centroids) for all movers in the data and as predicted by our model. This figure shows that the model replicates the hump-shaped distribution of distances moved, with the most frequent moves about 4 miles. The model slightly overpredicts moves between 4 and 10 miles in length and slightly underpredicts moves less than 4 miles.

Figure 7 shows a detailed comparison of model-predicted and actual annual migration rates for households that choose to move by poverty rate of Census tracts. The tracts from which people are moving are split into six groupings based on the poverty rate of the originating tract: 0-5, 5-10, 10-15, 15-20, 20-25 and >25. For each of these groupings, the probability of choosing a destination tract of a given poverty rate is plotted for the data (dark blue solid line) and as predicted by the model (light blue dotted line). Figure 7 shows model fit for very low-probability moves. The model tends to under-predict the probability that households living in low-poverty tracts move to a low-poverty tract, conditional on a move occurring. Aside from that, in our view the model fits the data well along this dimension.

\[\text{In the data we know the Census block of residence for each household. We eliminate any within-tract moves and for the remaining moves, we define distance moved as the distance between tract centroids of the sending and receiving tracts.}\]

\[\text{Recall the unconditional probability of any move is less than ten percent.}\]
2.4 Type-Specific Sensitivity to Rent

To understand the impact of a rent subsidy program such as MTO on neighborhood choice, we need to understand how utility of each neighborhood varies with rents paid to live in that neighborhood. Denote as $\tilde{\delta}_{j\tau}$ our estimate of indirect utility of neighborhood $j$ for a given type $\tau$. To make progress, we specify that $\tilde{\delta}_{j\tau}$ is a linear function of rent, observable characteristics of tract $j$, $O_j$, and unobserved characteristics of tract $j$, $\zeta_j$

$$\tilde{\delta}_{j\tau} = -\alpha_{\tau} \cdot \text{rent}_j + \lambda_{\tau} \cdot O_j + \zeta_j$$

The parameter $\alpha$, the rate at which indirect utility varies with rents, cannot be estimated using OLS because equilibrium rents will almost certainly be correlated with unobserved but valued characteristics of neighborhoods, $\zeta_j$. An instrument is required. We use a three-step IV approach to estimate $\alpha$ that is common in the IO and Urban literature, for example Bayer, Ferreira, and McMillan (2007).

In the first step of our procedure, we estimate $\alpha_{\tau}$ using two-stage least squares. We include characteristics of the housing stock 0-5 miles from tract $j$ in $O_j$ as controls (number of rooms, number of units in the housing structure and age of structure) and use characteristics of the
Figure 7: Poverty Category Transitions $t-1$ to $t$, Conditional on Moving

(a) From 0-5%

(b) From 5-10%

(c) From 10-15%

(d) From 15-20%

(e) From 20-25%

(f) From >25%
housing stock 5-20 miles from the tract as instruments for rent. The first-stage F-statistic is 7.

In the second step, we use estimates of $\alpha$ and $\lambda$ from the first step, call them $\tilde{\alpha}_\tau$ and $\tilde{\lambda}_\tau$, to construct a new surface of indirect utilities for each type abstracting from unobservables as

$$\hat{\delta}_{j\tau} = -\tilde{\alpha}_\tau \cdot rent_j + \tilde{\lambda}_\tau \cdot O_j$$

We simulate the model using this specification for indirect utility and adjust $rent_j$ for all $j$ until the simulated steady-state number of households in any tract is equal to the average number of households in our estimation sample in that tract. This procedure determines market-clearing rents in all tracts in the absence of unobserved amenities. We use these rents as instruments to estimate alpha in the third and final step with an F-statistic of 34. Intuitively, the F-statistic rises from 7 to 34 because the first step only uses information about the quality of substitutes for each tract individually whereas the third step uses similar information for all tracts.

We find remarkable variation in our estimates of $\alpha$ by type. We summarize this variation in Figure 8 which graphs the average value of $\alpha$ by initial Census tract of residence for the people in our estimation sample. The figure shows that people living in high poverty tracts are, on average, nearly three times more sensitive to changes in rent as people living in the lowest poverty areas.

3 Neighborhood Effects

In this section, we use confidential panel data from the Los Angeles Family and Neighborhoods Survey (LA FANS) to study how neighborhoods impact child cognitive abilities. The LA FANS study was designed specifically to investigate neighborhood influences on a variety of outcomes for families, adults, and children; see Pebley and Sastry (2011). The survey stratified 65 Census tracts using 1990 boundaries in Los Angeles County. Roughly 50 households in each Census tract were selected at random for inclusion in the survey. A randomly selected adult in the household was interviewed, as well as a randomly selected adult.

The intuition for the validity of these instruments arises directly from the Rosen-Roback model. Consider two pairs of tracts, $(A, B)$ and $(A', B')$, with $A$ and $A'$ providing identical direct utility and the housing stock in $B'$ of higher quality than the housing stock in $B$. Assume one set of households chooses between $A$ and $B$ and a different set of households chooses between $A'$ and $B'$. In equilibrium, $A$ will have a higher rental price than $A'$ because $B$ is of lower quality than $B'$, despite the fact that $A$ and $A'$ yield identical direct utility.

The average value of $\alpha$ varies by Census tract because the mix of types varies by tract.
child. If the household had more than one child, a randomly selected sibling was also interviewed. Further, if the selected child’s mother was in the household, she was interviewed as the primary caregiver. If she was absent, the actual primary caregiver was interviewed.

The LA FANS data has the advantage of sampling by Census tract, so that we observe many households within a small geographic region. The LA FANS oversamples poor neighborhoods, but the 65 Census tracts are distributed across much of Los Angeles. Figure 9 shows the distance of each tract in our Los Angeles sample, as defined earlier, to a tract in the LA FANS sample. Most tracts in Los Angeles are located within a few miles of an LA FANS tract, but on average high-poverty Census tracts are closer to an LA FANS tract, reflective of the LA FANS sampling design. 3,085 households were interviewed between 2000 and 2002 (wave 1), of which 1,242 were re-interviewed between 2006 and 2008 (wave 2). New households were admitted into the LA FANS sample in the second wave. Detailed information on the housing status (rentership versus ownership), family characteristics and child outcomes were collected from respondents and Census tract information was collected in both waves.

For cognitive skill measures we study the child’s score on Woodcock Johnson tests as

---

17This is in contrast with other geo-coded panel datasets such as the Panel Survey of Income Dynamics or the National Longitudinal Study of Youth.

18We are unable to show the spatial distribution of the sampled tracts due to confidentiality restrictions.
described in Schrank, McGrew, and Woodcock (2001) for applied problems ("math"), a test used in many MTO studies.\textsuperscript{19} We restrict our sample to children who had valid measurements for both waves and we eliminate from our sample children with missing observations in some of our control variables.\textsuperscript{20} This reduces our sample to 1,253, about 20 children per tract to estimate value-added. This is roughly the same sample size as studies of teacher value-added, i.e. one classroom of children.\textsuperscript{21}

We compute neighborhood value-added using standard techniques in the education literature for computing teacher value-added. Following, Kane and Staiger (2008) and Chetty, Friedman, and Rockoff (2014) for example, we work with the statistical model for the pro-

\textsuperscript{19}We have also studied results for passage comprehension. Our results are qualitatively and quantitatively very similar and as a result we do not discuss them in the paper.

\textsuperscript{20}We include all children, including those that change locations, in our estimation sample. Children that change locations between waves are assigned to the Census tract of their location in the first wave. We did not exclude movers from the sample for fear of sample selection. This choice was necessitated by the fact that LA FANS does not provide coverage for all Census tracts, including the tracts that are the destination of household moves in our sample. Our estimates can be interpreted as average annual value-added over a 5-year period for a given tract, conditional on starting the 5-year span in that tract.

\textsuperscript{21}A major reason for a lack of skill measurement in both waves is the child’s age. Only children under 18 were administered the Woodcock Johnson tests and thus only children who were under 18 in wave 2, i.e. aged 4 to 14 in wave 1 depending on the interview timing, are included. Additionally, new entrants to the survey would be disqualified since we only observe their test scores once.
duction of the change in child ability, $\Delta t-T A_{i,j,t}$, between periods $t-T$ and $t$ ,
\[ \Delta t-T A_{i,j,t} = Z'_{i,j,t-T} \psi + v_{i,j,t} ; \quad v_{i,j,t} = T [ \mu_j + \epsilon_{i,j,t} ] ; \quad (9) \]
where $i$ indexes children, $j$ indexes neighborhoods, $t$ indexes time, $Z_{i,j,t-T}$ is a vector of observable child and family characteristics measured at time $t-T$, $\mu_j$ is a causal (annualized) neighborhood “value-added” effect, $\epsilon_{i,j,t}$ is an idiosyncratic child/family effect and $T$ is the number of years between LA FANS waves.\(^{22}\)

Notice that in the absence of any control variables, $\mu_j$ would govern the average change in child ability over time for children living in neighborhood $j$. Consistent with the value-added approach, splines of lagged values of a behavioral problems index as described in Peterson and Zill (1986) are included as controls. Our other controls include variables covering family structure (number of children), age, race, gender of child, parental IQ, parental education and income and assets, all measured as of wave 1. We present descriptive statistics in Table 2.

The key insight to the value-added approach is that parents’ optimal neighborhood choice does not have to be uncorrelated with the observable control variables, including lagged child test scores, to produce unbiased estimates of neighborhood effects on child ability. Due to the presence of neighborhood fixed effects in equation (9), $\psi$ is identified purely by within-neighborhood variation of $Z_{i,j,t-T}$ and $\Delta t-T A_{i,j,t}$. Parents can select neighborhoods based on $Z_{i,j,t-T}$ and that will not bias estimates of $\psi$.\(^{23}\) For an unbiased estimate of $\mu_j$, the error term $\epsilon_{i,j,t}$ must be uncorrelated with $Z_{i,j,t-T}$. Parents can select neighborhoods based on the level of their child’s ability and/or other variables in $Z_{i,j,t-T}$, but not on the portion of expected growth of child ability that is not forecasted by $Z_{i,j,t-T}$.

Table 3 summarizes our regression results of equation (9), showing model fit across a number of specifications. The outcome variable is the change in the standardized test score between LA FANS waves. When tract-level fixed effects are the only regressors, model 1, the $R^2$ of the regression is just 9%. Once information on lagged child test scores is included as a regressor (model 2) the $R^2$ jumps to 41%. Adding child controls (model 3) and parent demographics (model 4) increases the $R^2$ to 52%. Adding information on parental income and assets (model 5) fails to further boost $R^2$ values. Given the $R^2$ value stays constant between models 4 and 5, we infer that for our results to be misleading, selection into neighborhoods based on $\epsilon_{i,j,t}$ must account for a significantly larger share of

\(^{22}\)We include the $T$ term when defining $v_{i,j,t}$ so that $\mu_j$ and $\epsilon_{i,j,t}$ are annualized.

\(^{23}\)Ioannides and Zanella (2008) estimate a model of location choice at the Census-tract level using panel data from the PSID and show that parents with young children are more likely to select neighborhoods with desirable observable characteristics used in the production of child human capital than other households.
Table 2: Descriptive Statistics, LA FANS

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in math score</td>
<td>-0.009</td>
<td>1.034</td>
<td>1253</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Control Variables</strong> (LA FANS Wave 1)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1 Test Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math score</td>
<td>0.000</td>
<td>1.000</td>
<td>1253</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Child Demographics</strong></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of Child (years)</td>
<td>8.148</td>
<td>4.919</td>
<td>1253</td>
</tr>
<tr>
<td>Hispanic (1=yes)</td>
<td>0.570</td>
<td>0.495</td>
<td>1253</td>
</tr>
<tr>
<td>Black (1=yes)</td>
<td>0.126</td>
<td>0.332</td>
<td>1253</td>
</tr>
<tr>
<td>Male (1=yes)</td>
<td>0.520</td>
<td>0.500</td>
<td>1253</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Parental Demographics and Education</strong></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of kids</td>
<td>2.570</td>
<td>1.222</td>
<td>1253</td>
</tr>
<tr>
<td>Parental IQ</td>
<td>87.690</td>
<td>15.082</td>
<td>1253</td>
</tr>
<tr>
<td>High School dropout</td>
<td>0.272</td>
<td>0.445</td>
<td>1253</td>
</tr>
<tr>
<td>High School graduate</td>
<td>0.197</td>
<td>0.398</td>
<td>1253</td>
</tr>
<tr>
<td>Some college</td>
<td>0.307</td>
<td>0.461</td>
<td>1253</td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>0.105</td>
<td>0.306</td>
<td>1253</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>0.063</td>
<td>0.243</td>
<td>1253</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Parental Income and Assets ($000s)</strong></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log income</td>
<td>3.799</td>
<td>1.174</td>
<td>1052</td>
</tr>
<tr>
<td>Log assets</td>
<td>2.404</td>
<td>2.005</td>
<td>1135</td>
</tr>
</tbody>
</table>

* Income and assets data are not always available for our estimation sample, explaining the smaller sample sizes for those variables.
Table 3: $R^2$ Values from LA FANS data

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Neighborhood Fixed Effects Only</td>
<td>0.09</td>
</tr>
<tr>
<td>2: + Splines in Lagged Child Scores</td>
<td>0.41</td>
</tr>
<tr>
<td>3: + Splines interacted w/ Child Controls</td>
<td>0.51</td>
</tr>
<tr>
<td>4: + Parent Ability and Demographics</td>
<td>0.52</td>
</tr>
<tr>
<td>5: + Lagged Income and Assets</td>
<td>0.52</td>
</tr>
</tbody>
</table>

65 tracts, 1,253 observations

observed differences in change in average ability across neighborhoods than selection into neighborhoods based on parental education, income and assets (Altonji, Elder, and Taber, 2005).

There are two issues we address before continuing. First, LA FANS only covers 65 tracts in Los Angeles but we require an estimate for all the 1,748 Census tracts in our sample. Second, following the teacher value-added literature (Chetty, Friedman, and Rockoff, 2014), we shrink the variance of the estimates of value-added arising from equation (9) to account for the fact that these estimates are derived from small samples and are noisy.

We perform the interpolation and shrinkage using a two-step process. To understand this process, let $k$ (or $k'$, as needed) denote an LA FANS Census tract. In the first step, we estimate equation (9) using the LA FANS data. Define $\hat{\mu}_k$ as the estimate of tract-k’s annual fixed effect, $\hat{\sigma}_\mu^2$ as the estimated variance of the tract-level fixed effects and $\hat{\sigma}_\epsilon^2$ as the estimate of the variance of annual changes in child ability after controlling for all $Z$ variables and neighborhood effects arising from this first step. Now let $j$ represent any tract in Los Angeles and define $\omega_{j,k}$ as a “weight” based on the physical distance between tracts $j$ to $k$, a “distance” between tracts $j$ and $k$ in attribute space, and the number of observations in tract $k$, $N_k$. Specifically, define

$$\omega_{j,k} = N_k \times \phi \left( \frac{\|j - k\|_{distance}}{h_1} \right) \times \phi \left( \frac{\|j - k\|_{attributes}}{h_2} \right)$$

where $h_1$ and $h_2$ are bandwidths and $\phi(.)$ is the standard Normal density function. The term $\|j - k\|_{distance}$ is the physical distance (in miles) between the centroids of tracts $j$ and $k$. The “distance” in attribute space $\|j - k\|_{attributes}$ is the difference between the value-added measures of $j$ and $k$ predicted by a regression of value-added on a host of observable tract
characteristics.\textsuperscript{24,25} We compute annual value-added for tract $j$ as

\[
\left( \frac{\sum_k \omega_{j,k} \hat{\mu}_k}{\sum_{k'} \omega_{j,k'} } \right) \left( \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2 + \hat{\sigma}_\epsilon^2 / \tilde{N}_j } \right) \tag{10}
\]

Interpolation \hspace{1cm} Shrinkage

where $\tilde{N}_j$ is defined as

\[
\left( \frac{\sum_k \omega_{j,k}}{\sum_{k'} (\omega_{j,k'}/N_{k'})} \right)^2 \tag{11}
\]

The interpolation term in equation (10) is straightforward, as it is a simple weighted average. To understand the shrinkage term and why it is standard in the teacher value-added literature, consider a simplified model where $\Delta a$ is the change in the next child’s test score, $\mu$ is the true neighborhood effect and $\epsilon$ is a child-specific shock. Suppose that a noisy estimate of $\mu$, call it $\mu^o$, is observed

Truth: $\Delta a = 1 \cdot \mu + \epsilon$
Observed: $\mu^o = \mu + \nu$ \tag{12}

with $\nu$ being measurement error. A regression of $\Delta a$ on $\mu^o$ will yield a biased coefficient of $\sigma_\mu^2 / (\sigma_\mu^2 + \sigma_\nu^2)$. Dividing estimates of $\mu^o$ by this expression will produce an unbiased regression coefficient of 1. In mapping the intuition of equation (12) to what we actually do, note that the variance of $\nu$ – the variance of the measurement error – will be a function of the sample size in the LA FANS data. The reason is that we estimate value-added as a fixed effect, which is a sample average. The greater the number of observations in each tract, the more precisely we estimate neighborhood value-added and the smaller the variance of $\nu$. This explains the presence of the $\tilde{N}_j$ term in equation (10). The fact that we use a weighted average of all LA FANS tracts in estimating value-added for any given Census tract leads to the functional form for sample size of equation (11).

Table 4 shows tract-level correlations of value-added estimates for the five different models

\textsuperscript{24}The list of explanatory variables includes tract poverty rate, median income, share receiving public assistance, crime rate, an index of transportation access, share Hispanic, and share black.

\textsuperscript{25}We use $h_1 = 1.5$ miles, and we set $h_2$ to the standard deviation of the predicted value-added measures across tracts. A wide range of bandwidths (i.e. a range of relative weights placed on physical and attribute distance in the interpolation) yield nearly identical results, consistent with the high degree of spatial correlation in observable characteristics across tracts.
Table 4: Correlation of Value-Added Estimates by Tract

All 1,748 tracts after interpolation and shrinkage has occurred

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>0.90</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.52</td>
<td>0.80</td>
<td>0.94</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>0.79</td>
<td>0.91</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Ann. Std. Dev. | 0.045 | 0.039 | 0.040 | 0.037 | 0.037 |

discussed in table 3 after interpolation and shrinkage have occurred for all 1,748 Census tracts in our study. This table reinforces the result that once lagged child controls are included as regressors (model 2), estimates of tract value-added from models that include more controls are very similar (models 3-5), as the correlations are 0.79 and above. The bottom rows report the estimated standard deviation of tract-level child value-added. In model 4, the specification we use in our counterfactual simulations later in the paper, the standard deviation of tract-level child value-added is 0.037. Note that the unconditional standard deviation of the level of the Woodcock-Johnson score is 1.0. Assuming linearly additive effects of neighborhood value-added over time, 10 years of exposure to a Census tract with a level of child value-added that is one standard deviation above the mean will cause a child’s Woodcock-Johnson test scores to increase 37% of one standard deviation.

Table 5 shows regressions of our value-added estimates on measures of local public school quality, tract poverty rates and tract-level racial percentages. We use a bootstrapping procedure to compute the standard errors shown in the table. The estimates of local school quality are estimates of math and reading value-added of the nearest elementary school as produced by the Los Angeles Times. The regressions show that our estimates of value-added are not simple transformations of race, poverty or public-school quality. There is considerable variation in value-added even after controlling for public school quality, tract

26 The results are very similar when we restrict the analysis to only the tracts with LA FANS data but still apply interpolation and shrinkage.

27 To compute bootstrap standard errors, we draw 1,000 LA FANS samples and for each LA FANS sample we draw 1,000 samples of 1,748 Census tracts. This gives us 1 million draws in total. In each LA FANS sample, we draw from all the 65 LA FANS tracts. The number of children drawn in each tract is fixed and equal to the LA FANS sample size. The LA FANS and Census tracts samples are both drawn with replacement.

28 See http://projects.latimes.com/value-added/ for details on how school value-added measures are computed. We assign the elementary school that is closest in distance to the centroid of the Census Tract.
Table 5: Neighborhood Traits and Value-Added

Regr. of Value-Added Estimates on Neighborhood Covariates, 1,748 Tracts
(Bootstrap Standard Errors in Parentheses)

<table>
<thead>
<tr>
<th>Variable</th>
<th>0.025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math School VA+</td>
<td>(0.045)</td>
</tr>
<tr>
<td>English School VA+</td>
<td>0.064</td>
</tr>
<tr>
<td>Poverty Rate</td>
<td>-0.003</td>
</tr>
<tr>
<td>(0.051)</td>
<td></td>
</tr>
<tr>
<td>Pct. Hispanic</td>
<td>0.063***</td>
</tr>
<tr>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>Pct. Black</td>
<td>-0.017</td>
</tr>
<tr>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>Pct. Hispanic x Poverty Rate</td>
<td>0.046</td>
</tr>
<tr>
<td>(0.066)</td>
<td></td>
</tr>
<tr>
<td>Pct. Black x Poverty Rate</td>
<td>0.069</td>
</tr>
<tr>
<td>(0.106)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.032</td>
</tr>
<tr>
<td>(0.010)</td>
<td></td>
</tr>
</tbody>
</table>

heightR2   0.141

+ LA Times Measure of Local Public Elementary School Value Add
*** Significant at a 1% confidence level
level poverty rates and racial percentages, as the $R^2$ of the regressions is only 14%.

Upon further review, a case can be made that our estimates of tract value-added are capturing an aspect of the neighborhood that is distinct from available estimates of public-school quality. Figure 10 plots our estimates of the average level of tract-level value-added by poverty rate in the top panel and the average level of public school quality as measured by the Los Angeles Times, also by poverty rate, in the bottom panel. There is considerable variation around the tract-level averages shown in figure 10 (not shown), but on average our estimates of value-added decline with tract poverty rates. In contrast, the Los Angeles Times’ estimates of school quality increase with tract poverty.

4 Out of Sample Validation: Analysis of MTO

In the next section of the paper, we use counterfactual simulations of our model to compute the impact of various hypothetical voucher policies on child outcomes. To lend credibility to those results, in this section we ask if our model can replicate the findings of the “Moving to Opportunity” (MTO) randomized experimental intervention, the first large-scale (experimental) program to explicitly link housing vouchers to specific neighborhoods. Since we use no MTO data in our estimation, we view this section as a test of out-of-sample fit.

Moving to Opportunity was a randomized control trial beginning in the 1990s that randomly assigned a group of households with children eligible to live in low income housing projects in five U.S. cities to three different groups: (i) a treatment group that received a Section 8 housing voucher that in the first year could be applied only in Census tracts with a poverty rate under 10% and could be applied unconditionally thereafter, (ii) a second treatment group that received a comparable Section 8 housing voucher with no location requirement attached, and (iii) a control group that received no voucher. Voucher amounts were set such that after applying the voucher, households spend no more than 30% of their income on rent. Households that wanted to rent a more expensive unit could only contribute up to an additional 10% of their income.

Summarizing the medium- to long-term impacts of MTO, Sanbonmatsu, Kling, Duncan, and Brooks-Gunn (2006), Kling, Liebman, and Katz (2007) and others show that on average the MTO treatment successfully reduced exposure to crime and poverty and improved the mental health of female children, but failed to improve child ability, educational attainment or physical health.
Figure 10: Tract Poverty, Value-Added and School Quality

(a) Avg. Tract Value-Added against Poverty Rates

(b) Avg. School Quality (Math) against Poverty Rates
To see if our estimated location-choice model can replicate these results, we simulate optimal decisions under several policy scenarios, restricting analysis to the households in our sample likely to have been eligible for MTO had they lived in an MTO area at the time of the experiment. Our three scenarios are as follows:  

- **(Baseline)** No subsidies or vouchers.

- **(MTO-A)** MTO style vouchers. Households who move to a Census tract with a poverty rate under 10% at $t = 1$ receive a Section 8 housing voucher. This voucher is received in perpetuity, even if the household moves out of a qualifying neighborhood in period $t > 1$. If a type-$\tau$ household is offered and accepts a voucher and subsequently lives in neighborhood $j$, we set the utility of that neighborhood equal to our original estimate, $\tilde{\delta}_{j\tau}$, plus $\alpha_\tau$ times the voucher amount. The annual voucher we use is $6,000, which we set such that the average MTO-eligible household can rent a 2-bedroom unit costing $766 per month after spending 30% of monthly income. We assume that households receiving a voucher spend the entire amount of the voucher each period.

- **(MTO-B)** Randomly assigned poverty reduction. We assign households to neighborhoods randomly according to the distribution of neighborhood poverty-rates that arises under scenario MTO-A. Comparisons of MTO-B and MTO-A highlight the role of neighborhood selection conditional on accepting a voucher.

To summarize the expected impact on child ability of the various MTO policies we consider, we compute an expected measure of accumulated neighborhood value-added exposure conditional on accepting a voucher. Let $i'$ denote a family that accepts a voucher in the MTO-A experiment, and assume there are $i' = 1, \ldots, I$ such families. For any given simulation draw $s$, we hold this set of families fixed for each of the three scenarios (policies) we consider. 

---

32 Our simulations target households residing at $t = 0$ in a Census tract with at least 250 non-senior citizen public housing units, 13 tracts total. While a few of the developments contain a small share of units set aside for senior citizens, these are predominately public housing developments for families with children. Note that we cannot restrict our simulations to households with children, as we do not know which households in the FRBNY Consumer Credit Panel / Equifax data have children.

33 Our calculation is $6,000 \approx 12 \times (766 - 0.30 \times 10,000/12)$, where $10,000$ is mean household income of the MTO-eligible population as computed by Galiani, Murphy, and Pantano (2015) and $766$ is the “payment standard” (max voucher amount) for a 2-bedroom apartments in Los Angeles in 2000.

34 Specifically, the procedure is: (1) pool the set of MTO-A simulated Census tract choices and the unconditional list of sample Census tracts. (2) Estimate a probit model predicting the probability that a record comes from the simulated data using only tract-poverty-rate categories as explanatory variables, and obtain the predicted probability $p_j$ (propensity score) that a record from tract $j$ comes from the simulated data. (3) Draw MTO-B simulated locations from the full set of Census tract with probability $Pr(j) = \frac{1}{J} \left( \frac{p_j}{1 - p_j} \right) \left( \frac{1 - p_j}{p_j} \right)$.

35 This is the impact of the treatment on the treated.
consider: Baseline, MTO-A and MTO-B. We then compute the expected impact of policy $p$ on child value-added measured over $\bar{T}$ periods (5, 10 or 18 years) as

$$\hat{\mu}^{TOT}_p = \frac{1}{S} \sum_{s=1}^{S} \left[ \frac{1}{\bar{T}} \sum_{t' = 1}^{\bar{T}} \sum_{t = 1}^{\bar{T}} \hat{\mu}_{\ell(i',t,s,p)} \right]$$

(13)

where $\ell(i', t, s, p)$ is the location chosen by family $i'$ in year $t$ under policy $p$ and for given simulation draw $s$ and $\hat{\mu}_{\ell(i',t,s,p)}$ is the value-added associated with $\ell(i', t, s, p)$. For each type, we run $S = 10,000$ simulations, yielding a total of 1.44 million simulations for each policy experiment. If, as suggested by Chetty and Hendren (2015), neighborhood effects are additive over time in the child ability production function (i.e. there are no complementarities across time periods) and neighborhood quality affects children equally at all ages, then these measures will characterize actual total neighborhood contributions to child ability. If child investments exhibit dynamic complementarities and early childhood investments are especially productive as in Cunha, Heckman, and Schennach (2010), these measures will understate neighborhoods’ long-term contributions to child ability. In either case, we view these measures as useful summaries for characterizing the impact of policy.

We compute standard errors around $\hat{\mu}^{TOT}_p$ to evaluate if the model-predicted outcomes from the baseline, MTO-A and MTO-B are statistically significantly different. Denote the number of types in estimation (144) as $T$ and the number of Census tracts (1,748) as $J$. Referring to notation in equation (8), we estimate the following sets of parameters

$$\{\theta_{\tau}\}_{\tau=1}^{T}, \{\alpha_{\tau}\}_{\tau=1}^{T}, \mathcal{M}$$

(14)

where $\theta_{\tau}$ is a vector of 180 parameters determining location choice for type $\tau$ and $\mathcal{M} = \{\mu_j\}_{j=1}^{J}$ is the vector of parameters determining child value-added in all Census tracts.

$\theta_{\tau}, \alpha_{\tau}$ and $\mathcal{M}$ are assumed to be drawn independently for all $\tau = 1, \ldots, T$. Denote $\Sigma_{\theta}$ as the variance-covariance matrix of $\theta_{\tau}, \sigma_{\alpha}$ as the variance of the estimate of $\alpha_{\tau}$ and $\Sigma_{\mathcal{M}}$ as the variance-covariance matrix of $\mathcal{M}$. The parameters in equation (14) are assumed to be

36We allow the set of families indexed by $i'$ to change across simulation draws.
Table 6: MTO Demonstration vs. Simulation Experiments

Impacts on Woodock-Johnson Math Scores (sd=1)

<table>
<thead>
<tr>
<th>Exposure time</th>
<th>MTO Demonstration (TOT)</th>
<th>Simulation Experiments</th>
<th>p-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 years</td>
<td>-0.019</td>
<td>-0.003</td>
<td>0.919</td>
<td>0.001</td>
</tr>
<tr>
<td>10 years</td>
<td>-0.052</td>
<td>-0.008</td>
<td>0.874</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>18 years</td>
<td>–</td>
<td>-0.002</td>
<td>–</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

distributed with a variance-covariance matrix of

\[
\begin{bmatrix}
\Sigma^\theta_1 & 0 & 0 & 0 \\
0 & \Sigma^\theta_2 & 0 & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \Sigma^\theta_T
\end{bmatrix}
\]

\[
\begin{bmatrix}
\Sigma^\tau_1 & 0 & 0 & 0 \\
0 & \sigma^\tau_2 & 0 & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \sigma^\tau_T
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 & 0 & 0 & \Sigma^M
\end{bmatrix}
\]

For \(\Sigma^\theta_T\) and \(\sigma^\tau_T\) we use asymptotic standard errors and for \(\Sigma^M\) we use a bootstrap procedure where we sample from the raw LA FANS data and run the sampled data through the process described in the previous section. To compute standard errors on our policy experiments, we draw parameters from this distribution 3,000 times and compute \(\hat{\mu}_{p}^{TOT}\) for each draw according to equation (13).

Table 6 shows our estimates of \(\hat{\mu}_{p}^{TOT}\). The first column shows results from the actual MTO demonstration, as reported by Sanbonmatsu, Kling, Duncan, and Brooks-Gunn (2006) and Sanbonmatsu, Ludwig, Katz, Gennetian, Duncan, Kessler, Adam, McDad, and Lindau (2011), and column 2 shows the simulated impact of MTO-A relative to the baseline. Column 1 highlights that MTO researchers found no impact of the voucher program on child ability.

31
after 5 and 10 years and column 2 shows that our model can replicate this finding, despite not using any MTO data in our analysis. Column 4 verifies that we can not reject the hypothesis that our MTO-A results are identical to the results from the actual MTO data. In our view our model passes this out-of-sample fit test.\footnote{One way in which our model does not match the data is in the take-up rate of the voucher. In our MTO-A simulation, only 45\% of the population eligible to receive a voucher accept it whereas the actual MTO take-up rate in Los Angeles was 67\%. We attribute the difference in take-up rates to the additional counseling that MTO offered as noted by \textit{Galliani, Murphy, and Pantano (2015)}. We can trivially modify the model to match the MTO take-up rate by adding one parameter that reduces the moving cost associated with the initial move required when accepting an MTO voucher. This would increase the MTO take-up rate but would not change the relative frequency with which any particular voucher-eligible neighborhoods are chosen.}

Column 3 reports the results of the MTO-B simulations. These demonstrate that when accumulated over a full 18-year childhood, the poverty reduction generated by MTO would improve math scores by 0.2-0.3 standard deviations if low-poverty neighborhoods were assigned at random to households accepting a voucher. These are substantial impacts, equivalent to closing about 20\% - 30\% of the black/white achievement gap according to \textit{Yeung and Pfeiffer (2009)}. Thus, the MTO experiment could have significantly improved child test scores if voucher-eligible tracts were chosen more selectively, motivating our work in the next section. Column 5 shows that we can reject the hypothesis that the results from MTO-B are the same as those in the MTO-A experiment.

The comparison of MTO-A to MTO-B shows that households receiving an MTO voucher selected into neighborhoods with low child value-added. Ultimately, for the types receiving housing vouchers, voucher-eligible neighborhoods with low child value-added have higher values of $\delta$ than those with high child value-added. One reason this result may occur is that the types of households likely to live in public housing are most sensitive to rent (see Figure 8) and the relative price of additional value-added is high in low-poverty neighborhoods. Figure 11 shows the relationship between composition-adjusted monthly rent in 2000 and neighborhood value-added for the 1,748 Census tracts in our study for three groups of Census tracts: Low poverty (0-10\%), middle (10-25\%), and high poverty (25\% and above).\footnote{We plot the expected rent in each Census tract for a 3-room unit built in 1960, computed as the outcome of a hedonic regression.} These figures show how the relative price of neighborhood quality changes with tract poverty rates. The change in rent associated with an increase in neighborhood quality is greatest in low poverty areas; that is, the slope of the green line (low poverty) is greater than the slope of the blue line (middle) and red line (high poverty). Even though neighborhoods with high value-added are relatively expensive in low poverty tracts, our analysis in the next section builds on the notion that households may be willing to pay to live in those neighborhoods if they receive a large enough rent subsidy.
5 Analyzing Other Voucher Policies

Abstracting from moving costs, the utility of living in tract $j$ for households of type $\tau$ given a voucher of size $V_j$ (for specific use in tract $j$) is

$$\delta_{j\tau} + \alpha_{\tau} V_j$$  \hfill (15)

If policy-makers want households to reside less frequently in (arbitrary) neighborhood $i$ and more frequently in $j$, they should either reduce voucher amounts to $i$, if any, or increase voucher amounts to $j$. As long as $\alpha > 0$, equation (15) shows that, holding preference shocks fixed, there is some voucher amount to live in location $j$ such that, no matter how large the initial difference in utility between tracts $i$ and $j$, households optimally accept the voucher and choose to live in $j$ rather than $i$.

Our MTO-A and MTO-B simulations show that MTO-subsidized households selected into especially low value-added tracts among the set of eligible low-poverty tracts. Ultimately, this occurred because the voucher amount did not sufficiently distinguish between low- and high-value-added tracts, and the lower value-added tracts among the eligible set provided higher utility. A partial explanation is that rents are relatively high in high-value-added

Figure 11: Scatterplots of Rental Prices and Child Value-Added by Poverty Rate

![Figure 11: Scatterplots of Rental Prices and Child Value-Added by Poverty Rate](image-url)
neighborhoods with low poverty rates (figure 11) and the types of households currently living in high poverty tract areas are especially sensitive to the level of rent (figure 8).

In the rest of this section, we consider the outcomes of a variety of possible voucher policies to see which policies are effective at inducing people to move to Census tracts with high measured value-added on child ability.\textsuperscript{39} Specifically, we compute the optimal location choices and child outcomes of four different policy experiments. We compute the responses and outcomes of the 10% of types (13 types) most likely to live in public housing and therefore most likely be eligible for a Section 8 housing voucher. In our baseline case, these types all receive a $600 monthly voucher independent of where they choose to live.\textsuperscript{40} In this section, we abstract from any general equilibrium effects. Our thought experiment is for a small, targeted voucher program that does not affect rents or housing supply in any Census tract.

Table 7 shows the distribution of characteristics for all renters in our sample (column 1), renters in tracts with housing projects (column 2) and the 10% of types in our sample most likely to live in tracts with housing projects (column 3).\textsuperscript{41} These statistics are all derived from all years of the FRBNY Consumer Credit Panel / Equifax data. When comparing columns 2 and 3, two differences jump out. In the types we consider in our experiments (column 3), no one starts our sample in a Census block with a white racial plurality. Additionally, the types we consider are more likely to have the household member chosen by the FRBNY Consumer Credit Panel / Equifax sampling design to be over the age of 65.

In each of the policy experiments we consider, households receive a voucher of a predetermined amount that only depends on where they live. Denote $V_{jp}$ as the voucher amount a household receives when it resides in tract $j$ in policy simulation $p$. $V_{jp}$ varies across tracts $j$ in a way we describe below, but within each policy simulation $V_{jp}$ is assumed to be fixed over time. Additionally, and different from the structure of the MTO experiment, in each period of the policy simulations households receive the voucher appropriate for their current tract of residence; the vouchers received are independent of history.

**VA-Index Targeting**

In the first two policies, we assume that policy-makers cannot directly observe tract-level value-added, but they can observe correlated characteristics and can design policies based on these characteristic values. We consider policies based on a mixture of three tract-level characteristics: The poverty rate, crime rate and accessibility to

\begin{footnotesize}
\textsuperscript{39}Of course policy-makers may have many objectives in mind when setting voucher policy. In this section, we focus on vouchers as a policy tool to improve child test scores.

\textsuperscript{40}Since our framework has no wealth effects, this is equivalent to a baseline in which no one receives a voucher.

\textsuperscript{41}See Bolton and Bravve (2012) for a description of the population receiving federally assisted housing.
\end{footnotesize}
<table>
<thead>
<tr>
<th></th>
<th>All L.A. Renters</th>
<th>Population of Housing-Project Tracts</th>
<th>10% Most Frequent Types in Housing-Project Tracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (non-Hispanic)</td>
<td>32.2</td>
<td>8.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Black (non-Hispanic)</td>
<td>7.4</td>
<td>12.2</td>
<td>33.2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>52.2</td>
<td>70.0</td>
<td>61.2</td>
</tr>
<tr>
<td>Other</td>
<td>8.2</td>
<td>9.1</td>
<td>5.6</td>
</tr>
<tr>
<td>Age:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 30</td>
<td>33.3</td>
<td>36.6</td>
<td>38.3</td>
</tr>
<tr>
<td>30-44</td>
<td>30.4</td>
<td>34.6</td>
<td>30.4</td>
</tr>
<tr>
<td>45-54</td>
<td>13.7</td>
<td>12.9</td>
<td>6.0</td>
</tr>
<tr>
<td>55-64</td>
<td>8.0</td>
<td>6.5</td>
<td>3.1</td>
</tr>
<tr>
<td>65+</td>
<td>14.5</td>
<td>9.4</td>
<td>22.3</td>
</tr>
<tr>
<td>Adults in H.H.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>24.4</td>
<td>26.8</td>
<td>16.6</td>
</tr>
<tr>
<td>2</td>
<td>18.5</td>
<td>18.6</td>
<td>27.9</td>
</tr>
<tr>
<td>3</td>
<td>17.2</td>
<td>17.1</td>
<td>26.5</td>
</tr>
<tr>
<td>4+</td>
<td>39.8</td>
<td>37.4</td>
<td>29.0</td>
</tr>
<tr>
<td>Car loan:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>67.5</td>
<td>73.9</td>
<td>80.2</td>
</tr>
<tr>
<td>Yes</td>
<td>32.5</td>
<td>26.1</td>
<td>19.8</td>
</tr>
<tr>
<td>Credit card:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>7.7</td>
<td>10.5</td>
<td>10.9</td>
</tr>
<tr>
<td>Yes</td>
<td>92.3</td>
<td>89.5</td>
<td>89.1</td>
</tr>
<tr>
<td>Student loan:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>88.0</td>
<td>89.5</td>
<td>88.0</td>
</tr>
<tr>
<td>Yes</td>
<td>12.0</td>
<td>10.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Consumer finance loan:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>71.1</td>
<td>69.3</td>
<td>64.0</td>
</tr>
<tr>
<td>Yes</td>
<td>28.9</td>
<td>30.7</td>
<td>36.0</td>
</tr>
<tr>
<td>Observations</td>
<td>1,632,696</td>
<td>15,486</td>
<td>145,974</td>
</tr>
</tbody>
</table>
For each of these characteristics, we compute for each tract a “z-score” equal to the number of standard deviations that tract’s characteristic lies above (or below) the sample mean. We construct a VA-index for each tract as the z-score for transportation less the z-scores for crime and for poverty. The correlation of this index and actual value-added at the tract level is 0.22 and a scatterplot is shown in figure 12. The thinking behind this policy is that even if child value-added is only noisily targeted, the vouchers steer families currently living in public-housing toward low-crime, low-poverty, high-transportation-accessible neighborhoods.

In both policies the voucher amount for a tract with a VA-Index of 0 is set to $600 per month. In “VA-Index Targeting,” changes in the voucher amount vary linearly with changes in the VA-Index. The slope coefficient is set such that the standard deviation of voucher amounts across all 1,748 tracts is identical to the standard deviation of composition-adjusted (constant-quality) rents across all tracts. This policy induces the same variation in voucher payments as a different policy that does not condition on value-added but instead simply adjusts voucher payments for observed variation in

---

42The poverty rate is taken from the Census, the crime rate is from Peterson and Krivo (2000) and the transportation-access data are from Ramsey and Bell (2013).
Table 8: Results of Counterfactual Policy Experiments

Cumulative impact on math scores (sd=1)

<table>
<thead>
<tr>
<th>Policy</th>
<th>All Housing-Project Types</th>
<th>Black</th>
<th>Hispanic</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Baseline</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>VA-index targeting</td>
<td>0.26</td>
<td>0.26</td>
<td>0.22</td>
<td>0.62</td>
</tr>
<tr>
<td>Aggressive VA-index targeting</td>
<td>0.42</td>
<td>0.27</td>
<td>0.48</td>
<td>0.65</td>
</tr>
<tr>
<td>Direct VA targeting</td>
<td>0.77</td>
<td>1.15</td>
<td>0.55</td>
<td>0.90</td>
</tr>
<tr>
<td>Aggressive direct VA targeting</td>
<td>1.22</td>
<td>1.47</td>
<td>1.05</td>
<td>1.59</td>
</tr>
</tbody>
</table>

rents. In “Aggressive VA-Index Targeting,” we double this coefficient.

**Direct VA Targeting**

In the final two policies, we assume policy-makers can directly observe each tract’s value-added and construct a z-score directly for value-added for each tract. As before, the voucher amount for a tract with a zero value-added z-score is set to $600 per month. In “Direct VA Targeting,” the coefficient relating changes in the voucher amount to changes in the value-added z-score is set such that the standard deviation of voucher amounts across all tracts is identical to the standard deviation of voucher amounts in the VA-Index Targeting policy. In “Aggressive VA Targeting,” we double this coefficient.

Table 8 reports the cumulative impact over 18 years, relative to the baseline, of each of the four policies we consider on child value-added. The first column reports the average impact across all types we simulate and the other three columns report the impact of simulated types sorted by race. The overall results are disproportionately reflective of the impact to Hispanic types because they account for more than 60% of our simulated population.

As table 8 clearly shows, the policies vary dramatically in their effectiveness at altering child value-added. The most effective policy, Aggressive Direct VA-Targeting, improves average child outcomes by nearly five times the least-effective policy, VA-index targeting. As the VA-Index Targeting experiment shows, a noisy or mildly increasing subsidy to high value-added neighborhoods is not sufficient to induce a large percentage of black- or Hispanic-type households to move. It is only when relatively large vouchers target higher value-added tracts

---

43To be clear, the columns report results for types where types are sorted based on the racial plurality of the Census block of their first residence in the sample.
that both black- and Hispanic-type households move to high-value-added neighborhoods in larger percentages. When this happens, policy shows sizable effects on child outcomes. The Aggressive Direct VA-Targeting policy has the potential to eliminate the black-white achievement gap which is equal to one standard deviation in test scores.

Given that we find that policies that aggressively target high value-added tracts are effective at improving child outcomes, we ask the question, “If neighborhoods can be perfectly targeted by policy-makers, what is the voucher amount that maximizes social surplus?” We tackle this question by making the restriction that vouchers may only be used in the top 5% of tracts by value-added. As with the previous counterfactual experiments, households receive a voucher only in the years in which they live in a targeted Census tract.

Figure 13 shows the costs and benefits of this voucher policy on an annual per-household basis, for various monthly voucher levels. For example, the black line shows that a monthly voucher of $700 is associated with an annual per-household cost of $3,830. From this, we can infer the take-up rate of the voucher among the eligible set of households is 46% = $3,830/(12 × $700). At this monthly voucher amount, the per-household benefit of one year of exposure to the targeted Census tracts is also $3,830. Given a take-up rate of 46%, the benefit to those accepting the voucher is $8,400, exactly the annual cost of vouchers paid to those accepting a voucher. We compute this benefit as the total impact of one year of exposure of a household to a targeted neighborhood, assuming 2.5 children in the household, to the net present value of the earnings of the children for 40 years after discounting the computed net present value by 8 years. Obviously, there may be benefits to children and households from living in these tracts in addition to the impact of value-added on child future wages, and if so policymakers should adjust calculations accordingly.

As figure 13 shows, $700/month is the maximum monthly voucher amount that can be offered where the total societal benefits of the voucher are at least as large as the total costs.

44For the 13 tracts in our sample with a poverty rate above 10% containing large (250 or more units) public housing developments, we regress the average number of children per-household (conditional on having a child) on a constant and the share of households in the Census tract that are renters. The constant is 2.78 and the coefficient on share renter is -0.32. This gives us an estimate of 2.46 children per renter household, conditional on having a child and on living in one of these 13 Census tracts. In our analysis, we round this to 2.5.

45Explaining, suppose that one year of exposure in a targeted Census tract increases the Woodcock-Johnson test score by 0.0723 standard deviations for each child. Each standard deviation improvement in the Woodcock-Johnson is assumed to increase adult earnings by $4,000 per year (Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan, 2011), implying the improvement to annual earnings from living in the targeted Census tract is $289 per-year per-child. The net present value of 40 years of $289 per-year improvement discounted at 5% is $4,964; discounting this present value again by 8 years at 5% yields $3,360 for each child. Our calculations from Census data suggest that households living in public housing with children have 2.5 children in the household, on average. Assuming 2.5 children per household gives a total benefit of $8,400.
This is not the surplus-maximizing voucher. At the surplus-maximizing voucher amount, the change in total cost of the program for a marginal change in the voucher is exactly equal to the change in total benefit. If we denote the annual voucher amount as $V$, the take-up rate (participation rate) of the voucher as $P(V)$, and the benefit to the future expected wages of children living a household from living in a targeted Census tract for one year as $B$, then the overall benefits of the voucher are \(^{46}\)

$$P(V)B - P(V)V$$

and assuming an interior solution the optimal voucher amount $V^*$ satisfies

$$V^* = B - \frac{P(V^*)}{P'(V^*)} \quad (16)$$

Under the assumption that the participation rate is never declining in the voucher amount, which should be satisfied given that we find $\alpha_\tau > 0$ for every type, then the surplus-maximizing voucher is always less than the benefit.

\(^{46}\)Note that these benefits exclude the monetary benefits to households from receiving the voucher of $\alpha_\tau V$. Recall, in this section we assume policy-makers distribute vouchers to improve child-outcomes.
Column (1) of table 9 shows our estimate of the surplus-maximizing monthly voucher amount for the entire population in our simulations, $300/month or $V = $3,600/year. At this voucher amount, the take-up rate is $P = 28\%$, shown in column (2). Column (3) reports our estimate of $P (B - V)$, from which the benefit of one year of exposure on the net present value of children’s wages can be computed as $B = $7,760. The implied value of $P'$ evaluated at $V^* = 6.61E-5$, implying a $151 per-year $(12.61 per-month) increase in the voucher amount increases the take-up rate of the voucher by one percentage point. Columns (4) and (5) of table 9 show the monthly voucher amount ($700) and take-up rate (46%) of the break-even voucher, results we discussed earlier.

In the other rows of the table, we compute the surplus-maximizing and break-even voucher amounts by racial types of households. Ignoring all considerations of equity, this table illustrates potential efficiency gains from tailoring public policy by type of household. The table shows there is significant variation in surplus-maximizing and break-even voucher amounts and take-up rates. Generalizing from this table, Hispanic-type households are much less likely to accept a voucher of any amount than black- and other-type households. For example, at a voucher of $200 per month, nearly 50\% of black types accept the voucher but at $500 per month, only 22\% of Hispanic types accept the voucher. This table suggests policymakers can offer relatively modest vouchers to black- and other-type households and expect to see significant participation and benefits in this program; whereas policymakers might need to consider a broader neighborhood choice set or perhaps a different program altogether (such as the “Aggressive direct VA targeting experiment” of table 8) to induce a majority of Hispanic-type households to accept vouchers to move out of public housing and into higher value-added neighborhoods.

6 Conclusion

In this paper, we use two new rich data sets to understand how households choose neighborhoods and the impact of neighborhoods on child ability. We find considerable heterogeneity of the population in the utility of different neighborhoods and we show meaningful variation in the impact of neighborhoods on child ability as measured by test scores. We also show that the utility of households residing in high-poverty neighborhoods, on-average, is much more sensitive to rental prices than the utility of households residing in low-poverty neighborhoods. This last finding helps explain the overall lack of improvement of child

\footnote{47 We compute this using grid search, searching over $50 increments per month.}

\footnote{48 We only consider 13 types of households in this experiment; dividing the types by race seemed a natural way to illustrate heterogeneity in the population.}
Table 9: Surplus-Maximizing and Break-Even Voucher Amounts, Targeted Vouchers

<table>
<thead>
<tr>
<th>Housing Types</th>
<th>Surplus-Maximizing Voucher</th>
<th>Break-Even Voucher+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monthly Voucher Amount</td>
<td>Steady-state Take-up (%)</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>All Public</td>
<td>$300</td>
<td>28%</td>
</tr>
<tr>
<td>Subgroups:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black:</td>
<td>$200</td>
<td>47%</td>
</tr>
<tr>
<td>Hispanic:</td>
<td>$400</td>
<td>18%</td>
</tr>
<tr>
<td>Other:</td>
<td>$500</td>
<td>52%</td>
</tr>
</tbody>
</table>

* Computed as the voucher take-up rate times the difference of the net present value of the impact on lifetime adult earnings from one year of exposure to the targeted neighborhoods and the cost of one year of vouchers paid to move to those neighborhoods. We assume households receiving a voucher have an average of 2.5 children.
+ The net benefit is zero in the break-even voucher scenario.

cognitive ability in the MTO experiment. Counterfactual simulations of our model of neighborhood choice strongly suggest that policy-makers can significantly affect child outcomes as long as housing vouchers directly target high-value-added neighborhoods. When housing vouchers are designed to directly target these neighborhoods, our estimate of the surplus-maximizing and break-even voucher amounts are $300 and $700 per month, respectively. Our analysis assumes that rents and housing supply remain constant after the vouchers are introduced. We think a promising avenue for future research will be to study the general equilibrium effects arising from the large-scale adoption of any of these voucher programs.

References


