Supplementary Material

Supplement to "Neighborhood effects and housing vouchers"

(*Quantitative Economics*, Vol. 12, No. 4, November 2021, 1307–1346)

Morris A. Davis
Department of Finance and Economics, Rutgers University

Jesse Gregory
Department of Economics, University of Wisconsin—Madison

Daniel A. Hartley
Economic Research Department, Federal Reserve Bank of Chicago

Kegon T. K. Tan
Department of Economics, University of Rochester

Appendix A: Comparing CCP to Census for Los Angeles

Table A1 compares sample statistics from the CCP 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 CCP 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 of Table A1 compares the CCP and Census data for 15 tracts containing large public housing developments, the residents of which will be the focus of some of the analysis of the paper.[1] That row shows the two data sets closely align for these tracts.

Morris A. Davis: mdavis@business.rutgers.edu
Jesse Gregory: jmgregory@ssc.wisc.edu
Daniel A. Hartley: Daniel.A.Hartley@chi.frb.org
Kegon T. K. Tan: ttan8@ur.rochester.edu

[1]We define large developments as those with at least 250 occupied, nonsenior public housing units in 2000. We also include the Census tracts containing Avalon Gardens and Hacienda Village, which are below the 250-unit threshold but are proximate to several large developments.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1664

TABLE A1. Comparison of Equifax and Census data.

Poverty Rate (%)	Avg. Population 2000–2010		Equifax Share	Pct. w/ Mortgage 2008–2012	
	Equifax[a]	Census[b]		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]	24,988	31,400	79.6%	19.1%	16.5%

Note: This table compares population in the Census (column 3) and ACS (column 6) with the implied equivalent population in the Equifax data (columns 2 and 5). Column (4) is the share of the Census population accounted for by the Equifax data, computed as column (2) divided by column (3).

[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 15 tracts with large public housing developments (250+ occupied, nonsenior public housing units in 2000).

APPENDIX B: DETAILS ON INSTRUMENTAL VARIABLES

This Appendix provides additional details on the instrumental variables method that we use to rescale preferences for consumption, housing, and amenities (σ_ϵ) given the variance of the model's i.i.d. preference shocks. Our maximum likelihood estimation of the location choice model identifies the indirect flow utility δ_ℓ provided by each tract for each household type. As shown in Section 2.5.2 of Davis, Gregory, Hartley, and Tan (2021), these estimated indirect utilities δ_ℓ for a given type can be related to tract rent and amenity levels by

$$\delta_\ell = \lambda \cdot \mathcal{O}_\ell - \left(\frac{1}{\sigma_\epsilon}\right) \cdot \alpha \ln r_\ell + \xi_\ell, \tag{A1}$$

where r_ℓ is the tract rent level, and $\mathcal{O}_\ell$ and ξ_ℓ are observed and unobserved tract characteristics. The impact of log-rent on δ_ℓ depends on the budget share devoted to housing (α) and the scale of the ϵ-shocks (σ_ϵ) in consumption utility units. Having already estimated budget shares α (as described in Section 2.5.1 of Davis, Gregory, Hartley, and Tan (2021), the parameter of interest ($\frac{1}{\sigma_\epsilon}$) can be thought of as the coefficient on α times log rent. Because equilibrium rents will almost certainly be correlated with unobserved but valued characteristics of neighborhoods, ξ_ℓ, consistent estimation of this coefficient requires an instrument Z that satisfies two conditions:

(A1) *Instrument relevance*: $\text{cov}(Z_\ell, \ln r_\ell) \neq 0$,

(A2) *Instrument exogeneity*: $\text{cov}(Z_\ell, \xi_\ell) = 0$.

The instruments must be predictive of tract rents but be uncorrelated with the unobservable component of tract amenity utility.[2]

Our choice of instruments follows the approach proposed by Bayer, Ferreira, and McMillan (2007) for adapting the idea of "BLP instruments" (Berry, Levinsohn, and Pakes (1995)) from the IO literature to the urban setting where amenities are spatially correlated. The standard BLP instruments for the market-specific price of a product ℓ are (functions of) the *characteristics* of other products in the same market competing with ℓ. In differentiated consumer product markets, characteristics of competing products will affect the price mark-up that can be charged for ℓ in equilibrium, thus satisfying the instrument relevance condition, but will not directly affect the utility that a consumer derives from product ℓ conditional on choosing ℓ over the competing products, satisfying instrument exogeneity.

In our location choice framework, the city's Census tracts are all "competing products" for one another whose characteristics are candidate instruments. As Bayer, Ferreira, and McMillan (2007) point out, however, the characteristics of tracts that are located very close to a tract ℓ are likely to be related to ξ_ℓ, because the quality of nearby housing can directly affect the utility one derives from living in a place. To address this concern, Bayer, Ferreira, and McMillan (2007) proposed using as instruments the characteristics of the housing stock outside of a 3-mile buffer and including the characteristics of the housing stock inside of the three mile buffer as controls ($\mathcal{O}_\ell$).

The choice of the buffer distance involves a practical tradeoff for the researcher. With a larger buffer, the exclusion restriction is easier to believe, because the housing stock characteristics of neighborhoods farther from ℓ are less likely to provide a direct amenity value to residents of ℓ. However, instrument relevance may decline with the size of the buffer, because neighborhoods further from ℓ are likely be less substitutable with ℓ and, therefore, have a smaller influence on equilibrium rents in ℓ via competition. To be conservative, we chose a buffer of 5 miles instead of the 3-mile buffer used by Bayer, Ferreira, and McMillan (2007) after verifying doing so did not dramatically reduce the power of the first stage.

The rental housing stock characteristics that we include in the instrument list are as follows: The number of bedrooms (shares with 1, 2, 3, and 4 bedrooms), the number of rental units per building (share of units in buildings with 2, 3–4, 5–49, and 50+ units), and the vintage of the rental stock (share constructed pre-1939, shares by decade of construction from the 1940s to the 1980s, and two categories in the 1990s). Specifically, the instruments are the average of each tract-level share among tracts whose centroids are between 5 and 20 miles from the centroid of tract ℓ. We include in the list of control variables $\mathcal{O}_\ell$ each of these variables measured in tract ℓ itself and the average of each tract-level share among tracts with centroids within 5 miles of the centroid of tract ℓ.[3]

[2]Note that the variation that must be instrumented is the log-rent variation exclusively, even though in practice we estimate a coefficient $\alpha \times \ln r_\ell$ to yield a coefficient with the desired structural interpretation. That is because for any given type of household the parameter α does not vary across tracts.

[3]Bayer, Ferreira, and McMillan (2007) also use variables related to land use in distant tracts as instruments.

We constrain the parameter $(\frac{1}{\sigma_\epsilon})$ in equation (A1) to be the same across types (τ) in the IV second stage regression by pooling the type-specific indirect utility measures into a single sample with one observation per type-tract pair (144 types $\times$ 1748 tracts = 251,712 observations). We allow for type-specific coefficients λ^τ on the tract-level controls $\mathcal{O}_\ell$ and estimate a single coefficient on $\alpha^\tau \times \ln r_\ell$, instrumenting for $\alpha^\tau \times \ln r_\ell$ with $\alpha^\tau \times Z_\ell$.

As described in Section 2.5.2 of Davis, Gregory, Hartley, and Tan (2021), the full IV procedure follows the three-step approach used by Bayer, Ferreira, and McMillan (2007). In the first step, we recover an initial estimate of $(\frac{1}{\sigma_\epsilon})$ using the buffered list of housing stock characteristics as instruments. Then, in a second step, we use the estimates of $(\frac{1}{\sigma_\epsilon})$ and λ^τ from the first step, call them $\widehat{\frac{1}{\sigma_\epsilon}}$ and $\widehat{\lambda}^\tau$, to construct a new surface of indirect utilities for each type abstracting from unobservables as

$$\widehat{\delta}_{\ell,\tau} = \widehat{\lambda}^\tau \cdot \mathcal{O}_j - \widehat{\left(\frac{1}{\sigma_\epsilon}\right)} \alpha^\tau \ln r_\ell.$$

We simulate the model using this specification for indirect utility and adjust r_ℓ for all ℓ tracts until the simulated total housing demand in any tract is equal to the observed housing demand in the estimation sample for that tract.[4] This procedure determines market-clearing rents in all tracts in the absence of unobserved amenities. We use these rents as instruments to estimate $(\frac{1}{\sigma_\epsilon})$ in the third and final step.

Table A2 presents the first stage coefficients on the excluded instruments. For simplicity, the coefficients reported are from a single regression of $\ln r_\ell$ on the Z_ℓ and $\mathcal{O}_\ell$, as opposed to the coefficients from the pooled first-stage regression with $\alpha^\tau \times \ln r_\ell$ on the left-hand side (which for each type equal the reported coefficients divided by each type's α^τ). Because each type contributes one observation per tract and α^τ is constant across each type's 1748 observations, this single regression summarizes all of the exogenous variation extracted from the first stage and is sufficient for discussing the strength of the instruments. Column (1) presents first stage estimates from the "first step" that uses all buffered housing stock variables as instruments. The variables describing the distribution of units per building 5 to 20 miles from ℓ are most predictive of ℓ's log-rent (joint p-value: 0.000), followed by the variables describing the age mix of the housing stock (joint p-value: 0.010). The F-stat for the joint significance of all excluded instruments is 5.35 (p-value: 0.000).

The second column reports first stage estimates from the "third step" where the identifying variation is summarized in the single simulated rent instrument. The first stage F-statistic in column (2) is 31.7 (p-value: 0.000). Intuitively, the F-statistic rises 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.

[4] Given Cobb–Douglas preferences, type-specific housing demand in tract ℓ is $\alpha^\tau w^\tau / r_\ell$, where w^τ is type-specific household income.

Table A2. Detail table of IV results.

	(1)	(2)
Share of rental units with X bedrooms		
Share with 1 bedroom	−10.070	
	(8.002)	
Share with 2 bedrooms	−1.288	
	(4.188)	
Share with 3 bedrooms	−8.472	
	(9.604)	
Share with 4 bedrooms	−14.370	
	(16.31)	
Joint significance of bedrooms: p =	0.6340	
Share of renter-occ. units consisting of X units		
2 unit buildings	3.739	
	(16.780)	
3–4 unit buildings	3.033	
	(6.226)	
5–49 unit buildings	3.254	
	(1.706)	
50+ unit buildings	−15.910	
	(6.628)	
Joint significance of units per building: p =	0.0000	
Share of all rental units by vintage		
Share built 1995–1998	29.580	
	(38.720)	
Share built 1990–1994	−87.470	
	(40.220)	
Share built 1980–1989	35.700	
	(30.650)	
Share built 1970–1979	−3.465	
	(33.640)	
Share built 1960–1969	−7.330	
	(31.280)	
Share built 1950–1959	16.130	
	(32.490)	
Share built 1940–1959	−39.630	
	(34.040)	
Share built 1939 or earlier	0.781	
	(31.55)	
Joint significance of rental vintage: p =	0.0102	
Simulated ln(rent) instrument		0.226
		(0.0402)
Controls for own tract housing characteristics	X	X
Controls for housing characteristics w/in 5 miles	X	X
Observations	1748	1748
R-squared	0.669	0.732

(*Continues*)

TABLE A2. *Continued.*

	(1)	(2)
All excluded instruments: F-statistic	5.35	31.69
All excluded instruments: p-value	0.000	0.000

Note: This table shows the results of the 1st and 2nd stages of the IV. Robust standard errors are in parentheses.

APPENDIX C: ASSUMPTIONS ABOUT CHILDREN IN MTO SIMULATIONS

The assumptions we make about the age of children in households in the simulations in order to replicate the results of CHK in our MTO simulations are shown in Table A3. Column (2) of this table shows the number of years we assume the child has lived in the initial location (one of the 15 tracts with public housing) prior to the simulation starting. Column (3) shows the number of years we simulate the model to determine optimal location choices in the baseline, MTO and MTO-R simulations. Column (1) shows the percentage of total simulations accounted for by the combinations shown in columns (2) and (3). We specify the distribution as shown in Table A3 to match three facts: First, the MTO experiment occurred between 1994 and 1998; second, CHK restrict their sample to children that are born on or before 1991; and third, the (significant) results of CHK are for children under the age of 13 at the time they are recruited for the MTO experiment. We assume households enter the MTO experiment uniformly between 1994 and 1998 and children of households in the MTO sample are born uniformly across years.

TABLE A3. Exposure by age in MTO simulations.

Percentage of Simulations (1)	Years Before Simulations Start (2)	Years of Simulations (3)
2.5%	3	15
5.0%	4	14
7.5%	5	13
10.0%	6	12
12.5%	7	11
12.5%	8	10
12.5%	9	9
12.5%	10	8
12.5%	11	7
12.5%	12	6

Note: Column (2) shows the age of the child at which housing assistance in the MTO program is first offered. Column (3) shows the number of years we track the household, such that (2) and (3) sum to 18. Column (1) shows the percentage of times the particular row is included in the simulations.

Appendix D: Opportunity Atlas regressions

For each of the two published Opportunity Atlas scores we use in our analysis—the child's expected percentile in the age-26 nationwide income distribution given household income of the 25th percentile and the 75th percentile of the nationwide income distribution—we regress the published score multiplied by 100 on average household income (in tens of thousands of dollars) and the percentages of the neighborhood that are African–American and Hispanic. The income and race regressors for each of the 1748 tracts are generated using data from our 144 household types.

The regression results are shown in Table A4. Explaining the coefficients, using column (2) as an example: All else equal, if average income increases by $10 thousand, then the predicted percentile in the income distribution of the child's income at age 26 increases by 0.74; if the share of African–American households increases by 10 percentage points then the predicted percentile falls by 2.037; and if the share of Hispanic households increases by 10 percentage points then the predicted percentile falls by 0.88. Remarkably, this simple regression can account for a large share of the variation of the Opportunity Atlas data, as the R2 values are 40% for the 25th percentile regression and 33% for the 75th percentile regression.

Table A4. Regressions of Opportunity Atlas data on income and race.

Regressor (1)	Child Expected Income Percentile (×100) if Household Income is at the: 25th Percentile (2)	75th Percentile (3)
Average household income ($0000s)	0.740	1.836
	(0.239)	(0.276)
African–American share	−20.370	−17.670
	(1.175)	(1.357)
Hispanic share	−8.824	−5.730
	(0.683)	(0.788)
Constant	47.060	49.760
	(1.333)	(1.540)
Observations	1748	1748
R-squared	0.399	0.326

Note: This table shows regressions of Census-tract-level Opportunity Atlas data on tract-level average income (in $0000s) and share of African–American and Hispanic households in the tract. The regressors (the Opportunity Atlas data) are the expected child percentile in the income distribution as an adult (times 100) given household income in the 25th percentile and the 75th percentile. Standard errors are in parentheses.

References

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. [3, 4]

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 63 (4), 841–890. [3]

Davis, M. A., J. Gregory, D. A. Hartley, and K. T. K. Tan (2021), "Neighborhood effects and housing vouchers." Working Paper. [2, 4]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 19 June, 2020; final version accepted 25 February, 2021; available online 29 April, 2021.