Neighborhood Choices and Neighborhood Effects*

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 Wisconsin - Madison
ttan8@wisc.edu

July 26, 2016

Abstract

We investigate how households choose where to live and how neighborhoods affect the ability of children. We use detailed panel data to estimate a dynamic model of neighborhood choice at the Census tract level for renters in Los Angeles county. We then use different panel data for Los Angeles to estimate tract-level "neighborhood effects," defined as the impact of neighborhoods on child cognitive ability. We conclude by simulating a Moving-to-Opportunity type experiment with our model, in which people residing in high poverty neighborhoods are given a rental voucher to move to a low-poverty Census tract. Child ability does not improve in these simulations, as households receiving vouchers tend to move to the least expensive eligible neighborhoods with the lowest neighborhood effects. If these households had chosen a Census tract randomly among the eligible set, child ability would have improved significantly.

JEL Classification Numbers: TBD
Keywords: Neighborhood Choice, Neighborhood Effects, Moving to Opportunity

*We thank Dionissi Aliprantis, Edward Coulson, Steven Durlauf, Edward Glaeser, John Kennan, Amy Ellen Schwartz, Chris Taber, Jim Walker 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 and how neighborhoods affect the cognitive ability of children. These topics have been studied individually before, but our approach is different and our data are new. We show that neighborhoods vary in their impact on child ability, and parents differ in willingness to pay to move to neighborhoods that, on average, significantly improve child ability. We estimate that optimal neighborhood choices are most sensitive to rental prices for the poorest households in our sample and this is key to understanding some of the empirical results of the Moving-to-Opportunity experiment.

Our paper has three main sections, and the first two reflect contributions to heretofore 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 minimizes the role of unobservable shocks in explaining differences in optimal location choices. 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) and Durlauf (2004) for recent surveys. While these studies typically attempt to account for selection issues,[1] the fact that individuals endogenously sort into neighborhoods

[1]For example, Cutler and Glaeser (1997) study the impact of segregation on outcomes of African-

leaves open the possibility of non-causal explanations for these patterns.[2]

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 substantially richer 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 and verbal 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 neighborhoods 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 10 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 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 third section, we reconcile the apparently contradictory conclusions of the neighborhood effects literature and the studying the impact of the Moving-to-Opportunity (MTO) experiment. The Moving to Opportunity experiment 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 there-

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.

after, (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 are set such that after applying the voucher, households spend no more than 30% of their income on rent. Summarizing the medium to long term impacts of MTO, Sanbonmatsu, Kling, Duncan, and Brooks-Gunn (2006), Kling, Liebman, and Katz (2007) and Ludwig, Duncan, Gennetian, Katz, Kessler, Kling, and Sanbonmatsu (2013) 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.[3]

Many view the results from MTO as evidence against the hypothesis that neighborhoods can have large effects on a child's development of skills and educational attainment. Our view is that the results are open to multiple interpretations. One possibility is that, indeed, the findings of large neighborhood effects from earlier observational studies are driven entirely by selection of families into neighborhoods, and that true but unobserved neighborhood effects are small or non-existant. A second interpretation is that families receiving a voucher in the MTO experiment chose their neighborhoods after considering how that voucher changed the relative price of eligible neighborhoods. Thus, depending on how neighborhoods were chosen by MTO participants after receiving a voucher, the intent-to-treat effect of the MTO subsidy offer may have differed substantially from the average treatment effect of lower poverty neighborhoods on outcomes: See Aliprantis and Richter (2016), Clampet-Lundquist and Massey (2008) and Pinto (2014) for related discussions.

In the spirit of Davis and Foster (2005) and Todd and Wolpin (2006), we run counterfactual simulations of our decision model to better understand why the MTO experiment did not improve child outcomes if neighborhood effects are in fact important.[4] To start, we replicate the environment created by the MTO experiment. We refer to this simulation as "MTO-A." To implement MTO-A, we require an estimate, by type, of how the utility of each neighborhood in Los Angeles would change given the change in rental prices induced by the voucher. We extract an estimate of the rental-price sensitivity of each of our types of households using the instrumental variables approach of Bayer, Ferreira, and McMillan

[3]Recent work by Chetty, Hendren, and Katz (2015) argues that MTO positively affected adult wages for children who were young at the time their family received a voucher, consistent with other research suggesting that housing vouchers positively affect female adult wages Andersson, Haltiwanger, Kutzbach, Palloni, Pollakowski, and Weinberg (2016). Given the findings from Sanbonmatsu, Kling, Duncan, and Brooks-Gunn (2006) that MTO did not cognitively affect the cognitive ability of this group, presumably the effects on earnings operate through a different channel.

[4]Galiani, Murphy, and Pantano (2015) estimate a structural model of location choice directly using MTO data, and run counterfactual experiments with their estimated model, but do not study the impact of MTO on child well-being.

(2007).

In MTO-A, we compute the optimal neighborhood choices of households that begin the sample in neighborhoods with public housing developments but are offered a rental voucher valid for use in low-poverty-rate neighborhoods in the first year and can live anywhere and continue to receive a voucher in subsequent years. Importantly, we show that households that use the voucher move to low-poverty neighborhoods with the lowest neighborhood value-added, on average. Since the chosen neighborhoods have low value-added, child test scores do not improve. So, why does the model predict this outcome? We demonstrate that the types of households likely to receive an MTO voucher are very sensitive to rental prices; and the highest-value-added neighborhoods in low-poverty tracts are, on average, the most expensive. To prove that selection on rental prices is key, we perform another simulation we call "MTO-B" in which we randomly assign MTO-eligible households to neighborhoods with similar poverty rates as those chosen in the MTO-A. In this simulation, we predict that child test scores significantly improve. In other words, our counterfactual simulations suggest that parents use the MTO voucher to move to low-poverty neighborhoods, but they choose relatively cheap and low value-added neighborhoods and child outcomes do not improve.

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 ℓ given a current location of j and current value of a shock ϵ_ℓ (to be explained later) as

$$V(\ell \mid j, \epsilon_\ell) \;=\; u(\ell \mid j, \epsilon_\ell) + \beta EV(\ell) \tag{1}$$

In the above equation $EV(\ell)$ is the expected future value of having chosen to live in ℓ today.

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 ℓ given a current location of j and a value for ϵ_ℓ. We assume u is the simple function

$$u(\ell \mid j, \epsilon_\ell) \;=\; \delta_\ell \;-\; \kappa \cdot 1_{\ell \neq j} \;+\; \epsilon_\ell \tag{2}$$

δ_ℓ is the flow utility the household receives this period from living in neighborhood ℓ, net of rents and other costs; κ is the sum of all costs (utility and financial) a household must pay when it moves to a different neighborhood i.e. when $\ell \neq j$; and ϵ_ℓ is a random shock that is known at the time of the location choice. ϵ_ℓ is assumed to be iid across locations, time and people. The parameters δ_ℓ and κ may vary across households, but for any given household δ_ℓ and κ are assumed fixed over time. ϵ_ℓ induces otherwise identical households living at the same location to optimally choose different future locations. Note that δ_ℓ is the type-specific indirect utility of living in neighborhood ℓ, 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 ϵ_1 as the shock associated with location 1, ϵ_2 as the shock with location 2, and so on. In each period after the vector of ϵ are revealed (one for each location), households choose the location that yields the maximal value

$$V(j \mid \epsilon_1, \epsilon_2, \ldots, \epsilon_J) \;=\; \max_{\ell \in 1,\ldots,J} V(\ell \mid j, \epsilon_\ell) \tag{3}$$

$EV(j)$ is the expected value of (3), where the expectation is taken with respect to the vector of ϵ.

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 δ with wages after adjusting for cost of living and allow κ to vary with distance. Bishop and Murphy (2011) and Bayer, McMillan, Murphy, and Timmins (2015) specify δ as a linear function of spatially-varying amenities with the aim of recovering individuals' willingness to pay for those amenities. We allow the δ's to vary flexibly across neighborhoods and across households, 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 ϵ 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

$$EV(j) \quad = \quad \log\left\{\sum_{\ell=1}^{J} \exp \widetilde{V}(\ell \mid j)\right\} + \zeta \tag{4}$$

where ζ is equal to Euler's constant and

$$\widetilde{V}(\ell \mid j) \quad = \quad \delta_\ell \; - \; \kappa \cdot 1_{\ell \neq j} \; + \; \beta EV(\ell) \tag{5}$$

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

$$p(\ell \mid j) \quad = \quad \widetilde{V}(\ell \mid j) \; - \; \log\left\{\sum_{\ell'=1}^{J} \exp\left[\widetilde{V}(\ell' \mid j)\right]\right\} \tag{6}$$

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

$$p(\ell \mid j) \quad = \quad \widetilde{V}(\ell \mid j) - \widetilde{V}(k \mid j) \; - \; \log\left\{\sum_{\ell'=1}^{J} \exp\left[\widetilde{V}(\ell' \mid j) - \widetilde{V}(k \mid j)\right]\right\} \tag{7}$$

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 (7) 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 equation (4). 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 to proceed. This approach does not require that we solve for the value functions. Note that equation (5) implies

$$\widetilde{V}(\ell \mid j) - \widetilde{V}(k \mid j) \quad = \quad \delta_\ell - \delta_k \; - \; \kappa\left[1_{\ell \neq j} - 1_{k \neq j}\right] \; + \; \beta\left[EV(\ell) - EV(k)\right] \tag{8}$$

But from equation (4),

$$EV(\ell) - EV(k) \quad = \quad \log\left\{\sum_{\ell'=1}^{J} \exp \widetilde{V}(\ell' \mid l)\right\} - \log\left\{\sum_{\ell'=1}^{J} \exp \widetilde{V}(\ell' \mid k)\right\} \tag{9}$$

Now note that equation (6) implies

$$p(k \mid \ell) = \widetilde{V}(k \mid \ell) - \log\left\{\sum_{\ell'=1}^{J} \exp\left[\widetilde{V}(\ell' \mid \ell)\right]\right\} \tag{10}$$

$$p(k \mid k) = \widetilde{V}(k \mid k) - \log\left\{\sum_{\ell'=1}^{K} \exp\left[\widetilde{V}(\ell' \mid k)\right]\right\} \tag{11}$$

and thus

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

is equal to

$$\begin{aligned} & \widetilde{V}(k \mid \ell) - \widetilde{V}(k \mid k) - [p(k \mid \ell) - p(k \mid k)] \\ = & -\kappa \cdot 1_{\ell \neq k} - [p(k \mid \ell) - p(k \mid k)] \end{aligned} \tag{12}$$

The last line is quickly derived from equation (5). Therefore,

$$EV(\ell) - EV(k) = -[p(k \mid \ell) - p(k \mid k) + \kappa \cdot 1_{\ell \neq k}] \tag{13}$$

and equation (8) has the expression

$$\begin{aligned} & \widetilde{V}(\ell \mid j) - \widetilde{V}(k \mid j) \\ & = \delta_\ell - \delta_k - \kappa[1_{\ell \neq j} - 1_{k \neq j}] - \beta[p(k \mid \ell) - p(k \mid k) + \kappa \cdot 1_{\ell \neq k}] \end{aligned} \tag{14}$$

Combined, equations (7) and (14) show that the log probabilities that choices are observed are simple functions of model parameters $\delta_1, \ldots, \delta_J$, κ and β 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 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

[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

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^{TM} which provides some information on the financial wherewithal of the household as demonstrated 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.[6] 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

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; however, a SSN is required to receive a housing voucher.

[6] On average, each Census tract in Los Angeles has about 4,000 people.

Table 1: Comparison of Equifax and Census Data

Poverty Rate (%)	Avg. Population 2000-2010 Equifax[a]	Census[b]	Equifax Share	Pct. w/ Mortgage 2008-2012 Equifax[c]	ACS[d]
(1)	(2)	(3)	(4)	(5)	(6)
0-5	610,336	654,004	93.3%	61.6%	62.6%
5-10	1,395,831	1,478,114	94.4%	50.0%	50.2%
10-15	1,033,076	1,135,194	91.0%	40.5%	39.2%
15-20	751,098	870,869	86.2%	37.3%	34.9%
20-25	630,830	761,841	82.8%	30.7%	26.9%
>25	1,085,466	1,497,545	72.5%	23.9%	19.0%
Public Housing[e]	34,988	42,431	82.5%	27.0%	23.9%

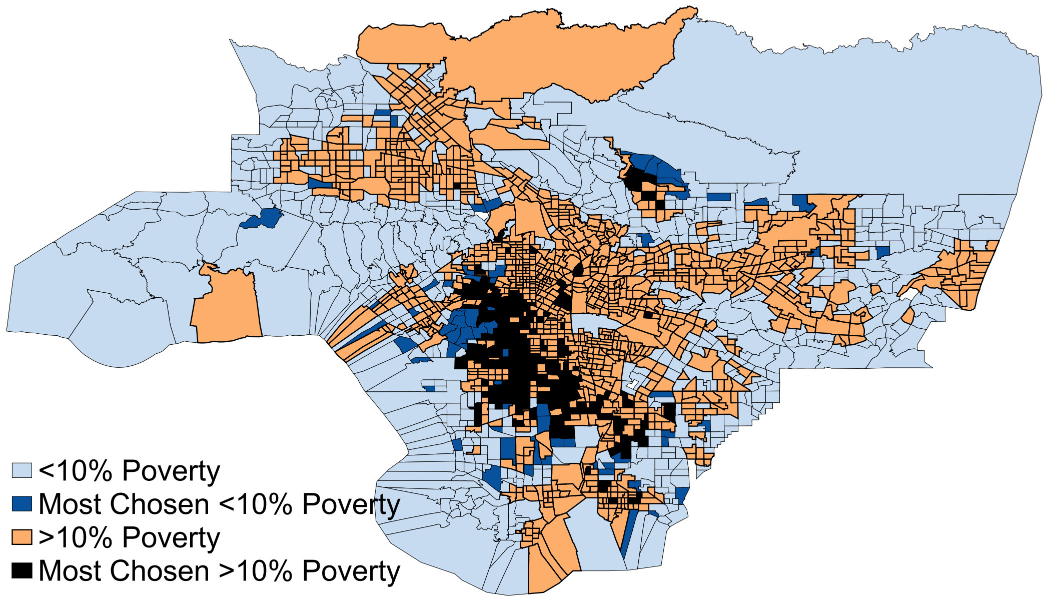
Notes:
a Data are computed as 20 times the average (1999-2014) number of Equifax primary individuals ages 18-64.
b Data shown are the average (2000 and 2010) of the Census tract population ages 18-64.
c Data are the average share of households in Equifax with a mortgage, 2008-2012.
d Data are the average 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+ public housing units and above 10% poverty rate in 2000.

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 quite close.

We stratify households into types using an 8-step stratifying procedure. We begin with the full sample, and subdivide the sample into smaller "cells" based on (in this order): The racial plurality of the 2000 Census block of residence (4 bins),[7] 5 age categories (cutoffs at 30, 45, 55, and 65), number of adults 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.

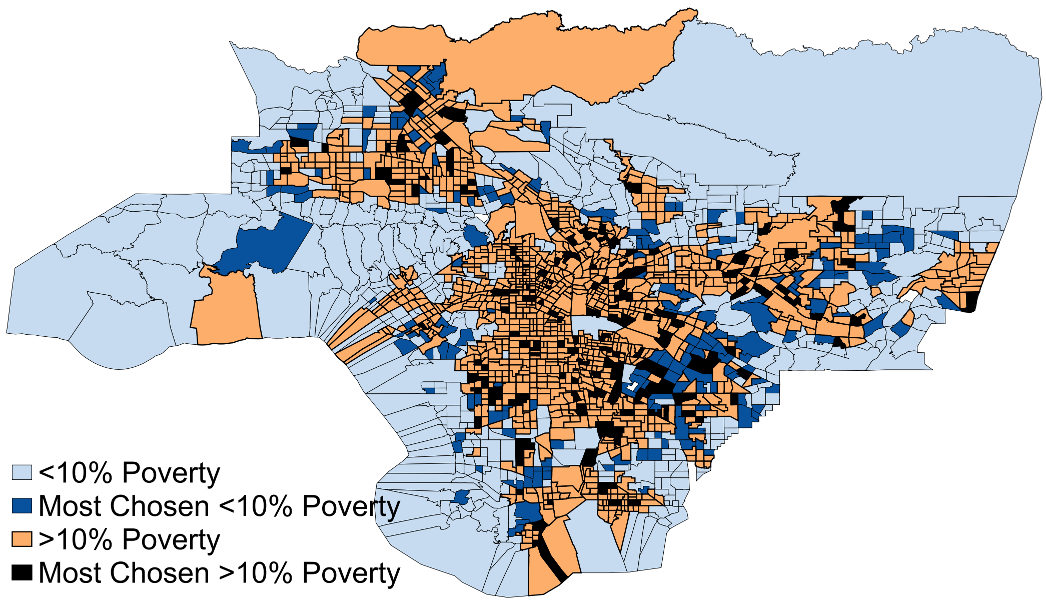
[7] We assign race based on the racial plurality of all persons in the Census block, owners and renters. We expect that the geography of the Census block is small enough that the racial plurality of renters will be identical to that of the entire block. For individuals who enter the sample after 1999, we classify them based on the racial plurality of the block where they are first observed, which in most cases is 1999.

Figure 1: Location Choices of 2-adult black households w/ <580 Equifax Riskscore



The benefit of working with a data set like the FRBNY Consumer Credit Panel / Equifax data is that its sheer size allows estimates of the substitutability of neighborhoods, i.e. the vector of δ_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. A few graphs 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 Riskscore 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 Riskscore 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 δ_j were the same, the model would attribute the systematic variation in optimal neighborhood choices of the two types entirely to differences in the i.i.d. utility shocks experienced.

Figure 2: Location Choices of 2-adult Hispanic households w/ 590-656 Equifax Riskscore



Our sample is comprised of 1,748 Census tracts. Allowing a separate value of δ for each tract and for each type would require estimating more than 250,000 parameters. For parsimony, for each type we specify that the utility of location j, δ_j, is a function of latitude (lat_j) and longitude (lon_j) of that location according to the formula

$$\delta_j = \sum_{k=1}^{K} a_k B_k \left(lat_j, lon_j\right) \tag{15}$$

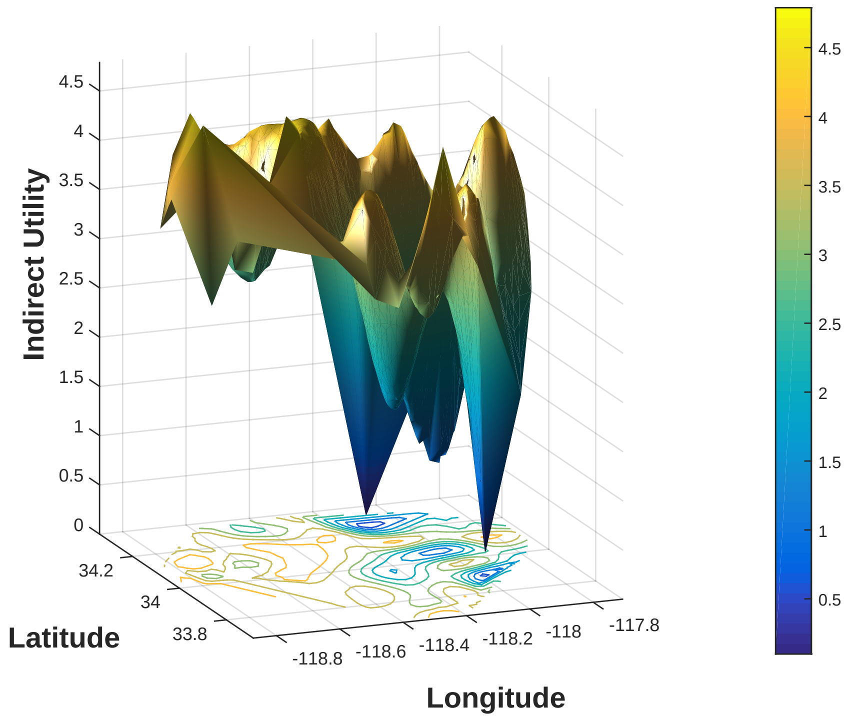
The B_k are parameter-less basis functions. For each type, we use $K = 178$ basis functions. Additionally, we allow the values of a_k to vary for tracts above and below 10% poverty threshold. Inclusive of the moving cost parameter, we estimate $2 \times 178 + 1 = 357$ parameters per type. With 144 types, we estimate a total of $51,408$ 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 ℓ_{it} as person $i's$ observed choice of location in year t. Denote τ as type and the vector of parameters to be estimated for each type as θ_τ. The log likelihood of the sample is

$$\sum_{\tau} \sum_{i \in \tau} \sum_{t} \quad p\left(\ell_{it} \mid j_{it}; \theta_\tau\right) \tag{16}$$

$p\left(.\right)$ is the model predicted log-probability of choosing ℓ_{it} given j_{it}. For each τ we use the

Figure 3: Indirect Utility, Type 133



quasi-Newton BFGS procedure to find the vector θ_τ that maximizes the sample log likelihood.

2.3 Estimates and Model Fit

Our estimation procedure ultimately yields estimates of δ_j and κ for each type to match model-predicted moving probabilities to those in the data. 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. Figure 5 compares actual and model-predicted migration rates in our sample. About 8-1/2 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 shows a more detailed comparison of model-predicted and annual migration

Figure 4: Indirect Utility, Type 20

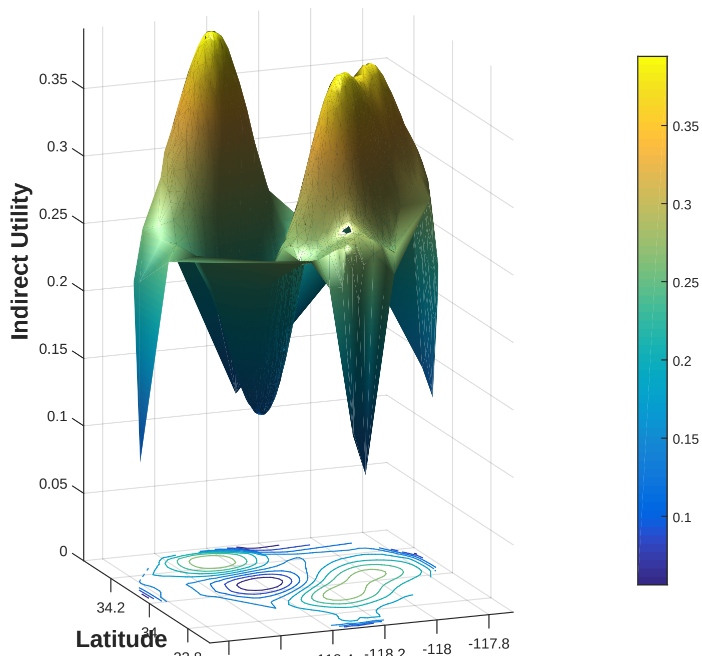
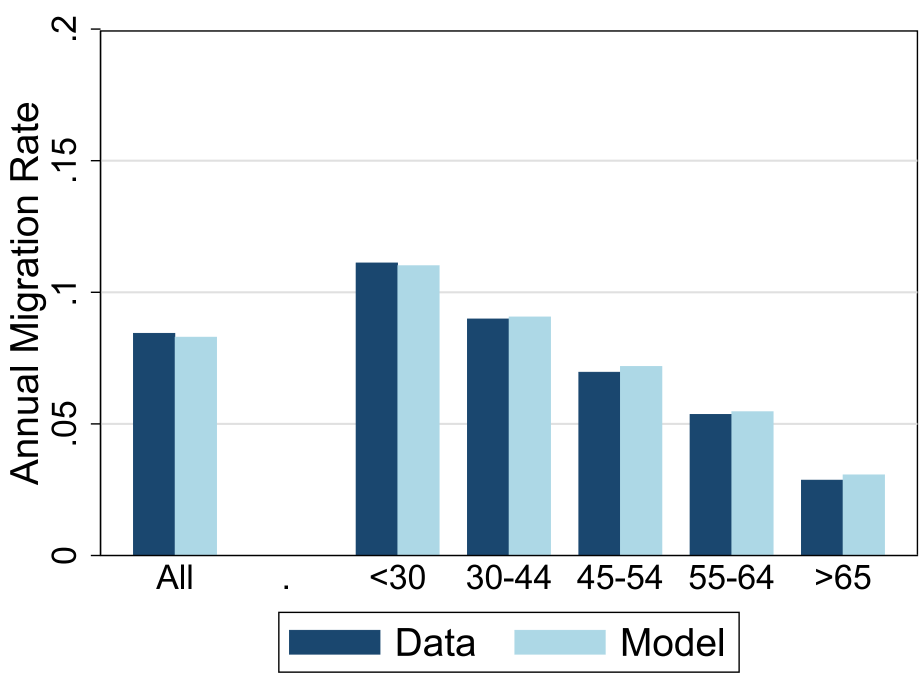


Figure 5: Model Fit: Moving Probabilities



rates for households that choose to move. 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 6 shows model fit for very low-probability moves,[8] and in our view, this figure shows that by and large the model fits the data impressively. Focusing on the negatives, the model tends to over-predict moves involving large changes in poverty rates. The panels in the top row show the model over-predicts moves from low poverty tracts to high poverty tracts and the bottom-right panel shows the model over-predicts moves from the highest to lower poverty tracts.

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 given type τ. To make progress, we specify that $\tilde{\delta}_{j\tau}$ is a linear function of rent, observable characteristics of tract j, $\mathcal{O}_j$, and unobserved characteristics of tract j, ζ_j

$$\tilde{\delta}_{j\tau} \;=\; -\alpha_\tau \cdot rent_j \;+\; \lambda_\tau \cdot \mathcal{O}_j \;+\; \zeta_j \tag{17}$$

The parameter α, 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, ζ_j. An instrument is required. We use a three-step IV approach to estimate α that is common in the IO and Urban literature, for example Bayer, Ferreira, and McMillan (2007).

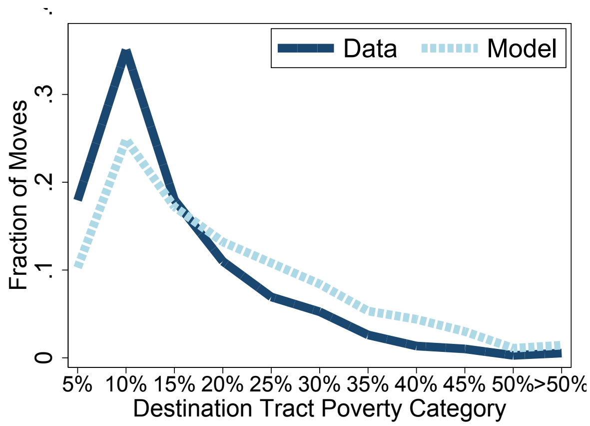
In the first step of our procedure, we include characteristics of the housing stock 0-5 miles from tract j in $\mathcal{O}_j$ and use characteristics of the housing stock 5-20 miles from the tract as instruments. These instruments are assumed to affect equilibrium rent in j but do not directly affect δ_j, the utility in tract j.[9] The F-statistic arising from this first step is 7.

In the second step, we use estimates of α and λ from the first step, call them $\widehat{\alpha}_\tau$ and $\widehat{\lambda}_\tau$,

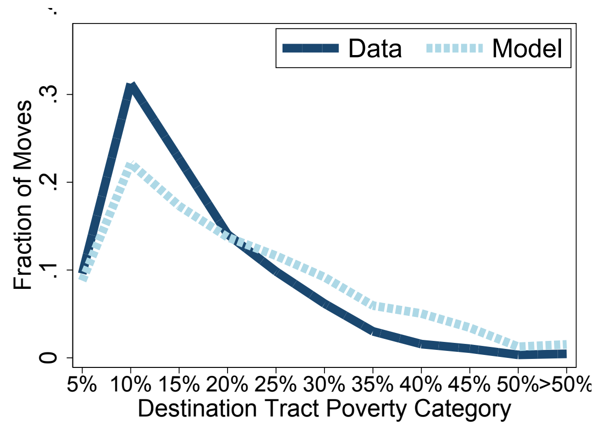
[8]Recall the unconditional probability of any move is less than ten percent.

[9]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 utility.

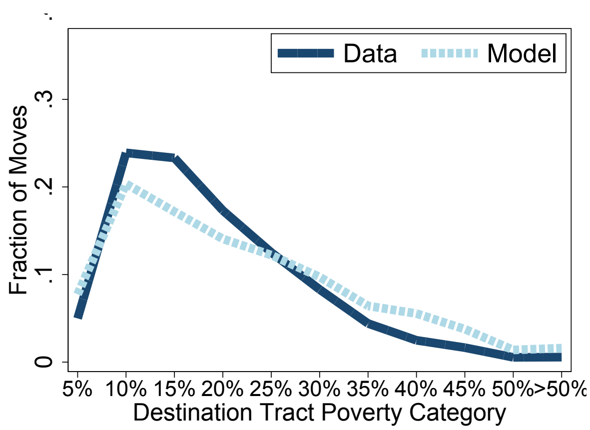
Figure 6: Poverty Category Transitions $t-1$ to t, Conditional on Moving



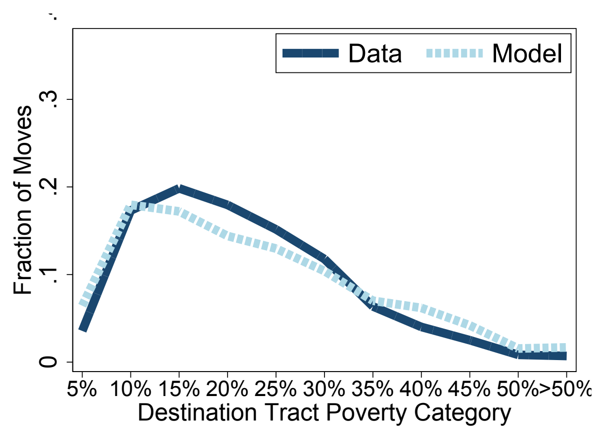
(a) From 0-5%



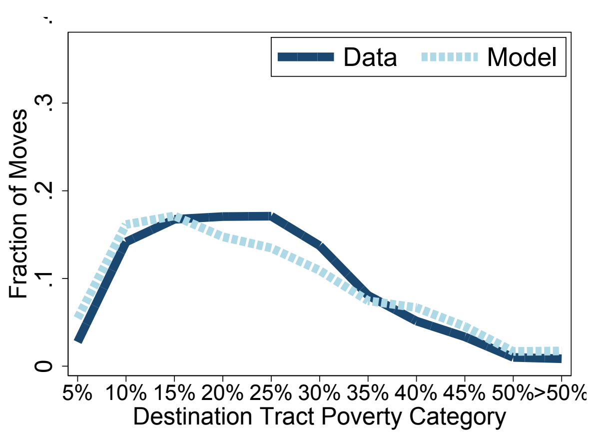
(b) From 5-10%



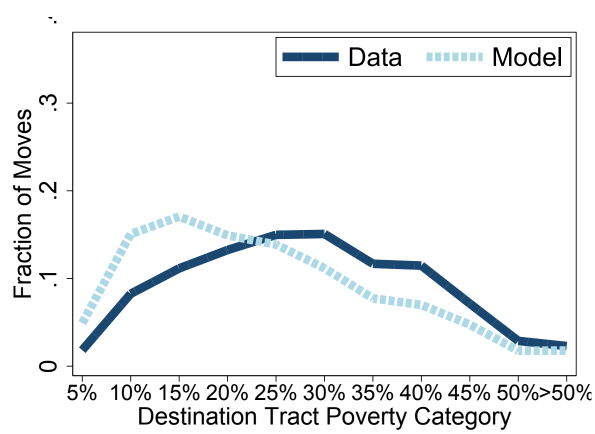
(c) From 10-15%



(d) From 15-20%



(e) From 20-25%



(f) From >25%

to construct a new surface of indirect utilities for each type abstracting from unobservables as

$$\widehat{\delta}_{j\tau} \;=\; -\widehat{\alpha}_{\tau} \cdot rent_j \;+\; \widehat{\lambda}_{\tau} \cdot \mathcal{O}_j \tag{18}$$

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 α by type. We summarize this variation in Figure 7 which graphs the average value of α by initial Census tract of residence for the people in our estimation sample.[10] The figure shows that people living in high poverty tracts are, on average, more than twice as 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) 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 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.

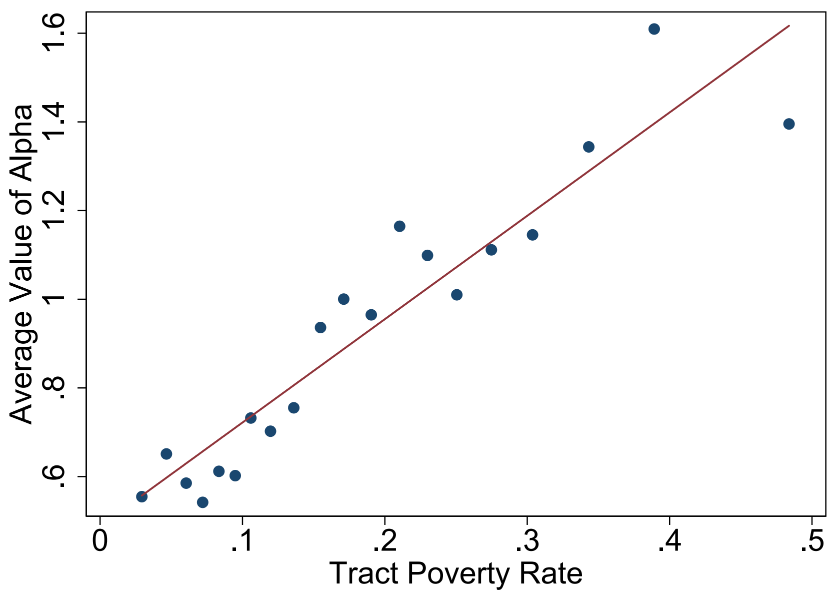
The LA FANS data has the advantage of sampling by Census tract, so that we observe many households within a small geographic region.[11] The LA FANS oversamples poor neigh-

[10] The average value of α varies by Census tract because the mix of types varies by tract.

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

Figure 7: Average Estimates of α by Tract Poverty Rate

$$\alpha \approx \frac{\partial \ln(\text{Choice Prob}_j)}{\partial(\text{rent}_j)/1000}$$

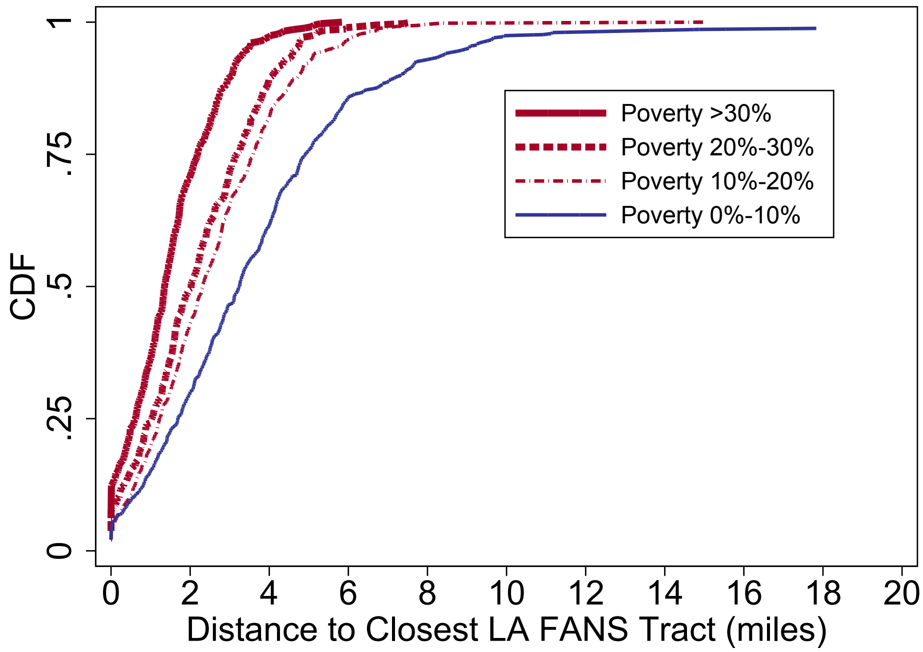


borhoods, but the 65 Census tracts are distributed across much of Los Angeles. Figure 8 shows the distance of each tract in our Los Angeles sample from the previous section to a tract in the LA FANS sample. Most tracts in Los Angeles are located within a few miles on LA FANS tract, but on average high-poverty Census tracts are closer to an LA FANS tract, reflective of 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 described in Schrank, McGrew, and Woodcock (2001) for applied problems ("math") and passage comprehension ("reading"), tests used in many MTO studies. 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.[12] This reduces our

[12]Children that change locations between waves are assigned to the Census tract of their location in the first wave. We include all children, including those that change locations, in our estimation sample.

Figure 8: Distance to Closest LA FANS Tract by Poverty Rate of Tract



sample to 1,253 for our math skill measure and 1,267 for our reading skill measure, 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.[13]

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 production of the change in child ability, $\Delta_{t-T}A_{i,j,t}$, between periods $t-T$ and t ,

$$\Delta_{t-T}A_{i,j,t} = Z'_{i,j,t-T}\psi + v_{i,j,t} \quad ; \qquad v_{i,j,t} = T\left[\mu_j + \epsilon_{i,j,t}\right] \ , \tag{19}$$

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 $t-T$, μ_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.[14]

[13]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 see their test scores once.

[14]We include the T term when defining $v_{i,j,t}$ so that μ_j and $\epsilon_{i,j,t}$ are annualized.

Table 2: Descriptive Statistics, LA FANS

	Mean	S.D.	Obs.
Dependent Variables			
Change in math score	-0.009	1.034	1253
Change in reading score	0.001	1.051	1267
Control Variables (LA FANS Wave 1)			
Wave 1 Test Scores			
Math score	0.000	1.000	1271
Reading score	0.000	1.000	1271
Child Demographics			
Age of child (years)	7.783	4.827	1271
Hispanic (1=yes)	0.663	0.473	1271
Black (1=yes)	0.094	0.291	1271
Male (1=yes)	0.515	0.500	1271
Parental Demographics and Education			
Number of kids	2.427	1.176	1271
Parental IQ	86.845	15.624	1271
High school dropout	0.354	0.478	1270
High school graduate	0.169	0.375	1270
Some college	0.268	0.443	1270
Bachelor degree	0.100	0.300	1270
Graduate degree	0.063	0.243	1270
Parental Income and Assets ($000s)			
Log income	2.511	3.117	1263
Log assets	2.052	2.015	1263

Notice that in the absence of any control variables, μ_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, income and assets, all measured as of Wave 1. We present descriptive statistics of our key dependent and independent variables 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

Table 3: $R2$ of Models of Neighborhood Effects

$R2$ Values from LA FANS data
65 tracts, 1,253 observations for math and 1,217 for reading

Model	Math	Reading
1: Neighborhood Fixed Effects Only	0.09	0.10
2: + Splines in Lagged Child Scores	0.41	0.40
3: + Splines interacted w/ Child Controls	0.51	0.49
4: + Parent Ability and Demographics	0.52	0.51
5: + Lagged Income and Assets	0.52	0.52

the presence of neighborhood fixed effects in equation (19), ψ 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 ψ.[15] For an unbiased estimate of μ_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 (19), showing model fit across a number of specifications. The outcome variable is the change in the relevant standardized test score between LA FANS waves. When tract-level fixed effects are the only regressors, model 1, the $R2$ of the regression is just 10%. Once information on lagged child test scores is included as a regressor (model 2) the $R2$ jumps to 40%. Adding child controls (model 3) and parent demographics (model 4) increases the $R2$ to 50%. Adding information on parental income and assets (model 5) fails to further boost $R2$ values. Given the $R2$ 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 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 (19) to account

[15] 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 otherwise desirable observable characteristics than other households.

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 (19) using the LA FANS data. Define $\hat{\mu}_k$ as the estimate of tract-k's annual fixed effect, $\widehat{\sigma}^2_\mu$ as the estimated variance of the tract-level fixed effects and $\widehat{\sigma}^2_\epsilon$ 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

$$\omega_{j,k} \quad = \quad N_k \times \phi\left(\frac{\|j-k\|_{distance}}{h_1}\right) \times \phi\left(\frac{\|j-k\|_{attributes}}{h_2}\right) \tag{20}$$

where h_1 and h_2 are bandwidths and $\phi\left(.\right)$ is the standard Normal density function. The term $\|j-k\|_{distance}$ is the physical distance (in miles) between the centroids of 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.[16][17] We compute annual value added for tract j as

$$\underbrace{\left(\frac{\sum\limits_k \omega_{j,k}\ \hat{\mu}_k}{\sum\limits_{k'} \omega_{j,k'}}\right)}_{\text{Interpolation}} \ \underbrace{\left(\frac{\widehat{\sigma}^2_\mu}{\widehat{\sigma}^2_\mu + \widehat{\sigma}^2_\epsilon / \widetilde{N}_j}\right)}_{\text{Shrinkage}} \tag{21}$$

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

$$\frac{\left(\sum\limits_k \omega_{j,k}\right)^2}{\sum\limits_{k'}\left(\omega^2_{j,k'}/N_{k'}\right)} \tag{22}$$

The interpolation term in equation (21) is straightforward, as it is a simple weighted average. To understand the shrinkage term and why it is standard in the teacher value-

[16]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.

[17]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

	Math Value Added				
Model	1	2	3	4	5
1	1.00				
2	0.75	1.00			
3	0.68	0.90	1.00		
4	0.52	0.80	0.94	1.00	
5	0.50	0.79	0.91	0.99	1.00
Ann. Std. Dev.	0.045	0.039	0.040	0.037	0.037

	Reading Value Added				
Model	1	2	3	4	5
1	1.00				
2	0.78	1.00			
3	0.76	0.96	1.00		
4	0.75	0.92	0.96	1.00	
5	0.75	0.90	0.95	0.99	1.00
Ann. Std. Dev.	0.061	0.059	0.056	0.053	0.055

All Census tracts (1,748) after interpolation and shrinkage has occurred

added literature, consider a simplified model where Δa is the change in the next child's test score, μ is the true neighborhood effect and ϵ is a child-specific shock. Suppose that a noisy estimate of μ, call it μ^o, is observed

$$\begin{array}{lrcl} \text{Truth:} & \Delta a & = & 1 \cdot \mu + \epsilon \\ \text{Observed:} & \mu^o & = & \mu + \nu \end{array} \tag{23}$$

with ν being measurement error. A regression of Δa on μ^o will yield a biased coefficient of $\sigma_\mu^2 / \left(\sigma_\mu^2 + \sigma_\nu^2\right)$. Dividing estimates of μ^o by this expression will produce an unbiased regression coefficient of 1. In mapping the intuition of equation (23) to what we actually do, note that the variance of ν – 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 ν. This explains the presence of the $\widetilde{N}_j$ term in equation (21). 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 (22).

Table 4 shows tract-level correlations of value-added estimates for the five different models discussed in table 3 after interpolation and shrinkage have occurred for all 1,748 Census tracts in our study.[18] The results for math are shown in the top half and reading in the bottom half. 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 for math and 0.90 and above for reading. 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 for math and 0.053 for reading. 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 between 37% and 50% 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.[19] 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.[20] 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 level poverty rates and racial percentages, as the $R2$ of the regressions are less than 17%.

Upon further review, a case can be made that our estimates of tract value-added are capturing something very different from available estimates of public-school quality. Figure 9 plots, for mathematics, 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.[21] There is considerable variation around the tract-level averages shown in figure 9 (not shown), but on average our estimates of value-added decline with tract poverty rates and the Los Angeles Times estimates of school quality

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

[19]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.

[20]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.

[21]The same graphs for reading look very similar (not shown).

Table 5: Neighborhood Traits and Value Added

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

Variable	Math	Reading
Math School VA+	0.025	-0.067
	(0.045)	(0.056)
English School VA+	0.064	0.218**
	(0.071)	(0.096)
Poverty Rate	-0.003	0.073
	(0.051)	(0.053)
Pct. Hispanic	-0.063***	-0.097**
	(0.024)	(0.026)
Pct. Black	-0.017	0.017
	(0.025)	(0.025)
Pct. Hispanic x Poverty Rate	0.046	0.195**
	(0.066)	(0.074)
Pct. Black x Poverty Rate	0.069	-0.246**
	(0.106)	(0.120)
Constant	0.032	0.022
	(0.010)	(0.010)
R2	0.141	0.168

+ LA Times Measure of Local Public Elementary School Value Add
** Significant at a 5% confidence level
*** Significant at a 1% confidence level

increase.

4 Reconciling Large Neighborhood Effects with MTO

Our finding of large neighborhood effects is squarely in line with an earlier literature that estimates these effects: See Leventhal and Brooks-Gunn (2000) and Durlauf (2004) 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.[22]

Recognizing the limitations of observational studies, the literature on neighborhood effects has devoted considerable attention recently to the "Moving to Opportunity" randomized experimental intervention. Moving to Opportunity was a randomized control trial begining 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 are set such that after applying the voucher, households spend no more than 30% of their income on rent.[23] 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.[24]

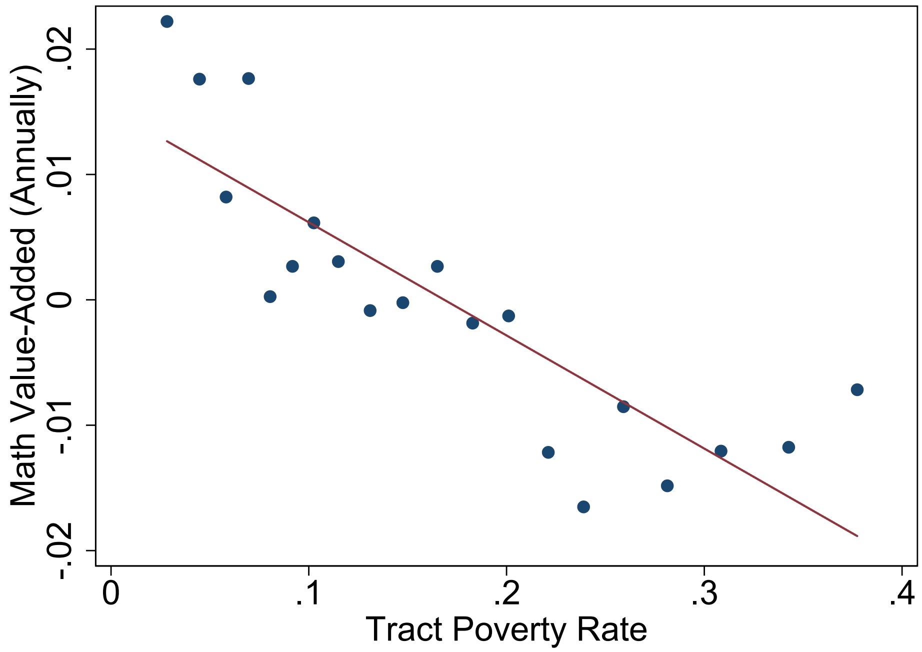
But do the MTO results prove that all neighborhood effects are small? Perhaps not. Suppose there is variation in neighborhood value-added in tracts with a poverty rate under 10%; and, suppose that rents are higher for tracts with greater value-added. Once households receive a voucher to live in a tract with pverty rate under 10%, they must decide whether to move to a high-rent, high-value-added tract or a low-rent, low-value-added tract. Figure 10

[22] 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. See Aaronson (1998) for examples of instruments used by other researchers in this field and their potential limitations.

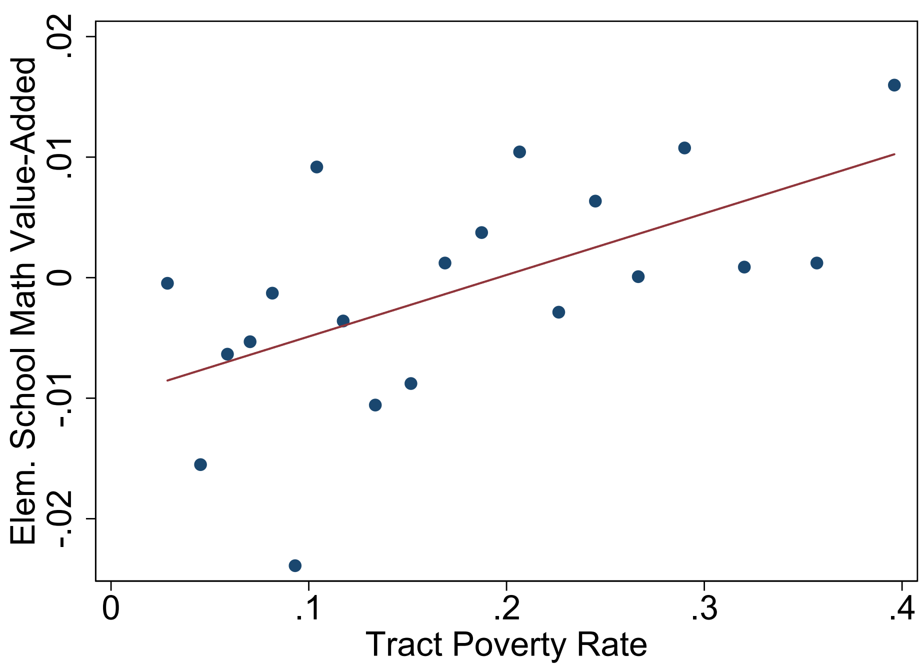
[23] Households that wanted to rent a more expensive unit could only contribute up to an additional 10% of their income.

[24] Recent work by Chetty, Hendren, and Katz (2015) demosntrates that MTO positively affected adult wages.

Figure 9: Tract Poverty, Value Added and School Quality (Math)



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



(b) Avg. School Quality (Math) against Poverty Rates gives a stylized graphical illustration of the range of possible outcomes after an MTO-style intervention. In both panels, the x-axis represents neighborhood value-added; the y-axis represents housing rent; the solid black line shows the set of available combinations for the high-poverty neighborhoods; the dashed line shows the set of available combinations for the low-poverty neighborhoods; and the red lines show indifference curves.[25] The top panel shows one possible outcome from MTO: As households move from high-poverty to low-poverty tracts via the MTO rent subsidy, their rent falls and their child value-add rises. The bottom panel shows a case where child value-added falls after the MTO rent subsidy. The ultimate change in child outcomes after the rent subsidy is received depends on ideas from classic microeconomics: Changes to the slope of the budget line, and income and substitution effects.

Our data suggests relative prices and income and substitution effects may be of first order importance. 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).[26] The top panel shows the relationship for math value added and the bottom panel shows reading. 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, households may be willing to pay to live in those neighborhoods conditional on receiving a large enough rent subsidy.

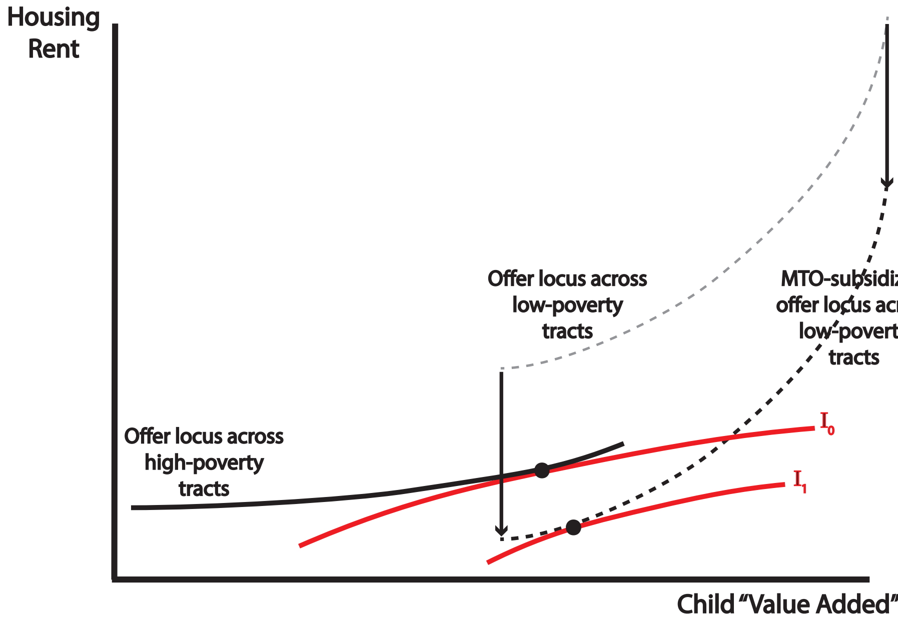
Figures 7 and 11 foreshadow our results. Figure 7 suggests the types of people currently living in high poverty tract areas are quite sensitive to the level of rent; and, Figure 11 suggests the price of acquiring additional child value added is much higher in low-poverty-rate tracts than in high-poverty tracts. It seems quite possible that child outcomes may not change or perhaps worsen if we subsidize families to move from high poverty neighborhoods to low poverty neighborhoods without further restricting which low poverty neighborhood they move to. If this were indeed the case, it would reconcile the apparent contradiction of large neighborhood effects in the observational literature and small experimental results of MTO.

For a quantitative analysis, we simulate optimal decisions of our estimated location-choice

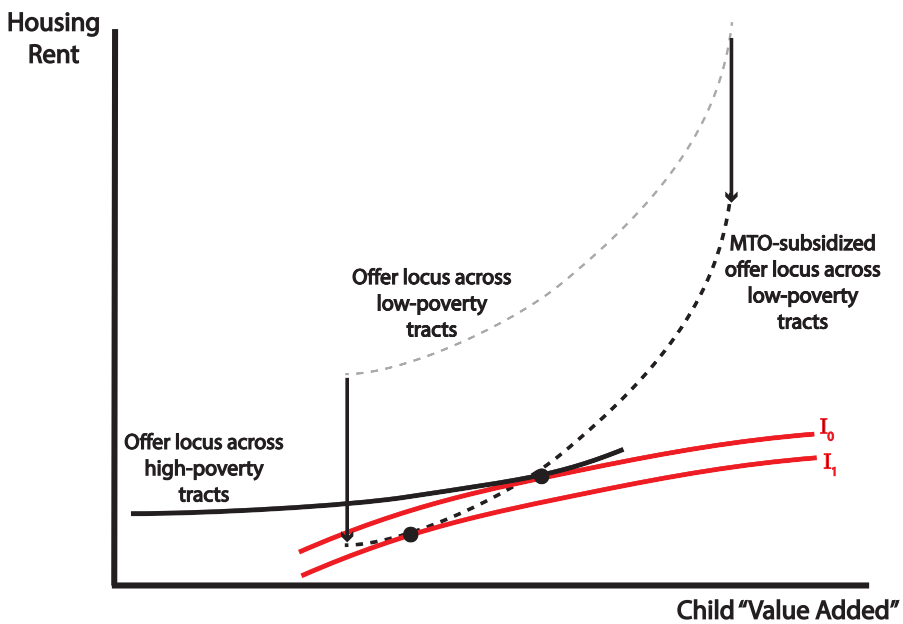
[25]Households dislike housing rent and like value added, so households are best off in the south-east corner of the graph.

[26]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.

Figure 10: MTO's Predicted Effect on Child Value-Added when the Hedonic Rent/Value-added gradient is steeper in low-poverty areas than high poverty areas

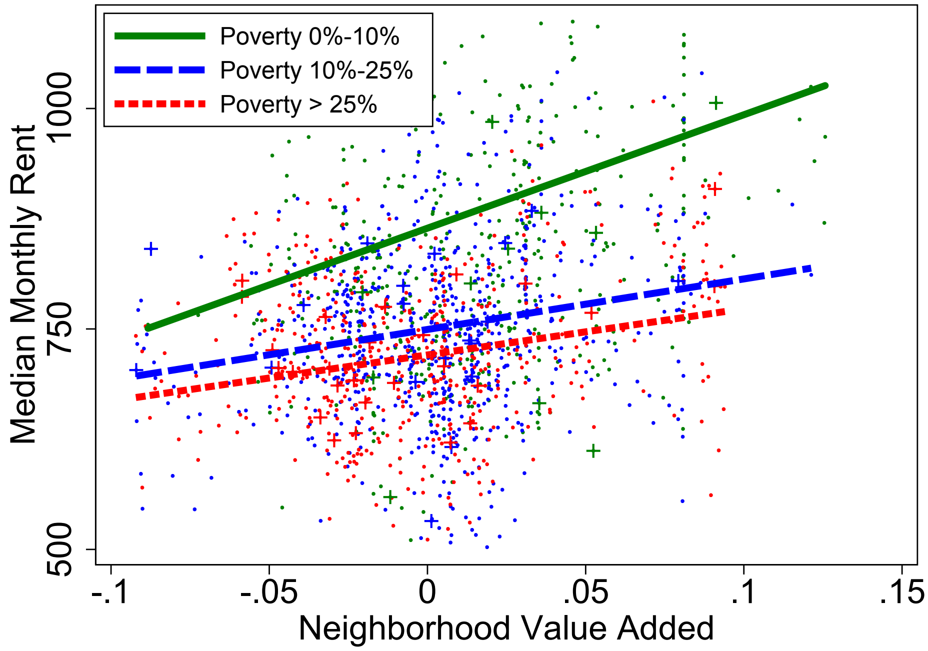


(a) Scenario where MTO increases child outcomes

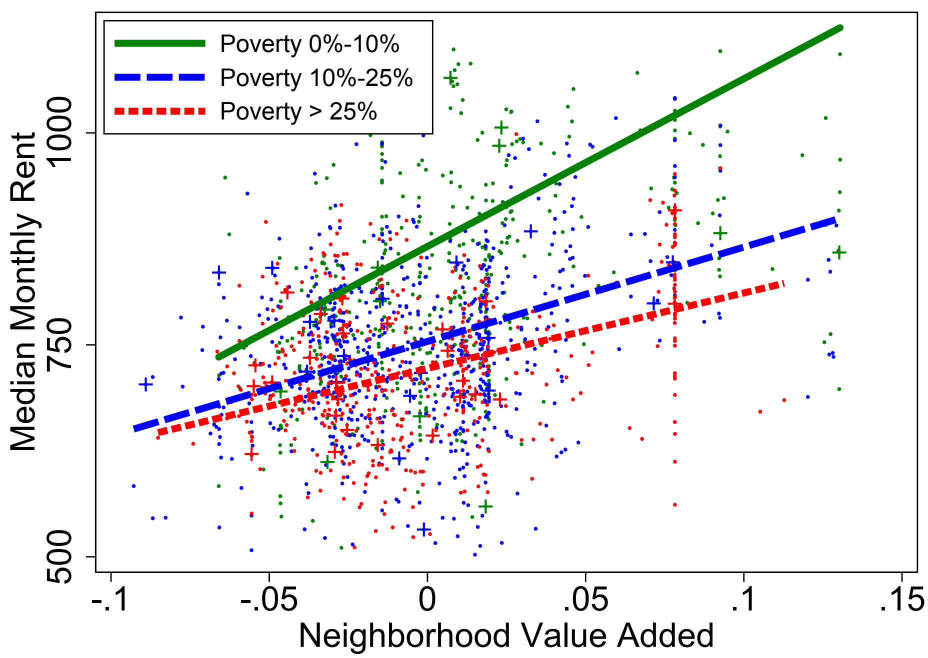


(b) Scenario where MTO decreases child outcomes

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



(a) Math Value Added



(b) Reading Value Added

model 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:[27]

- (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-τ 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 α_τ 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.[28]
- (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.[29]

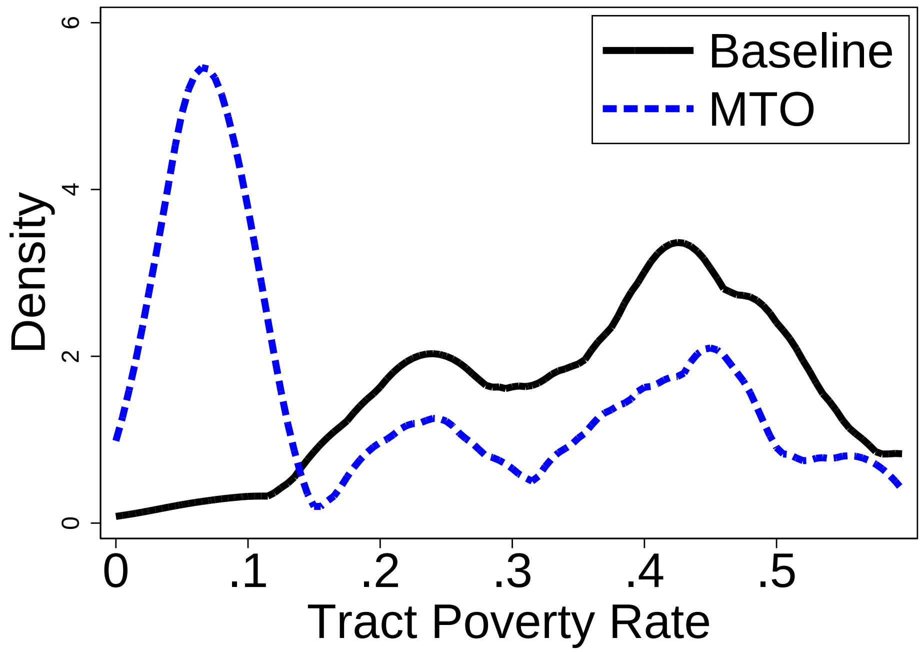
Figure 12 shows the simulated distribution of the poverty rate of locations chosen in all periods of the simulation in our baseline case and in the MTO-A simulations. Simulations last for 18 years following implementation of the MTO policy. The distributions that are shown in figure 12 are for all households eligible to receive a voucher in the MTO-A experiment, not just those that accept the voucher. In our MTO-A simulation, 70% of the population eligible to receive a voucher accept it; the actual MTO take-up rate in Los Angeles was 67%. The fact that we can nearly match the MTO voucher take-up rate suggests our type-specific

[27]Our simulations target households residing at $t = 0$ in a Census tract with at least 100 public housing units. Alternative targeting rules (results not shown) targeting eligibility to residents of tracts with very high poverty rates and/or rates of public assistance yield similar results. 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.

[28]Our calculation is $\$6,000 \approx 12\,[\$766 - 0.30\,(\$10,000/12)]$, 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.

[29]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 $Pr(j) = \frac{1}{J}\Big(\frac{p_j}{1-p_j}\Big)\Big(\frac{1-\overline{p}}{\overline{p}}\Big)$.

Figure 12: MTO vs. Baseline (Poverty), Intent-to-Treat



estimates of α_τ are reasonable.[30] Comparing the black solid line (baseline) with the blue dashed line (MTO-A), our simulations find that most of the people offered a voucher chose to move to a low poverty neighborhood and remain in that neighborhood for an extended period of time.

To summarize the expected impact on child ability of this reduction to poverty exposure, we compute an expected measure of accumulated neighborhood value-added exposure conditional on *accepting* a voucher in the MTO-A experiment. Let i' denote a family that accepts a voucher in the MTO-A experiment, and assume there are $i' = 1, \ldots, \mathcal{I}$ such families. For any given simulation draw s, we hold this set of families fixed for each of the three scenarios (policies) we consider: Baseline, MTO-A and MTO-B.[31] We then compute the expected impact of policy p on child value added measured over $\bar{T}$ periods (5, 10 or 18 years) as

$$\widehat{\mu}_p^{TOT} = \frac{1}{S}\sum_{s=1}^{S}\left[\frac{1}{\mathcal{I}}\sum_{i'=1}^{\mathcal{I}}\sum_{t=1}^{\bar{T}}\widehat{\mu}_{\ell(i',t,s,p)}\right] \tag{24}$$

[30] Galiani, Murphy, and Pantano (2015) also verify their estimates of housing rent sensitivity by comparing predicted and actual take-up rates.

[31] We allow the set of families included in i' to change across simulation draws.

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 $\widehat{\mu}_{\ell(i',t,s,p)}$ 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 $\widehat{\mu}_p^{TOT}$ 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 $\mathcal{T}$ and the number of Census tracts (1,748) as J. Referring to notation in equation (16), we estimate the following sets of parameters

$$\{\theta_\tau\}_{\tau=1}^{\mathcal{T}},\ \{\alpha\}_{\tau=1}^{\mathcal{T}},\ \mathcal{M} \tag{25}$$

where θ_τ is a vector of 357 parameters determining location choice for type τ and $\mathcal{M} = \{\mu_j\}_{j=1}^{J}$ is the vector of parameters determining child value-added in all Census tracts.

θ_τ, α_τ and $\mathcal{M}$ are assumed to be drawn independently for all $\tau = 1, \ldots, \mathcal{T}$. Denote Σ_τ^θ as the variance-covariance matrix of θ_τ, σ_τ^α as the variance of the estimate of α_τ and $\Sigma^{\mathcal{M}}$ as the variance-covariance matrix of $\mathcal{M}$. The parameters in equation (25) are assumed to be distributed with a variance-covariance matrix of

$$\begin{bmatrix} \begin{matrix} \Sigma_1^\theta & 0 & 0 & 0 \\ 0 & \Sigma_2^\theta & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \Sigma_{\mathcal{T}}^\theta \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} \sigma_1^\alpha & 0 & 0 & 0 \\ 0 & \sigma_2^\alpha & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \sigma_{\mathcal{T}}^\alpha \end{matrix} & 0 \\ 0 & 0 & \Sigma^{\mathcal{M}} \end{bmatrix} \tag{26}$$

Table 6: MTO Demonstration vs. Simulation Experiments

	(1)	(2)	(3)	(4)	(5)
	A. Impacts on Woodcock-Johnson ***Math*** Scores (sd=1)				
	MTO Demonstration	Simulation Experiments			
Exposure time	***TOT***	"MTO-A" (***TOT***)	"MTO-B" (***ATE*** of <10% Pov)	p-value H0: (2) = (1)	p-value H0: (3) ≤ (2)
5 years	-0.019	0.019	0.098	0.747	0.001
10 years	-0.052	0.025	0.180	0.743	< 0.001
18 years	--	0.023	0.31	--	< 0.001
	B. Impacts on Woodcock-Johnson ***Reading*** Scores (sd=1)				
	MTO Demonstration	Simulation Experiments			
Exposure time	***TOT***	"MTO-A" (***TOT***)	"MTO-B" (***ATE*** of <10% Pov)	p-value H0: (2) = (1)	p-value H0: (3) ≤ (2)
5 years	0.064	0.034	0.0731	0.539	0.066
10 years	0.006	0.046	0.134	0.634	0.053
18 years	--	0.065	0.231	--	0.059

For Σ_τ^θ and σ_τ^α we use asymptotic standard errors and for $\Sigma^{\mathcal{M}}$ 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 (24).

Table 6 shows our estimates of $\widehat{\mu}_p^{TOT}$ for math scores (top panel) and reading scores (bottom panel). 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 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. 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 and reading 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 Yeung and Pfeiffer (2009). 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 MTO-A and MTO-B results suggest that MTO-subsidized households selected into especially low value-added tracts among the set of eligible low-poverty tracts. The economics of this result are straightforward. Rents are relatively high in high-value-added 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 7). Households in the MTO-A experiment that accepted the housing voucher moved to relatively low-cost, low value-added neighborhoods in tracts with low poverty rates.

5 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 in the population in the utility of different neighborhoods; and we also 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 result reconciles two results in the literature on child outcomes that seem contradictory: The existence of large effects of neighborhoods on changes to child ability, and the overall lack of improvement of children in the MTO experiment. Counterfactual simulations of our model of neighborhood choice strongly suggest that parents receiving vouchers in the MTO experiment moved to the lowest-cost, lowest-value-added neighborhoods among the eligible set. If parents had randomly chosen low-poverty neighborhoods after receiving a voucher, our analysis suggests their children would have shown a remarkable improvement in ability.

References

Aaronson, D. (1998): "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," *The Journal of Human Resources*, 33(4), 915–946. 3, 26

Aliprantis, D., and F. Richter (2016): "Evidence of Neighborhood Effects from Moving to Opportunity: LATEs of Neighborhood Quality," Federal Reserve Bank of Cleveland Working Paper. 4

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on Observed and Unobserved

Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113(1), 151–184. 21

Andersson, F., J. Haltiwanger, M. J. Kutzbach, G. Palloni, H. Pollakowski, and D. H. Weinberg (2016): "Childhood Housing and Adult Earnings: A Between-Siblings Analysis of Housing Vouchers and Public Housing," US Census Bureau Center for Economic Studies Paper No. CES-WP-13-48R. 4

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115(4), 588–638. 4, 15

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2015): "A Dynamic Model of Demand for Houses and Neighborhoods," Duke University Working Paper. 2, 5, 6, 9

Bhutta, N., and B. J. Keys (2015): "Interest Rates and Equity Extraction During the Housing Boom," Kreisman Working Papers Series in Housing Law and Policy No. 3. 9

Bishop, K. C. (2012): "A Dynamic Model of Location Choice and Hedonic Valuation," Working Paper, Washington University in St. Louis. 7

Bishop, K. C., and A. D. Murphy (2011): "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach," *American Economic Review*, 101(3), 625–629. 6

Board of Governors of the Federal Reserve System (2007): "Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit," . 9

Brown, M., S. Stein, and B. Zafar (2013): "The Impact of Housing Markets on Consumer Debt: Credit Report Evidence from 1999 to 2012," Federal Reserve Bank of New York Staff Report No. 617. 9

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104(9), 2593–2632. 3, 19, 21

Chetty, R., and N. Hendren (2015): "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates," Harvard University Working Paper. 33

Chetty, R., N. Hendren, and L. F. Katz (2015): "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," National Bureau of Economic Research Working Paper 21156. 4, 26

Clampet-Lundquist, S., and D. S. Massey (2008): "Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment," *American Journal of Sociology*, 114(1), 107–143. 4

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931. 33

Cutler, D. M., and E. L. Glaeser (1997): "Are Ghettos Good or Bad?," *The Quarterly Journal of Economics*, 112(3), 827–872. 2, 26

Davis, M. A., and E. M. Foster (2005): "A Stochastic Dynamic Model of the Mental Health of Children," *International Economic Review*, 46(3), 837–866. 4

Durlauf, S. N. (2004): *Handbook of Regional and Urban Economics*chap. Neighborhood Effects, pp. 2174–2230. Elsevier B.V. 2, 26

Galiani, S., A. Murphy, and J. Pantano (2015): "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment," University of Maryland Working Paper. 4, 31, 32

Gallagher, J., and D. A. Hartley (2014): "Household Finance after a Natural Disaster: The Case of Hurricane Katrina," Federal Reserve Bank of Cleveland Working Paper No. 14-06. 9

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60(3), 497–529. 7

Ioannides, Y. M., and G. Zanella (2008): "Searching for the Best Neighborhoods: Mobility and Social Interactions," Tufts University Working Paper. 21

Kane, T. J., and D. O. Staiger (2008): "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," National Bureau of Economic Research Working Paper 14607. 3, 19

Kennan, J., and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251. 2, 5, 6, 11

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1), 83–119. 4, 26

Leventhal, T., and J. Brooks-Gunn (2000): "The Neighborhoods They Live in: The Effects of Neighborhood Residence on Child and Adolescent Outcomes," *Psychological Bulletin*, 126(2), 309–337. 2, 26

LUDWIG, J., G. J. DUNCAN, L. A. GENNETIAN, L. F. KATZ, R. C. KESSLER, J. R. KLING, AND L. SANBONMATSU (2013): "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity," *American Economic Review*, 103(3), 226–31. 4

PEBLEY, A. R., AND N. SASTRY (2011): "Los Angeles Family and Neighborhood Study (L.A. FANS)," RAND Corporation, Restricted File, Version 2.5. 17

PETERSON, J. L., AND N. ZILL (1986): "Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children," *Journal of Marriage and Family*, 48(2), 295–307. 20

PINTO, R. (2014): "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity," University of Chicago Working Paper. 4

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033. 7

SANBONMATSU, L., J. R. KLING, G. J. DUNCAN, AND J. BROOKS-GUNN (2006): "Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment," *Journal of Human Resources*, 41(4), 649–691. 4, 26, 34

SANBONMATSU, L., J. LUDWIG, L. F. KATZ, L. A. GENNETIAN, G. J. DUNCAN, R. C. KESSLER, E. ADAM, T. W. MCDAD, AND S. T. LINDAU (2011): "Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation," Discussion paper, U.S. Department of Housing and Urban Development, Office of Policy Development and Research. 34

SCHRANK, F. A., K. S. MCGREW, AND R. W. WOODCOCK (2001): *Woodcock-Johnson III Assessment Service Bulletin Number 2*Riverside Publishing, Itasca, IL. 18

TODD, P. E., AND K. I. WOLPIN (2006): "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 96(5), 1384–1417. 4

YEUNG, W.-J. J., AND K. M. PFEIFFER (2009): "The black-white test score gap and early home environment," *Social Science Research*, 38(2), 412–437. 34