Spring 5-6-2019

Essays on Migration, Energy Use, Emissions, and School Assignment

Cody Reinhardt

ABSTRACT

ESSAYS ON MIGRATION, ENERGY USE, EMISSIONS, AND SCHOOL ASSIGNMENT

By

CODY KARL REINHARDT

4/1/2018

Committee Chair: Dr. H. Spencer Banzhaf

Major Department: Economics

This dissertation has two essays. The first studies migration patterns in the U.S. and the relationship between migration patterns and energy use and carbon emissions. It uses a two-city model of energy use and household migration to analyze emission implications from city level green policies. Per-household emissions are calculated for the largest 49 MSA's in the U.S. and data on migration patterns used to assign substitute locations to migrating households. Results show large differences in net carbon emissions from migration, which has implications for a wide range of policies affecting migration decisions.

The second essay studies how school quality is assigned to properties through various methods. It first replicates methods in the literature, such as assignment by distance and district means, and adds new methods to assign measures of school quality to census blocks. Next, these assignments are compared to a new dataset of school assignment to determine accuracy. Both distance matching and assignment by district means are shown to be relatively inaccurate methods of assignment. The accuracy also varies over space and district size.

ESSAYS ON MIGRATION, ENERGY USE, EMISSIONS, AND SCHOOL ASSIGNMENT

BY

CODY KARL REINHARDT

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2019

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair:	Dr. H. Spencer Banzhaf
Committee:	Dr. Kyle Mangum
	Dr. Daniel Kreisman
	Dr. John Winters

Electronic Version Approved:

Dr. Sally Wallace, Dean
Andrew Young School of Policy Studies
Georgia State University
May, 2019

Contents

List of Figures

List of Tables

1. U.S. Internal Migration Networks, Energy Use, and Emissions

1.1 Introduction

Climate change as a result of carbon emissions is a highly studied and broad topic in the economics literature. As noted in Glaeser and Kahn (2010), a significant proportion of US carbon emissions come from household energy use, and urban structure plays a prominent role in how much energy households consume. Mangum (2017) and Glaeser and Kahn (2010) have shown that cities vary greatly in per household levels of emissions, with the high-emission U.S. cities having nearly twice the per-household emissions as the low-emission cities. Glaeser and Kahn examine differences in urban structure and both within city and between city variation in household energy use. This paper extends this literature by using historic internal migration data to examine the role migration plays in the total emissions for the U.S. Given the plethora of local policies on housing and zoning, and the popularity of local green regulations, it is highly unlikely that emissions will be optimally taxed. As noted by Glaeser and Kahn, even an otherwise perfectly calibrated Piguvian carbon tax is not sufficient for optimal location decisions in the presence of local policies or incentives which restrict development in green areas and subsidized development in less green areas. In reality, the U.S. has many such policies and incentives. According to Glaeser, "By restricting new development, the cleanest areas are pushing development to areas of higher emissions" (Glaeser and Kahn). So migration will play a key role in how optimal emission decisions are made from a country perspective, because how the population is distributed and moving among the cities of various emissions levels affects the total country level of emissions.

As household migrate between cities, they change their housing consumption, carbon content of electricity and heating, and driving patterns as they change locations. Any local policies directly or indirectly taxing carbon emissions would have to consider the potential migration effects on emissions an how movement of households to and from their neighbors contributes to the national carbon account. Policies in all of the cities are important, as well as a city's location in the sense of it's largest migration neighbors. While Mangum (2017) considers simulations of national level policies, this paper focuses on local policies with migration effects following historic migration patterns. The purpose of this paper is to examine the role migration plays in the total carbon emissions in the U.S. This paper extends a two-city model first developed in Glaeser and Kahn (2008). It does this by using city pairs constructed from data on MSA emissions and MSA-to-MSA migration data. This will represent the migration effect of the MSA by weighting its migrants with the per-household emissions of their destination MSA. Each MSA will thus have different migration effects, for both out- and in- migration, due to their place in the migration network and the greenness of substitute cities in their part of the network. The paper proceeds as follows. Section 2 presents the two-city model and the generation process for the representative migration city. Section 3 describes the data used in the paper. Section 4 details the results and implications. Section 5 concludes and discusses opportunities for further research.

1.2 The Two-City Model

This section expands on the two-region model presented in Glaeser and Kahn (2008). The original model is introduced and then expanded by considering the changes on energy use. The model contains two regions (which will be defined as cities in this paper) where individuals are free to move between them to maximize utility. They maximize utility by choosing location and energy service consumption. The individual wishes to live in the location where they can get the most utility from energy service consumption, which depends on the price of energy services and that location's utility function with respect to energy. For example, heating and cooling expenses can be expensive in an area with a very mild climate and total energy service consumption could be lower and yield a higher total utility. With income and total population being held constant, the model shows that the distribution of population between regions with different energy prices, energy uses, and external costs of energy service consumption affects total utility. New zoning or tax policies cause a movement between cities as well as a change in energy service consumption within.

The two regions are expanded from abstract areas to constructed empirical areas using migration data to represent the migration effect of a city. The model is presented and then followed by the representative migration city construction. The two-city model begins with individuals maximizing a quasi-linear utility function $Y_i - P_i^H - (P_i^E + t)E_i + t\hat{E} + V_i(E_i; X_i) - C(N\hat{E})$ where Y_i is income, $P_i^H$and P_i^E are prices of housing and energy services for city i; t is an energy use tax; E is energy use in city i; $\hat{E}$ is the national average energy consumption; $V_i(.;.)$ is a function for city-specific benefits from energy services; X_i is a vector of exogenous attributes for location i; $C(N\hat{E})$ is the external cost of energy use by the whole country, which can be thought of as the national contribution to climate change; and N is population. Note that in modeling energy services, I am looking at the cost of, e.g., maintaining a given temperature in the home, which will be a function of energy prices but also house size, weather, and so forth. Finally, note that the tax is revenue neutral, since individuals are receiving a lump sum rebate of $t\hat{E}$. Next, each city i has Q_i^F identical employers, with revenues $f(.)$ increasing and concave in the the number of people hired. Each city has builders Q_i^B, with costs $k(.)$ increasing and convex in buildings constructed. Now wage income is $f'(\frac{N_i}{Q_i^F})$, or the marginal revenue product of labor (MPL), and housing cost is $k'(\frac{N_i}{Q_i^B})$, the marginal cost of supplying housing. Individuals hold equal rights to all business profits. The two equilibrium conditions are as follows: individuals choose privately optimal energy consumption E_i^* to maximize their utility, so $P_i^E + t = V_1(E_i^*; X_i)$, with $V_1(E_i^*; X_i)$ being the first derivative of $V(.;.)$ with respect to E. The next condition is a locational equilibrium, so $f'(\frac{N_i}{Q_i^F}) - k'(\frac{N_i}{Q_i^B}) - (t + P_i^E)E_i^* + V(E_i^*; Z_i)$ must be equal for all cities. Individuals in this model are identical, and the social welfare function used is additive:

$$\sum_i Q_i^F f(\frac{N_i}{Q_i^F}) - Q_i^B k(\frac{N_i}{Q_i^B}) + N_i(V(E_i; X_i) - P_i^E E_i - C(N\hat{E})) \qquad (1)$$

So this yields two first order conditions. The first, for energy consumption, is

$$P_i^E E_i - NC'(N\hat{E}) = V_1(E_i; X_i) \qquad (2)$$

so that the private optimality condition is socially optimal at a tax of $t = NC'(N\hat{E})$. For the last unit of energy service consumption, the price of energy services plus the optimal tax equals the marginal benefit for the city of that unit of energy services.

The first order condition for location decisions is that

$$f'(\frac{N_i}{Q_i^F}) - k'(\frac{N_i}{Q_i^B}) + V(E_i^*; X_i) - E_i(P_i^E + NC^{'}(N\hat{E})) \tag{3}$$

is constant over space. Income plus the benefits from energy services, minus the cost of energy (both price cost and external cost) and cost of housing must be equal for all locations. This gives a locational equilibrium and there is no arbitrage opportunity from changing location.

Consider the case of environmentally inspired land use restrictions. A location can impose a "zoning tax" z_i on new construction. Builders in location 1 now have a first order condition $P_1^H = z_1 + k'(\frac{N_1}{Q_1^B})$. Assume that the tax is returned to inframarginal residents so as to be revenue neutral. Here, Glaeser and Kahn (2008) assume that zoning can affect population sizes but not energy use or energy prices. However, as noted in Mangum (2017), zoning regulations affect the patterns of energy consumption, and are not merely an impediment to the movement of households. The effect of zoning on patterns of energy use in City 1 will be modeled though the cost of energy services, P_1^E. Zoning increases the cost of energy related services, P_1^E. Height restrictions, for example, decrease the ratio of interior living space to exterior building space, known in the literature as the floor-area-ratio (FAR), lowering heating and cooling efficiency and making it more expensive to achieve the same level of energy services E_1; it has been shown that such restrictions are welfare decreasing for the urban resident (Bertaud and Brueckner 2005, Borck and Brueckner 2018). Any zoning which reduces density, such as a minimum lot size, green belt, or height restriction (such as a limit on the FAR) means that the network for electricity must consist of a higher ratio of infrastructure (such as wires and cables) to buildings they service. Electricity transfer over such infrastructure is less than perfect, so increasing this ratio increases costs of providing any level of electricity. In the same way, the fuel requirements for transportation would be higher. Thus $\frac{\partial P_1^E}{\partial z_1} > 0$. However, it is also possible that added green space reduces cooling costs and that zoning decreases dwelling unit size, which would have the opposite effect.

The zoning tax reduces the number of people in location 1. Starting with the locational equilibrium condition for two cities 1 and 2 after adding the zoning cost for city 1,

$$f'\frac{(N_1}{(Q_1^F}) - (k'(\frac{N_1}{Q_1^B}) + z_1) - (t + P_1^E)E_1^* + V(E_1^*; X_1) =$$
$$f'(\frac{N_2}{Q_2^F}) - k'(\frac{N_2}{Q_2^B}) - (t + P_2^E)E_2^* + V(E_2^*; X_2).$$

It is possible to differentiate this condition with respect to zoning z_1:

$$\frac{\partial}{\partial z_1}\Bigg[f'(\frac{N_1}{Q_1^F}) - (k'(\frac{N_1}{Q_1^B}) + z_1) - (t + P_1^E)E_1^* + V(E_1^*; X_1) =$$
$$f'(\frac{N_2}{Q_2^F}) - k'(\frac{N_2}{Q_2^B}) - (t + P_2^E)E_2^* + V(E_2^*; X_2).\Bigg]$$

which yields the expression:

$$(\frac{1}{Q_1^F})f''(\frac{N_1}{Q_1^F})(\frac{\partial N_1}{\partial z_1}) - (\frac{1}{Q_1^B})k''(\frac{N_1}{Q_1^B})(\frac{\partial N_1}{\partial z_1}) - 1 - t(\frac{\partial E_1^*}{\partial z_1}) - (\frac{\partial P_1^E}{\partial z_1})E_1^* - (\frac{\partial E_1^*}{\partial z_1})P_1^E + (\frac{\partial E_1^*}{\partial z_1})V_1(E_1^*; X_1) = (\frac{1}{Q_2^F})f''(\frac{N_2}{Q_2^F})(\frac{\partial N_2}{\partial z_1}) - (\frac{1}{Q_2^B})k''(\frac{N_2}{Q_2^B})(\frac{\partial N_2}{\partial z_1})$$

First, note that with only two cities, $\frac{\partial N_2}{\partial z_1} = -\frac{\partial N_1}{\partial z_1}$. Population gained by city 2 is population lost by city 1 and vice versa. Secondly, recall the private energy optimization $P_i^E + t = V_1(E_i^*; Z_i)$; this cancels terms and leaves the equation ready to be solved for $\frac{\partial N_1}{\partial z_1}$

$$(\frac{1}{Q_1^F})f''(\frac{N_1}{Q_1^F})(\frac{\partial N_1}{\partial z_1}) - (\frac{1}{Q_1^B})k''(\frac{N_1}{Q_1^B})(\frac{\partial N_1}{\partial z_1}) - 1 - (\frac{\partial P_1^E}{\partial z_1})E_1^* =$$

$(-\frac{1}{Q_2^F})f''(\frac{N_2}{Q_2^F})(\frac{\partial N_1}{\partial z_1}) + (\frac{1}{Q_2^B})k''(\frac{N_2}{Q_2^B})(\frac{\partial N_1}{\partial z_1})$

And thus the resulting equation for $\frac{\partial N_1}{\partial z_1}$ is:

$$\frac{\partial N_1}{\partial z_1} = \frac{-1 - (\frac{\partial P_1^E}{\partial z1})E_1^*}{(\frac{1}{Q_1^B})k''(\frac{N_1}{Q_1^B}) + (\frac{1}{Q_2^B})k''(\frac{N_2}{Q_2^B}) - (\frac{1}{Q_1^F})f''(\frac{N_1}{Q_1^F}) - (\frac{1}{Q_2^F})f''(\frac{N_2}{Q_2^F})} < 0. \quad (4)$$

Zoning regulations increase the price of energy services and will cause additional reduction in population 1 relative to a model where zoning has no impact on the price of energy services. The impact from the zoning migration effect on welfare is $((E_2 - E_1)(NC^{'}(N\hat{E}) - t) + z_1)(\frac{\partial N_1}{\partial z_1})$. $(E_2 - E_1)$ is the change in energy consumption from the household moving from city 1 to city 2. $(NC^{'}(N\hat{E}) - t)$ is the external cost of energy use in the zoned city, net of energy taxes. This is positive as long as $(E_1 - E_2)(NC^{'}(N\hat{E}) - t) > z_1$. This effect is welfare improving if 1) city 1 was the high energy use city $((E_1 - E_2) > 0)$ and 2) z_1 is smaller than the difference in energy use times the difference in between social cost of energy use and the energy tax. This is to say that the zoning tax should not be greater than the external cost of energy consumption net of taxes. Assuming energy taxes which are smaller than external cost of energy $((NC^{'}(N\hat{E}) - t) > 0)$, if city 1 is the low-energy city $((E_1 - E_2) < 0)$ then z_1 must be welfare reducing. In other words, if zoning taxes are imposed on low energy use city, they will be counterproductive: they force population away from low energy-use areas and into high energy-use areas. Next consider the effect of a zoning tax on energy services E_1.

Energy service can be broken down into two main types: in-home energy and gasoline from driving. Thus E_1 can be represented as a function: $E_1 = f(Heating(p_h(z_1), p_e, Z_1), Electricity(p_h(z_1), p_e, Z_1), Driving(p_h(z_1), p_e, Z_1$. Z_1 is a vector of city characteristics such as climate. In-home energy services are comprised of heating and electricity, both of which depend on the price of housing, the price of energy services, and city characteristics. Driving depends on price of housing, the price of energy services, and city characteristics. The primary interest for energy is the relationship between per-household energy services and zoning. Thus $\frac{\partial E_1}{\partial z_1}$ depends on zoning's effect on heating, electricity, and driving through price of housing. $\frac{\partial p_h}{\partial z_1}$ is positive; as zoning regulations increase, housing prices increase. And for heating and electricity, $\frac{\partial Heating(.)}{\partial z_1}$ and $\frac{\partial Electricty(.)}{\partial z_1}$ are negative because of two effects: higher housing prices lead to smaller houses built and consumed, reducing energy consumption in-home, because smaller houses will require less energy to heat and cool and use less electricity. Zoning increases the price of energy services P_1^E, reducing quantity demanded of these services. Smaller houses built increases density and reduces average commute distance, reducing driving. Price of energy services includes gasoline and other transport related expenditures, and thus reduces consumption of these services via driving. Finally, simulations of zoning regulations on energy use in Mangum (2017) show a negative correlation at the national level for both in-home energy use and for driving. Thus $\frac{\partial E_1}{\partial z_1}$ is negative. When zoning z_1 is changed, there are effects on the extensive $\frac{\partial N_1}{\partial z_1}$ and intensive $\frac{\partial E_1}{\partial z_1}$ margins. As noted in Mangum(2017), any simulation of national policy necessarily involves changes on both margins. What this means is that high-emission cities will have two carbon-reducing effects from increased zoning: shifting population to cleaner cities (carbon decreasing) and lowering per-household carbon use within the city (carbon decreasing.) However, low-emission cities will have opposing effects from zoning: they can trade higher per-household energy use for more population by decreasing zoning, or trade lower per-household energy use for lower population by increasing zoning. The effect of zoning policies on energy use can be written as:

$$\frac{\partial(NE)}{\partial z_1} = \frac{\partial N_1}{\partial z_1}[E_1 - E_2] + \frac{\partial E_1}{\partial z_1}N_1. \tag{5}$$

The first half is the effect of migration on total energy use; this comes from multiplying the number of people who move out of city 1, $\frac{\partial N_1}{\partial z_1}$, by the energy use differential between city 1 and city 2, $[E_1 - E_2]$. The second half is the effect of zoning policies on per-household energy use within city 1, $\frac{\partial E_1}{\partial z_1}$, times the population of city 1 N_1. Thus equation (5) captures the tradeoffs mentioned above when considering zoning policies and energy use.

Whereas Glaeser and Kahn (2010) consider the carbon intensity of living in arbitrarily compared cities, and whereas Mangum (2017) estimates an equilibrium model without regards to observed patterns of inter-city substitution, I propose to calibrate the carbon intensity of a city's relevant substitutes using the matrix of intercity migration patterns. Thus to expand the two-city model, and to quantify the counterproductive effects described in the two-city model, pairs will be constructed for an MSA and its representative migration city. Two types of representative cities are constructed for each MSA: one representing the target of that MSA's out-migration, and one representing the origin of that MSA's in-migration. The representative out-migration city is a migration-weighted city using all of the cities which receive migration from the MSA. This represents the yearly flow carbon footprint of all migrants moving out of MSA i at year t. For each MSA_k which receives migrants from MSA_i, the percent of out-migration of MSA_i which goes to MSA_k is multiplied by the per-household emissions for MSA_k. This is done for multiple years t. So for $MSA_{i,t}$, the representative out-migration city $R_{i,t}$ is defined:

$$R_{i,t} = \sum_k \frac{Migration_t \, MSA_i \; to \; MSA_k}{\sum_l Migration_t \, MSA_i \; to \; MSA_l} * Emissions(MSA_{k,t}) \forall l \neq i, k \neq i. \tag{6}$$

The representative out-migration city does not include the people who do not move ($k \neq i$ and $l \neq i$). For each $MSA_{i,t}$, the net effect on national emissions from out-migration is:

$(Emissions(MSA_{i,t}) - Emissions(R_{i,t})) * \sum_k Migration_t(MSA_i to MSA_k)$ $for \; k \neq i$,

which is the difference in emissions per household between the MSA and its representative out-migration city times the number of households which migrated out of that MSA. A second set of representative migration cities is also constructed for in-migration. This represents the yearly flow carbon footprint of all migrants who move to MSA_i at year t. For MSA_i, the representative in-migration city R^{IN} it is defined:

$$R^{IN}_{i,t} = \sum_k \frac{Migration_t \, MSA_k \; to \; MSA_i}{\sum_l Migration_t \, MSA_l \; to \; MSA_i} * Emissions(MSA_{k,t}), \forall l \neq i, k \neq i. \tag{7}$$

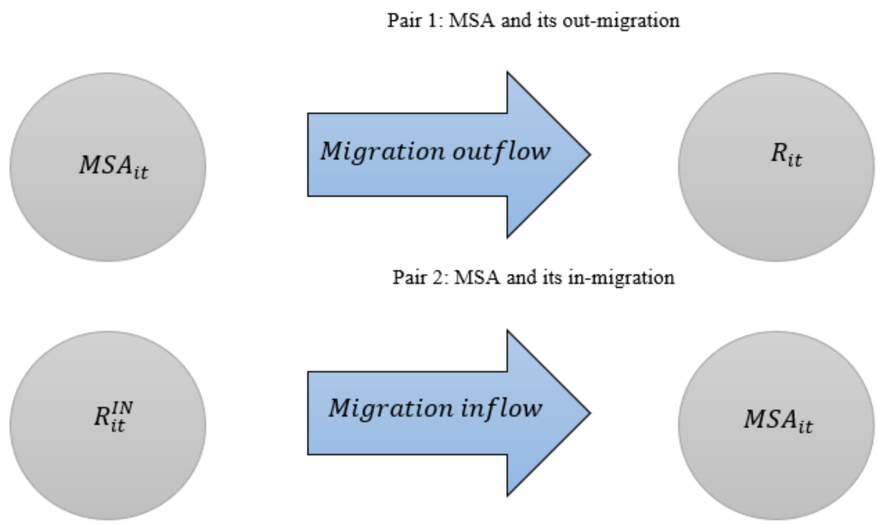
The net effect on national emissions from in-migration is:

$(Emissions(R^{IN}_{i,t}) - Emissions(MSA_{i,t})) * \sum_k Migration_t MSA_k to MSA_i$

$for \; k \neq i$,

which is the difference in emissions per household between the representative in-migration city and the MSA times the number of households which migrated into that MSA. There are two possible pairs of cities to use the two-city model for. These two pairs will be analyzed to show the impact on national emissions from migration to and from major metro areas in the US.They can be seen in Figure 1.

Figure 1: Representative Migration Cities



Notes: The first pair represents out-migration, the households leaving a given MSA. The second pair represents in-migrations, the households entering a given MSA.

1.3 Data

This section describes the data used in this paper. Migration data comes from the IRS tax returns data. These are reported to the IRS as a change in household address from year to year on the head of household tax return. This data has both to and from city, and so gives flows for every county-to-county pair in the US. These counties are aggregated up to the MSA level so that moves in the data represent changes in labor markets rather than local moves. Data for all MSA pairs in the US exists, though only those MSAs with adequate emissions data are included in the analysis. As better emissions data becomes available, more MSAs can be added to the migration network data. These data are a panel of one-way flows for years 1991-2010. Data on energy use closely follows the methodology of Glaeser and Kahn (2010) and Mangum (2017). The goal of this data is to assign per household carbon emissions to each MSA in the analysis for each year in the time horizon. Data for gasoline use comes from the National Highway Transportation Survey (NHTS), which has 5 waves from 1983 until 2009. Total gallons per household are calculated in the same way as Mangum (2017) by regressing gas usage on location and time dummy variables, and then scaled for city household size and proportion of households with personal vehicles obtained from public use census files. This is to be able to use the average driving emissions of the city household rather than the NHTS household. In-home energy use comes from the Residential Energy Consumption Survey (RECS). The energy sources used are fuel oil, natural gas, and electricity. The RECS has seven waves from 1987-2009. Geographic data is relatively limited, including census subregion and metro status. Older homes are known to have higher energy use than newer homes, so the average energy use tends to be higher than the marginal new home energy use. However, newer homes are more often built in the suburbs and are associated with higher gasoline consumption (Glaeser and Kahn 2010). It is possible to distinguish between average energy use and marginal energy use by restricting the sample to homes built in the last 20 years for marginal energy use.

With energy usage data assigned, it is now necessary to standardize energy use in terms of carbon emissions. Glaeser and Kahn (2010) assign 23.46 pounds per gallon of gasoline, 120.6 pounds per 1000 cubic feet of natural gas, and 26.86 per gallon for fuel oil. Carbon content for electricity is determined by state using the North American Electric Reliability Corporation (NERC) carbon content per kilowatt hour. Now each MSA has a household level average annual carbon emission for each year in the time horizon. For a summary of assigned household carbon emissions at the MSA level over time, see Table 8 in the Appendix.

Data for the Wharton Regulation Index is published online by the authors of Gyourko, Saiz, and Summers (2008). This data is used as an indicator of strictness of housing regulations. This analysis is limited to more recent years by the time limitations of the Wharton Index.

1.4 Results

This section details the estimates and results for the representative migration cities. The 49 largest U.S. MSA's have their average household carbon content, representative migration city, and total carbon content from migration calculated for the years 1992, 2000, and 2008. All estimates use the top 49 largest metropolitan areas in the US. These are the cities which have the best available data for emissions at the household and individual level. First, all representative migration cities are calculated by taking the shares of out-migration and multiplying by household level emissions. Table 1 in the appendix details the findings for 2008. Each MSA is identified by numeric MSA code and name. The second column is the population rank of the MSA. Within the sample of cities, they are ranked on average population between 1990 and 2010. The fourth column is the carbon emissions per household of the MSA (origin city). This is for the city listed in the same row. A household is calculated using a representative number of household members which is constant for all MSAs. The fifth column is the total number of households moving out of the MSA for 2008. Note that these only include moves within the sample of MSAs. The sixth column is the per-household carbon of the representative out-migration city (i.e. the aggregate substitute to which migrants from the city are going to). For an out-migrant from the row MSA chosen at random, this is the average new carbon per household of the destination. For comparison, the seventh column shows the per-household difference in carbon emissions between the representative out-migration city and the MSA (Rep - MSA.) A positive number indicates that the MSA has a lower carbon per household emission, and thus each out migrant on net will add to national carbon emissions. The seventh column shows the difference between the representative out-migration city and that national average carbon emission (Rep - National Avg); a positive number indicates that the out-migrants for that MSA are going to places with higher than average carbon emissions, and a negative number indicated the out-migrants are going to places with lower than average emissions. The last column is the total carbon footprint for migration out of the MSA in millions of pounds, and the table is sorted by this value. The average net carbon from out-migration is weighted by that MSA's out migration to return a total carbon footprint for all out migration for that year. Positive numbers show that out-migration (caused by policy or any other reason) increases national carbon emissions, and negative numbers show that out migration is instead carbon reducing. The magnitudes are related to the total migration flows and the other MSAs to which these flows are sent to; some specific cases are discussed later. In Tables 2 and 3, this same layout is repeated for years 2000 and 1992 to see how metropolitan areas and their representative migration cities change over time. Note that carbon emissions for these calculations are in terms of annual emissions added: a result of 10 million pounds means that the migrants from that year add 10 million pounds of carbon to the national emissions *every year*.

It is important to note that the calculations here are using carbon emission values for the average resident. Ideally, data would exist that allow analysis of the marginal mover, in the case that the household moving from city A to city B has consumption patterns different from the average of either city. This analysis will assume that the average moving household had carbon emissions equal to the average of the origin city before moving and then will have carbon emissions equal to the average of the destination city after moving. It is possible that the movers do not represent the average emissions; this could be the case either because of important differences in carbon emissions from marginal uses (economies of scale in electricity generation, increasing congestion from additional drivers, ect) or because the households that move have important characteristics or retain patterns of energy consumption. Glaser and Kahn (2010) use recent construction to try to get closer to the marginal figure, though the occupants of recently constructed homes are a combination of incumbent residents and new movers. Results are generally robust and similar between the average household and those in more recently constructed housing. Mangum (2017) finds that conditioning on demographics, consumption patterns do not depend on the origin of the household (they do not keep their habits from city A.) Having the analysis following the average household is a limitation of not being able to identify movers and their energy consumption patterns.

What the tables show is that the MSAs with the largest footprint for out-migration are those cities with relatively low household level emissions for their region and large total outflows. Los Angeles is at the top of the table for all three years, and for 2008 households leaving LA added over 700 million pounds of carbon emissions per year to the national footprint. New York City and Philadelphia in the northeast, Chicago in the Midwest, Miami in the Southeast, and Seattle in the Northwest all have similar roles in the regions, though have a much smaller outflow and a lower household emissions differential with their representative out-migration cities, and thus a lower footprint. This is not always the case, however, and the migration network plays a large role in carbon footprint from migration. For example, Miami is ranked 5th in carbon footprint from out migration at 177 million annual pounds per year, despite having almost identical household emissions as Salt Lake City, which is ranked 19th in this table and is carbon saving in out-migration. The difference lies in their relevant substitutes. Also notable is that NYC started off in the middle of the pack in 1992, almost carbon neutral, but has risen to the second highest footprint for leavers at over 477 million pounds of carbon in 2008.

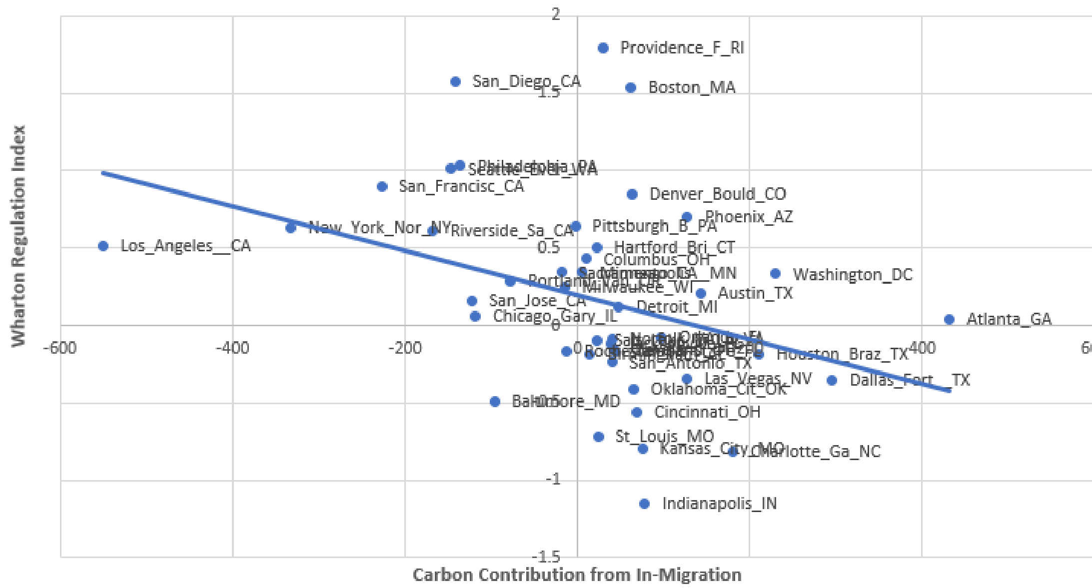
The bottom of the table, occupied by those cities most carbon-saving in out-migration, is occupied primarily by MSAs in the south. Atlanta, Washington DC, Houston, and Dallas are all near the bottom, and thus are carbon-saving from their out-migration. Oklahoma City, Boston, and Las Vegas are also notable examples from other regions, though to a lesser extent. Again it is interesting to see the role of the migration network at play. Washington DC is a close second in terms of carbon saving from out-migration, coming in at a 230 million annual pounds per year reduction, even though it has lower per household emissions than San Antonio, which is essentially carbon neutral in out-migration. Boston is an interesting case because most of its representative out-migration city is New York City, and so becomes grouped as a high-emission city due to its proximity and migration flow relationship with one of the lowest-emission cities.

Next are the results for representative in-migration cities. This represents the emissions for migrants to a city. High-emission cities have a high in-migration footprint. The same tables are constructed for 2008, 2000, and 1992. Southern MSAs dominate the top of the list. The magnitudes have changed but the regional pattern remains similar. Again NYC goes from being an average footprint city in 1992 to a low-emission city in 2008.

Two-City Model

Attention is now turned to the two-city model. MSAs with the highest addition to the carbon footprint, either with a low carbon footprint for leavers or a high carbon footprint for newcomers, are ones which would want the highest housing regulations, all else equal, if the goal is reducing the overall carbon footprint of the US. This would provide the most incentive for households to migrate away from these high emissions MSAs and provide disincentives for migrants to move to these MSAs. There can also be gains from the intensive margin, as stricter zoning can reduce carbon emissions within the city. Unfortunately, according the Wharton Regulation Index, the reality is almost exactly the opposite. LA, which has far and away both the highest contribution to carbon footprint from out-migration and highest carbon footprint savings from in-migration, is the city with far and away the highest Wharton index value, meaning it is the strictest on new housing development. San Francisco, San Diego, Seattle, San Jose, NYC and Miami are all near the top of the regulation list and the top in terms of their contribution from out-migration to carbon footprint and savings from in-migration. Those MSAs which are the most carbon-saving from out-migration and carbon-costing from in-migration (standing to reduce emissions the most on the extensive margin), Atlanta, Dallas, Washington DC, and Oklahoma City, are at the very bottom of the Wharton index, meaning they are the most friendly towards new housing development. This confirms the conjecture of Glaeser and Kahn (2010). Regression results for Figure 2 are shown in Table 7 in the appendix.

Figure 2: Wharton Regulation Index and Carbon from In-Migration



Notes: 1. The vertical axis shows MSA WRI, the horizontal axis shows millons of pounds from in-migration.

2. WRI, a measure of land use regulations, comes from Gyourko et al 2008.

3. A negative relationship shows that more regulated MSAs have larger carbon savings from in-migration.

4. Results significant at 0.05 level.

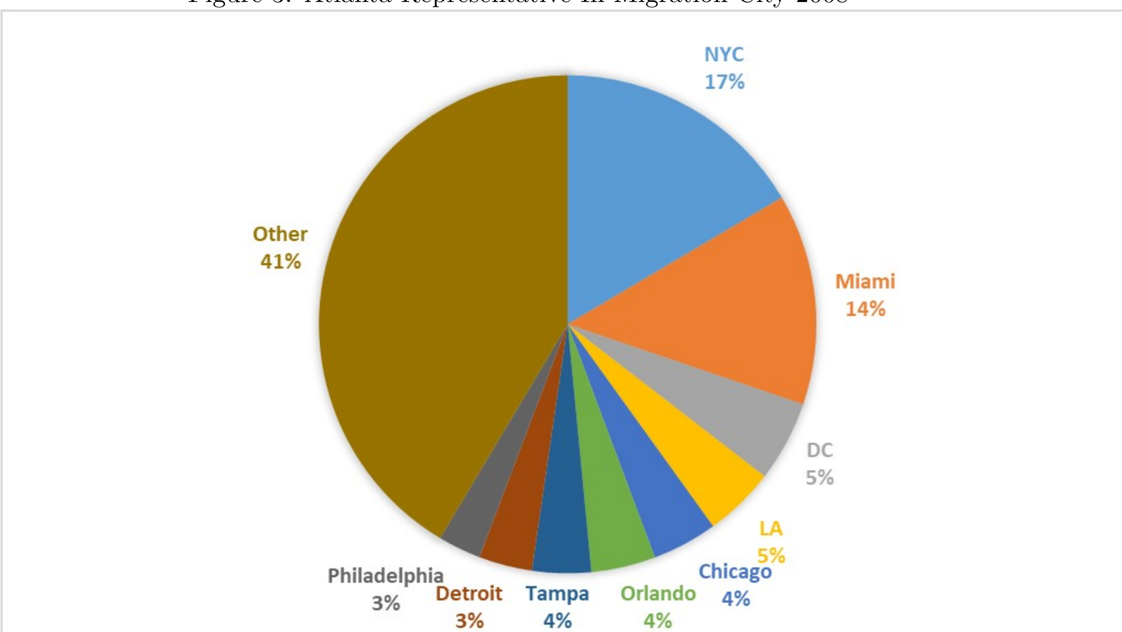
Next special attention is paid to both high-emission and low-emission cities. The high-emission cities to be examined are Atlanta, Washington DC, and San Antonio. These three are selected to show the differing influences of migration flows and within-city carbon emissions on total footprints. In-migration is considered for these cities. Los Angles is selected as the low-emission city. Out-migration is considered for LA.

Atlanta, Georgia

Atlanta is a prototypical high-emission city by the known factors which increase per-household carbon emissions: It is located in the south, people spend a lot of time driving, and people live in large houses. As noted in Mangum (2017), Atlanta is near the top of MSAs in terms of carbon from in home sources and from driving. In terms of migration flows, Atlanta was near the top of the list in total households migrating to it (nearly 42,000 in 2008). All of these factors combine to make Atlanta the dirtiest MSA in the country in 2008 in terms of carbon emissions from in-migration. Note that when discussing migration for a particular MSA, only the migration to and from the top 49 MSAs are considered as a base. 10% of Atlanta's in-migrants means 10% of the total in-migrants from the 48 other MSAs used for the sample. For most cities in the sample, the top 48 other MSA's constitute nearly all of the migration flows. Where are the households moving to Atlanta coming from? For 2008, about 17% of Atlanta's in-migrants in the city sample come from New York City, and about 14% come from Miami. After these two, no other MSAs represent more than 5% of Atlanta's in-migrants. The difference in annual household carbon emissions for Atlanta and NYC is over 15,000 pounds per year; in other words, Atlanta households emit 50% more carbon than NYC households. The difference in annual household carbon emissions for Atlanta and Miami is over 12,000 pounds per year; households in Atlanta emit around 36% more carbon than households in Miami. Accounting for just over 30% of Atlanta's in-migrants, NYC and Miami account for over 40% of its carbon emissions from in-migration. So over 160 million pounds of annual carbon emissions was added by movers from Miami and NYC to Atlanta in 2008. Los Angeles, with one of the lowest carbon emissions per household, accounts for around 5% of Atlanta's in-migration (1,900 households in 2008).

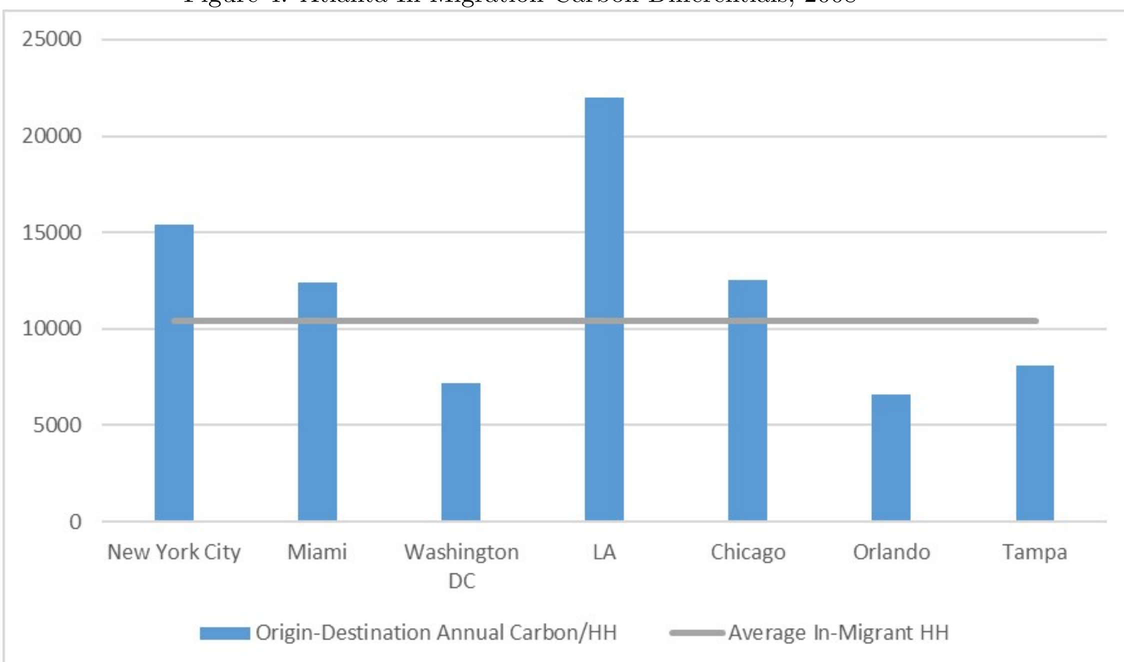
L.A. accounts for over 10% of the carbon contribution from Atlanta's in-migration. The per-household carbon emissions in Atlanta is 93% higher than in L.A., a gap of around 22,000 pounds per year. If L.A. were to send a similar amount of households to Atlanta as NYC does, this would mean an additional 5,000 households moving from L.A. to Atlanta and would increase national annual carbon emissions by 110 million pounds. Figure 3 details Atlanta's in-migration in households for 2008, while Figure 4 details Atlanta's in-migration carbon contributions for 2008. Atlanta's carbon contributions from in-migration for 2000 and 1992 can be found in the appendix.

Figure 3: Atlanta Representative In-Migration City 2008



Notes: This chart shows the proportions of Atlanta's in-migration in 2008. The in-migration for Atlanta for 2008 is 41,318 households.

Figure 4: Atlanta In-Migration Carbon Differentials, 2008



Notes: 1. The bars represent the difference between the destination (Atlanta) and the sending MSA (i.e. New York City) in carbon emissions per household. A bar above zero indicates a net increase in emissions from the move.

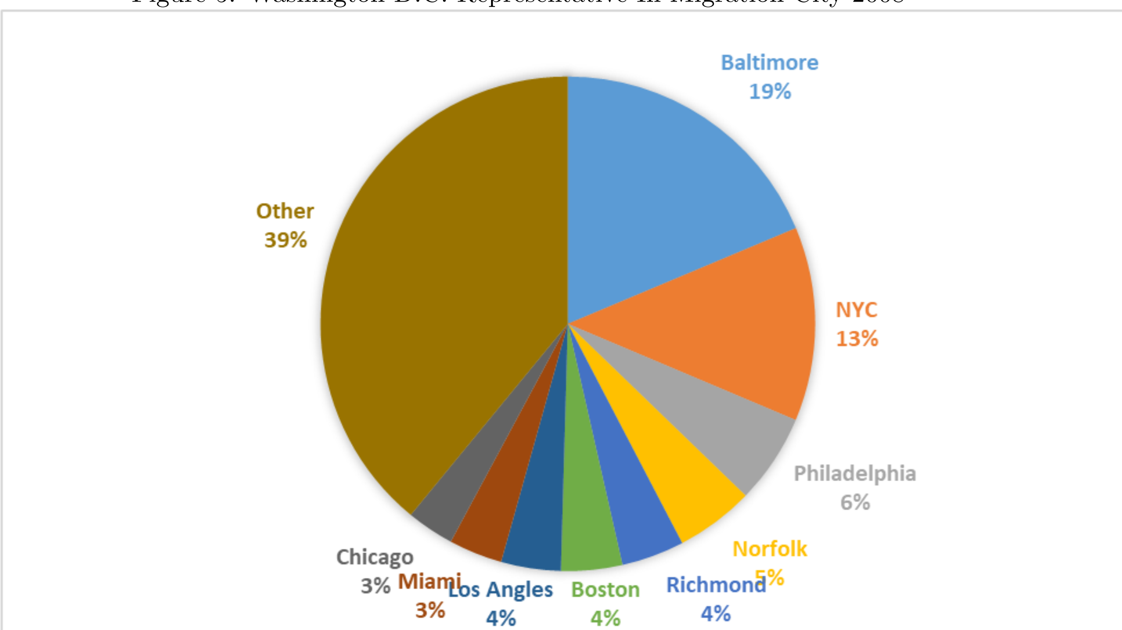
2. The gray line indicates the difference between the destination MSA and its average in-migrant in 2008.

3. MSA's are ordered by their migration share, largest to smallest left to right.

Washington, D.C.

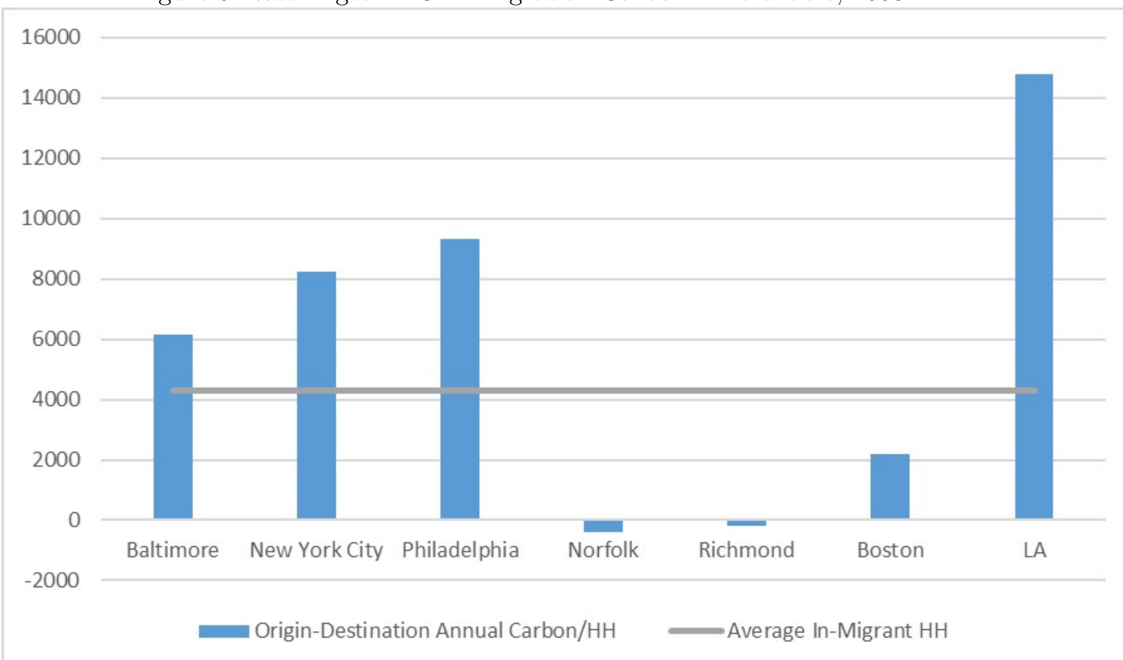
Washington DC has consistently contributed one of the highest carbon totals from in-migration, despite being significantly cleaner in terms of carbon emissions per-household than other cities near the top of in-migration footprint. In 2008 DC had a per household carbon emission of 38,375 pounds. By comparison, Atlanta had a per-household carbon emission of over 45,500 pounds, and Dallas, Houston, Charlotte, and Austin, all cities near the top of in-migration carbon footprint, all had per-household emissions between 42,000 and 45,000 pounds per year. Despite being lower emission than these cities, Washington DC has a large in-carbon footprint because of its migration network, as shown in Figures 5 and 6. The biggest migration senders to DC are Baltimore and NYC, and they have significantly lower emissions per household (32,227 and 30,157 respectively.) This means that these two channels of migration contribute more than half of DC's in-migration carbon footprint, adding 120 million pounds of carbon per year in 2008. Washington DC's carbon contributions from in-migration for 2000 and 1992 can be found in the appendix.

Figure 5: Washington D.C. Representative In-Migration City 2008



Notes: This chart shows the proportions of DC's in-migration in 2008. The in-migration for DC for 2008 is 53,623 households.

Figure 6: Washington DC In-Migration Carbon Differentials, 2008



Notes: 1. The bars represent the difference between the destination (Washington D.C.) and the sending MSA (i.e. Baltimore) in carbon emissions per household. A bar above zero indicates a net increase in emissions from the move.

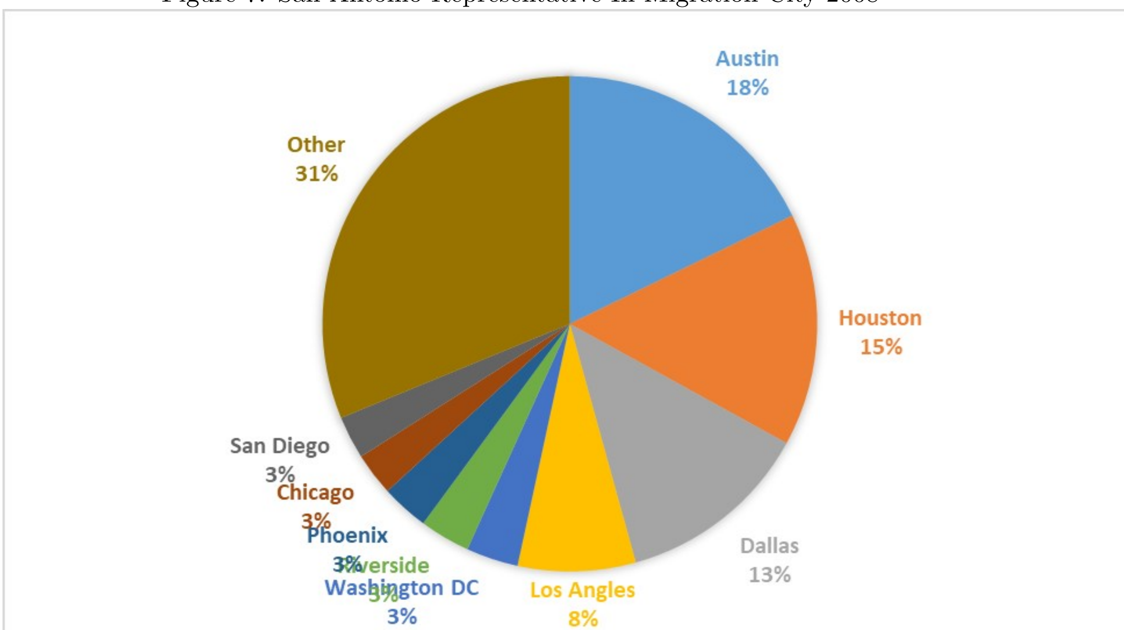
2. The gray line indicates the difference between the destination MSA and its average in-migrant in 2008.

3. MSA's are ordered by their migration share, largest to smallest left to right.

San Antonio, TX

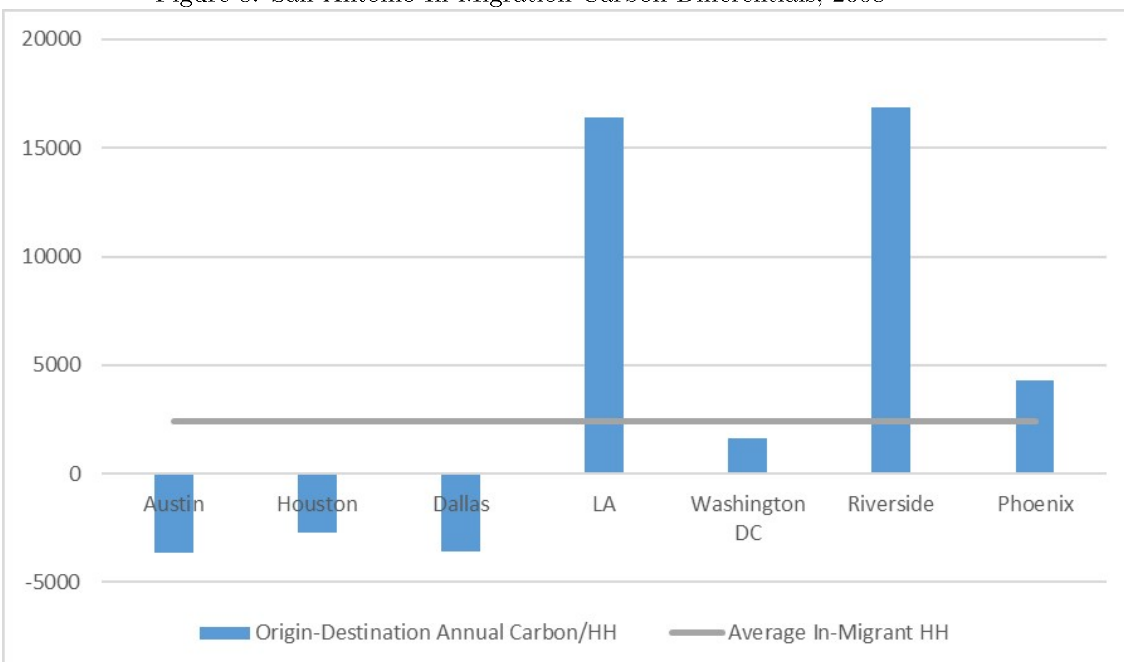
San Antonio is an interesting city and another example of the importance of the interplay between migration network and per-household emissions. In 2008, San Antonio had a per-household emission of 39,994 pounds per year; this is higher than that of Washington DC. However, the carbon footprint of in-migration in San Antonio was only 40 million pounds per year, or about 17% of the footprint from in-migration for Washington DC. Its top 3 migration senders, Austin, Houston, and Dallas (see Figure 7) are all higher than San Antonio in per-household emissions. The in-carbon footprint for these cities is -25 million pounds per year for 2008. Relative to its Texas neighbors, San Antonio is a low-emission city, and so is actually carbon-reducing for these migrants. However, the carbon footprint for movers from Los Angeles to San Antonio is around 21 million pounds per year, or over half of the total net footprint for San Antonio's in-movers. Riverside and San Diego are also large contributors to the in-migration footprint, 9 million pounds per year and 6 million pounds per year respectively, despite only being 3% each of the total in-movers to San Antonio. San Antonio's carbon contributions from in-migration for 2000 and 1992 can be found in the appendix.

Figure 7: San Antonio Representative In-Migration City 2008



Notes: This chart shows the proportions of San Antonio's in-migration in 2008. The in-migration for San Antonio for 2008 is 16,658 households.

Figure 8: San Antonio In-Migration Carbon Differentials, 2008



Notes: 1. The bars represent the difference between the destination (San Antonio) and the sending MSA (i.e. Austin) in carbon emissions per household. A bar above zero indicates a net increase in emissions from the move.

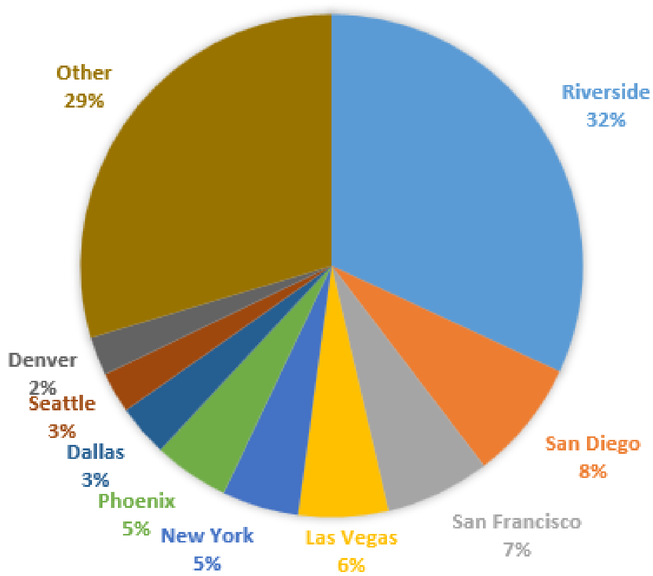
2. The gray line indicates the difference between the destination MSA and its average in-migrant in 2008.

3. MSA's are ordered by their migration share, largest to smallest left to right.

Los Angeles, California

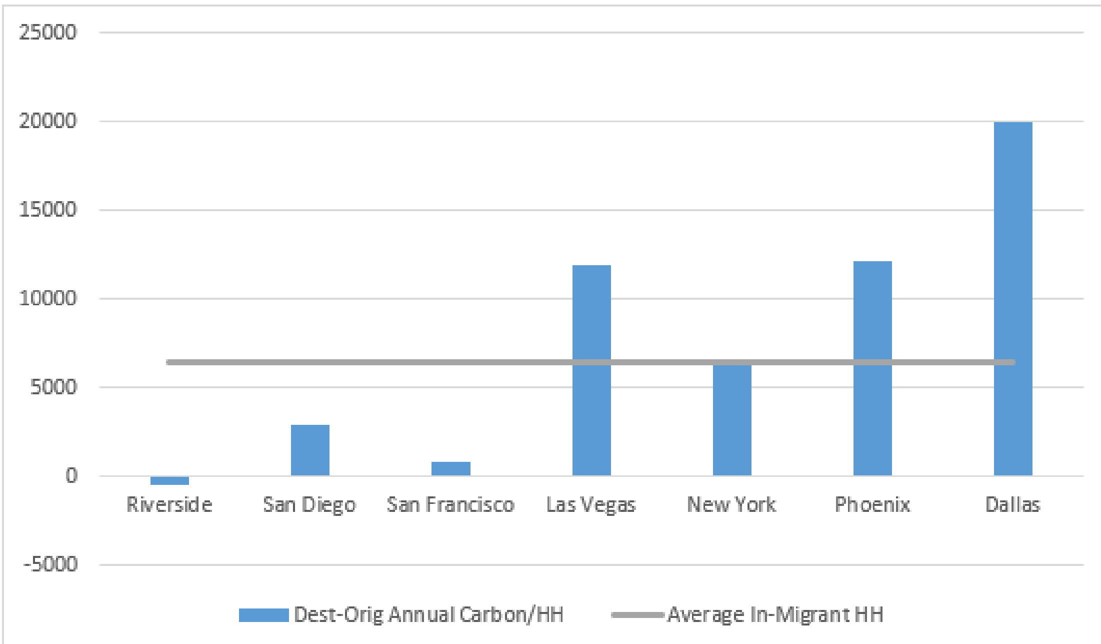
L.A. is a low-emission city and contributes to carbon emissions through out-migration. Thus the representative out-migration city will be used for L.A. In 2008, L.A. had a per-household emission of 23,590 pounds per year, one of the lowest in the country. Homes in Los Angeles don't require much cooling and heating, and the sources of electricity and heating are low carbon in California. Each year 100,000 households migrate out of L.A., and the destination cities have on average 6400 pounds per ear higher emissions per household. The total annual carbon increase from out-migration for L.A. in 2008 was over 700 million pounds. Almost half of the out-migrants are to other cities in California, mostly Riverside, San Diego, and San Francisco. Only a small part of the total carbon footprint comes from these cities. Destinations which receive a smaller portion of the out-migrants from L.A., such as Las Vegas, Phoenix, and Dallas, and Atlanta as we say previously, all have very large carbon footprints. A small percent of L.A.'s migration is still a very large number of households, and the increase in carbon emissions can be nearly double.

Figure 9: L.A. Out-Migration City 2008



Notes: This chart shows the proportions of LA's out-migration in 2008. The out-migration for LA for 2008 is 100,238 households.

Figure 10: L.A. Out-Migration Carbon Differentials, 2008



Notes: 1. For Los Angeles, out-migration is considered. This is different from the previous examples.

2. The bars represent the difference between the destination (i.e. Riverside) and the sending MSA (LA) in carbon emissions per household. A bar above zero indicates a net increase in emissions from the move.

3. The gray line indicates the difference between the average out migration city and the origin MSA in 2008.

4. MSA's are ordered by their migration share, largest to smallest left to right.

1.5 Conclusion

This paper investigates the relationship between the intercity migration network in the US and carbon emissions at the household level. It's not simply the case that some cities are cleaner than others in emissions, but as people move from city to city, they affect the overall carbon output of the country. Thus is it important to know not only the emissions levels of cities, but also their relative position in the migration network and the carbon emissions associated with migration. Certain cities, notably Atlanta and Washington, DC are in a position where they receive many migrants from other cities and have a high per-household emissions factor, and thus growth in these cities increases total carbon output. Certain cities in a large part of the migration network can vastly improve the national carbon footprint by attracting people to migrate there from higher emission cities. Los Angeles, Chicago, and New York City are particularly striking examples of this phenomena, together reducing the annual national carbon footprint by nearly one billion tons per year from in-migration. When it comes to policies which can affect internal migration, we see from analysis using the Wharton Regulation Index that current housing policies greatly add to national carbon emissions on the extensive margin, since the places which are most carbon-saving as destinations are those more heavily regulated than the cities which are most carbon-saving as origins of movers. In the attempts to reduce total national carbon footprint, the ultimate way to reduce the consequences from climate change, it is clear that policies must be aimed at both the household emissions margin and the migration flow margin. Attempting to tax or regulate cities such as New York City or Los Angeles will cause a substantial increase in total national carbon from migration sources.

In 2018, a new regulation was passed in California which requires new homes to be constructed with solar panels, with an increased construction cost z_1 estimated between \$8,000 to \$12,000 per house (Penn 2018). It was passed by unanimous vote by the California Energy Commission with wide public approval. While sure to provide some energy savings from solar energy, the increase to an already regulated and expensive housing market is also sure to have trade-offs not considered by the commission. The gains come in an area which has the best climate and thus lowest need for in-home energy, and replaces energy generated from among the lowest carbon-heavy sources in the country. The increase in housing costs are sure to drive would-be movers and some current residents to migrate elsewhere, and migrating out of California cities will increase the national carbon footprint substantially (Glaeser and Kahn 2010). Local policies passed on their "green merit" can in fact not be green at all, and understanding these trade-offs in terms of energy use and migration flows is the key to evaluating such policies now and in the future.

One way in which policies can be analyzed in the framework of this paper is to consider the net migration effect and increase to carbon emissions nationally through a "ripple" effect. Starting with a change to city A, migration can be simulated to increase proportionally through the migration network for city A. Then, a secondary effect or round would take place, with each of this first group of cities sending additional migration through their networks. Depending on the assumptions for number of rounds and strength of the effect, this can simulate a migration counter-effect locally (1 round), regionally (2-3 rounds) or nationally (large number of rounds). The assumptions for how persistent and how strong each iteration would become can be tested using exogenous events which triggers such a migration ripple (such as major natural disasters.) A well calibrated model would thus be able to analyze policies in terms of their migration and carbon effects based on place, size of the policy, and assumed migration impact.

Appendix

Notes for tables 1-3: 1. The first column contains population rank. 2. MSA Carbon per Household is yearly carbon emissions in pounds. 3. Total Outflow measuring the number households which moved out of the MSA. 4. Representative City Carbon per Household shows the weighted average of carbon emissions for a household leaving the MSA, considering all of the destinations and their migration flow. 5. Rep - MSA Carbon subtracts row 3 from row 5; it shows the per-household increase in emissions nationally from out-migration for the MSA. 6. Rep - National Average Carbon shows the difference between the representative out-migration city carbon emissions and the national average emissions. 7. Total Rep Footprint is the total national increase in carbon emissions, annually, from this MSA's out migration (in millions of pounds.)

Table 1: Representative Out-Migration Cities 2008

Population Rank	MSA	MSA Carbon per HH	Total Outflow	Rep. Carbon per HH	Rep - MSA Carbon per HH	Rep - National Avg Carbon	Total Rep Footprint
2	Los Angeles, CA	23,590	109,238	30,017	6,428	-5,935	702.1
1	New York, NY	30,158	101,943	34,842	4,684	-1,110	477.5
13	Riverside, CA	23,150	49,455	27,930	4,780	-2,228	236.4
12	San Francisco, CA	24,405	45,789	28,756	4,351	-7,196	199.2
6	Miami, FL	33,161	46,192	36,995	3,834	1,043	177.1
4	Philadelphia, PA	29,066	36,596	33,716	4,649	-2,236	170.1
17	San Diego, CA	26,449	40,138	30,407	3,958	-5,545	158.8
3	Chicago, IL	33,031	46,012	35,571	2,540	-381	116.9
15	Seattle, WA	26,857	22,055	32,104	5,247	-3,848	115.7
28	San Jose, CA	23,117	23,159	27,741	4,624	-8,211	107.1
19	Baltimore, MD	32,227	23,737	36,011	3,784	59	89.8
25	Portland, OR	25,706	11,392	30,935	5,229	-5,017	59.6
27	Sacramento, CA	26,098	17,035	27,728	1,630	-8,224	27.8
48	Rochester, NY	30,444	5,886	34,747	4,303	-1,205	25.3
35	Milwaukee, WI	32,052	7,405	35,240	3,188	-712	23.6
41	Buffalo, NY	32,397	6,255	34,684	2,287	-1,268	14.3
20	Pittsburgh, PA	34,623	9,832	35,177	554	-775	5.4
29	San Antonio, TX	39,994	13,591	39,332	-662	3,380	-9
49	Salt Lake City, UT	33,996	6,960	32,579	-1,417	-3,378	-9.9
45	Richmond, VA	38,578	7,286	37,170	-1,408	1,218	-10.3
16	Minneapolis, MN	34,886	13,273	34,050	-836	-1,902	-11.1
47	Birmingham, AL	43,932	3,210	40,088	-3,844	4,136	-12.3
31	Columbus, OH	38,984	10,895	37,479	-1,505	1,527	-16.4
21	Tampa, FL	37,470	22,602	36,608	-862	656	-19.5
44	Jacksonville, FL	39,529	10,649	37,060	-2,469	1,108	-26.3
42	Louisville, KY	45,956	3,812	38,892	-7,064	2,940	-26.9
43	Hartford, CT	38,308	7,181	34,455	-3,853	-1,497	-27.7
18	St Louis, MO	39,487	11,363	36,647	-2,840	695	-32.3
40	Memphis, TN	44,112	6,417	39,030	-5,082	3,078	-32.6
33	Norfolk, VA	38,759	15,498	36,527	-2,232	575	-34.6
32	Providence, RI	38,086	12,275	35,155	-2,931	-797	-36
22	Denver, CO	36,313	18,419	33,754	-2,559	-2,198	-47.1
38	Nashville, TN	46,342	6,734	38,373	-7,968	2,421	-53.7
23	Cleveland, OH	42,533	11,002	36,661	-5,872	709	-64.6
30	Orlando, FL	38,929	22,940	36,062	-2,867	110	-65.8
46	Oklahoma City, OK	51,257	5,122	38,367	-12,891	2,415	-66
34	Indianapolis, IN	45,666	7,530	36,672	-8,994	720	-67.7
37	Charlotte, NC	45,263	8,098	36,432	-8,832	480	-71.5
14	Phoenix, AZ	35,725	27,753	32,962	-2,763	-2,990	-76.7
36	Las Vegas, NV	35,465	20,901	31,665	-3,800	-4,287	-79.4
10	Boston, MA	36,172	34,315	33,832	-2,340	-2,120	-80.3
39	Austin, TX	43,617	16,194	38,500	-5,117	2,548	-82.9
24	Cincinnati, OH	47,185	8,766	37,506	-9,678	1,554	-84.8
26	Kansas City, MO	46,095	8,563	36,167	-9,928	215	-85
9	Detroit, MI	39,388	26,400	35,773	-3,615	-179	-95.4
8	Houston, TX	42,705	27,003	38,028	-4,677	2,076	-126.3
5	Dallas, TX	43,557	33,448	37,443	-6,114	1,491	-204.5
7	Washington, DC	38,376	55,166	34,201	-4,174	-1,751	-230.3
11	Atlanta, GA	45,556	27,493	35,748	-9,807	-204	-269.6

Table 2: Representative Out-Migration Cities 2000

Population Rank	MSA	MSA Carbon per HH	Total Outflow	Rep. Carbon per HH	Rep - MSA Carbon per HH	Rep - National Avg Carbon	Total Rep Footprint
2	Los Angeles, CA	23,672	102,128	30,147	6,475	-7,430	661.3
12	San Francisco, CA	25,123	51,813	29,768	4,645	-7,809	240.7
3	Chicago, IL	33,240	50,385	36,761	3,521	-816	177.4
13	Riverside, CA	23,721	36,008	28,038	4,317	-1,569	155.4
28	San Jose, CA	23,628	29,850	28,460	4,832	-9,117	144.2
17	San Diego, CA	27,056	36,214	30,916	3,860	-6,661	139.8
6	Miami, FL	32,810	24,211	38,359	5,549	782	134.3
15	Seattle, WA	28,130	23,859	32,908	4,778	-4,669	114
4	Philadelphia, PA	33,948	33,558	36,740	2,792	-837	93.7
19	Baltimore, MD	35,983	20,890	39,091	3,108	1,514	64.9
25	Portland, OR	27,380	12,885	31,580	4,200	-5,997	54.1
1	New York, NY	35,538	93,803	36,075	537	-1,502	50.4
27	Sacramento, CA	26,954	14,582	28,884	1,931	-8,693	28.2
35	Milwaukee, WI	32,964	7,220	36,128	3,165	-1,449	22.8
29	San Antonio, TX	40,852	13,502	41,472	620	3,895	8.4
48	Rochester, NY	36,486	5,951	37,573	1,087	-4	6.5
44	Jacksonville, FL	38,831	9,702	39,422	591	1,845	5.7
21	Tampa, FL	38,103	19,097	38,240	137	663	2.6
41	Buffalo, NY	38,538	6,616	37,828	-711	251	-4.7
47	Birmingham, AL	44,738	3,646	42,399	-2,339	4,822	-8.5
31	Columbus, OH	39,717	9,753	38,636	-1,080	1,059	-10.5
30	Orlando, FL	38,656	16,209	37,721	-935	144	-15.2
49	Salt Lake City, UT	35,132	8,630	33,316	-1,816	-4,261	-15.7
18	St Louis, MO	39,294	11,714	37,670	-1,623	93	-19
45	Richmond, VA	43,332	6,487	40,227	-3,106	2,650	-20.1
42	Louisville, KY	46,570	3,931	39,917	-6,653	2,340	-26.2
43	Hartford, CT	41,096	6,720	36,959	-4,136	-618	-27.8
40	Memphis, TN	44,874	6,360	40,420	-4,454	2,843	-28.3
32	Providence, RI	40,250	9,966	36,990	-3,261	-587	-32.5
23	Cleveland, OH	40,964	10,700	37,774	-3,190	197	-34.1
14	Phoenix, AZ	35,565	24,375	34,089	-1,476	-3,488	-36
20	Pittsburgh, PA	41,052	10,992	37,683	-3,368	106	-37
36	Las Vegas, NV	35,249	13,893	32,398	-2,851	-5,179	-39.6
9	Detroit, MI	38,937	18,745	36,783	-2,154	-794	-40.4
33	Norfolk, VA	42,412	13,608	39,218	-3,193	1,641	-43.5
16	Minneapolis, MN	38,930	12,392	35,054	-3,876	-2,523	-48
22	Denver, CO	37,685	19,599	35,053	-2,632	-2,524	-51.6
34	Indianapolis, IN	45,138	7,355	37,832	-7,305	255	-53.7
38	Nashville, TN	48,283	6,735	40,043	-8,240	2,466	-55.5
10	Boston, MA	37,831	33,616	36,110	-1,721	-1,467	-57.9
37	Charlotte, NC	47,916	6,845	39,146	-8,770	1,569	-60
24	Cincinnati, OH	46,174	8,635	38,337	-7,837	760	-67.7
26	Kansas City, MO	45,486	8,831	37,655	-7,831	78	-69.2
39	Austin, TX	45,354	13,734	39,799	-5,555	2,222	-76.3
46	Oklahoma City, OK	53,785	5,706	40,043	-13,742	2,466	-78.4
8	Houston, TX	44,225	28,068	39,670	-4,555	2,093	-127.8
5	Dallas, TX	45,104	34,101	38,973	-6,131	1,396	-209.1
11	Atlanta, GA	48,299	24,888	37,517	-10,782	-60	-268.3
7	Washington, DC	42,371	47,770	36,341	-6,031	-1,236	-288.1

Table 3: Representative Out-Migration Cities 1992

Population Rank	MSA	MSA Carbon per HH	Total Outflow	Rep. Carbon per HH	Rep - MSA Carbon per HH	Rep - National Avg Carbon	Total Rep Footprint
2	Los Angeles, CA	24,625	115,531	29,995	5,371	-6,169	620.5
12	San Francisco, CA	26,068	45,045	29,714	3,646	-6,450	164.2
4	Philadelphia, PA	30,237	30,800	35,298	5,061	-866	155.9
3	Chicago, IL	32,103	39,249	35,409	3,306	-755	129.7
13	Riverside, CA	24,907	36,439	28,394	3,487	-7,770	127
17	San Diego, CA	27,559	36,552	30,777	3,219	-5,387	117.6
6	Miami, FL	31,149	19,535	36,549	5,400	385	105.5
28	San Jose, CA	24,802	24,014	28,836	4,033	-7,328	96.9
19	Baltimore, MD	34,354	17,642	37,670	3,316	1,506	58.5
15	Seattle, WA	28,307	6,167	33,491	5,185	-2,673	32
25	Portland, OR	27,693	8,052	30,698	3,005	-5,466	24.2
29	San Antonio, TX	38,151	11,071	39,979	1,828	3,815	20.2
35	Milwaukee, WI	31,387	5,902	34,734	3,347	-1,430	19.8
27	Sacramento, CA	27,664	14,034	29,016	1,352	-7, 148	19
44	Jacksonville, FL	35,516	8,069	37,453	1,937	1,289	15.6
21	Tampa, FL	35,903	18,293	36,482	579	318	10.6
48	Rochester, NY	33,453	4,777	35,281	1,828	-883	8.7
1	New York, NY	34,097	89,247	34,176	79	-1,988	7.1
20	Pittsburgh, PA	35,643	8,927	35,910	267	-254	2.4
41	Buffalo, NY	34,908	5,401	35,303	394	-861	2.1
18	St Louis, MO	36,502	11,220	36,495	-8	331	-0.1
31	Columbus, OH	37,131	7,412	37,081	-50	917	-0.4
30	Orlando, FL	36,272	13,465	36,226	-47	62	-0.6
49	Salt Lake City, UT	33,587	5,363	32,721	-866	-3,443	-4.6
14	Phoenix, AZ	33,507	19,487	33,171	-336	-2,993	-6.5
36	Las Vegas, NV	33,493	9,568	32,439	-1,054	-3,725	-10.1
47	Birmingham, AL	43,758	2,898	39,874	-3,883	3,710	-11.3
45	Richmond, VA	41,107	5,249	38,508	-2,600	2,344	-13.6
9	Detroit, MI	36,265	17,614	35,392	-872	-772	-15.4
23	Cleveland, OH	37,987	8,729	35,924	-2,063	-240	-18
33	Norfolk, VA	39,559	8,347	37,309	-2,250	1,145	-18.8
42	Louisville, KY	44,432	3,206	38,238	-6,195	2,0714	-19.9
37	Charlotte, NC	44,591	4,580	38,050	-6,541	1,886	-30
40	Memphis, TN	43,314	6,178	38,439	-4,875	2,275	-30.1
34	Indianapolis, IN	41,479	5,873	36,326	-5,154	162	-30.3
32	Providence, RI	39,772	10,069	35,979	-3,793	-185	-38.2
26	Kansas City, MO	42,434	7,219	36,884	-5,550	720	-40.1
43	Hartford, CT	40,610	7,685	35,359	-5,251	-805	-40.4
16	Minneapolis, MN	38,208	9,845	34,094	-4,114	-2,070	-40.5
38	Nashville, TN	47,354	4,889	38,798	-8,556	2,634	-41.8
24	Cincinnati, OH	42,737	7,487	36,830	-5,907	666	-44.2
39	Austin, TX	44,361	10,267	39,560	-4,801	3,396	-49.3
46	Oklahoma City, OK	52,115	4,986	38,691	-13,424	2,527	-66.9
22	Denver, CO	38,768	15,377	34,170	-4,598	-1,994	-70.7
10	Boston, MA	37,557	31,490	34,877	-2,681	-1,287	-84.4
8	Houston, TX	43,611	23,082	38,308	-5,303	2,144	-122.4
11	Atlanta, GA	45,992	18,221	36,669	-9,323	505	-169.9
5	Dallas, TX	44,870	31,000	37,958	-6,913	1,794	-214.3
7	Washington, DC	41,133	44,419	35,156	-5,978	-1,008	-265.5

Notes for tables 4-6: 1. The first column contains population rank. 2. MSA Carbon per Household is yearly carbon emissions in pounds. 3. Total Inflow measuring the number households which moved into the MSA. 4. In-Rep Carbon per Household shows the weighted average of carbon emissions for a household entering the MSA, considering all of the origins and their migration flow. 5. MSA - In-Rep Carbon subtracts row 3 from row 5; it shows the per-household increase in emissions nationally from in-migration for the MSA. 6. Rep - National Average Carbon shows the difference between the representative in-migration city carbon emissions and the national average emissions. 7. Total Rep Footprint is the total national increase in carbon emissions, annually, from this MSA's in migration (in millions of pounds.)

Table 4: Representative In-Migration Cities 2008

Population Rank	MSA	MSA Carbon per HH	Total Inflow	In-Rep. Carbon per HH	MSA - In-Rep. Carbon per HH	Rep - National Avg Carbon	Total Rep Footprint
11	Atlanta, GA	45,556	41,318	35,119	10,437	-833	431.2
5	Dallas, TX	43,557	41,747	36,479	7,078	527	295.5
7	Washington, DC	38,376	53,623	34,096	4,279	-1,856	229.5
8	Houston, TX	42,705	35,046	36,695	6,010	743	210.6
37	Charlotte, NC	45,263	17,572	34,981	10,282	-971	180.7
39	Austin, TX	43,617	22,842	37,367	6,250	1,415	142.8
14	Phoenix, AZ	35,725	37,435	32,311	3,414	-3,641	127.8
36	Las Vegas, NV	35,465	27,821	30,895	4,570	-5,057	127.1
30	Orlando, FL	38,929	25,035	35,033	3,896	-919	97.5
38	Nashville, TN	46,342	9,626	37,750	8,592	1,798	82.7
34	Indianapolis, IN	45,666	8,066	35,984	9,682	32	78.1
26	Kansas City, MO	46,095	7,401	35,894	10,201	-58	75.5
24	Cincinnati, OH	47,185	7,092	37,406	9,779	1,454	69.3
46	Oklahoma City, OK	51,257	4,670	37,309	13,949	1,357	65.1
22	Denver, CO	36,313	24,371	33,706	2,607	-2,246	63.5
10	Boston, MA	36,172	30,893	34,162	2,010	-1,790	62.1
9	Detroit, MI	39,388	11,627	35,277	4,111	-675	47.8
21	Tampa, FL	37,470	26,098	35,761	1,709	-191	44.6
23	Cleveland, OH	42,533	6,785	36,394	6,139	442	41.7
33	Norfolk, VA	38,759	12,804	35,605	3,154	-347	40.4
29	San Antonio, TX	39,994	16,658	37,582	2,413	1,630	40.2
44	Jacksonville, FL	39,529	12,441	36,365	3,165	413	39.4
40	Memphis, TN	44,112	5,111	38,040	6,072	2,088	31
42	Louisville, KY	45,956	4,003	38,520	7,435	2,568	29.8
32	Providence, RI	38,086	9,840	35,075	3,011	-877	29.6
18	St Louis, MO	39,487	8,939	36,775	2,712	823	24.2
49	Salt Lake City, UT	33,996	8,601	31,376	2,621	-4,576	22.5
43	Hartford, CT	38,308	5,221	34,059	4,249	-1,893	22.2
45	Richmond, VA	38,578	8,885	36,581	1,997	629	17.7
47	Birmingham, AL	43,932	3,211	39,367	4,565	3,415	14.7
31	Columbus, OH	38,984	9,894	37,972	1,012	2,020	10
16	Minneapolis, MN	34,886	10,632	34,373	513	-1,579	5.5
20	Pittsburgh, PA	34,623	7,492	34,995	-372	-957	-2.8
41	Buffalo, NY	32,397	4,170	33,869	-1,471	-2,083	-6.1
48	Rochester, NY	30,444	3,726	33,780	-3,336	-2,172	-12.4
35	Milwaukee, WI	32,052	5,581	34,714	-2,662	-1,238	-14.9
27	Sacramento, CA	26,098	17,925	27,137	-1,039	-8,815	-18.6
25	Portland, OR	25,706	15,843	30,596	-4,890	-5,356	-77.5
6	Miami, FL	33,161	39,352	35,193	-2,032	-759	-80
19	Baltimore, MD	32,227	24,569	36,155	-3,927	203	-96.5
3	Chicago, IL	33,031	40,844	35,914	-2,883	-38	-117.8
28	San Jose, CA	23,117	23,176	28,393	-5,276	-7,559	-122.3
4	Philadelphia, PA	29,066	33,519	33,132	-4,065	-2,820	-136.3
17	San Diego, CA	26,449	39,192	30,058	-3,608	-5,894	-141.4
15	Seattle, WA	26,857	27,109	32,264	-5,406	-3,688	-146.6
13	Riverside, CA	23,150	53,307	26,295	-3,144	-9,657	-167.6
12	San Francisco, CA	24,405	50,554	28,897	-4,492	-7,055	-227.1
1	New York, NY	30,158	82,814	34,172	-4,014	-1,780	-332.4
2	Los Angeles, CA	23,590	88,497	29,804	-6,214	-6,148	-549.9

Table 5: Representative In-Migration Cities 2000

Population Rank	MSA	MSA Carbon per HH	Total Inflow	In-Rep. Carbon per HH	MSA - In-Rep. Carbon per HH	Rep - National Avg Carbon	Total Rep Footprint
11	Atlanta, GA	48,299	36,598	37,755	10,545	178	385.9
7	Washington, DC	42,371	49,977	36,710	5,661	-867	282.9
5	Dallas, TX	45,104	35,654	38,949	6,155	1,372	219.4
8	Houston, TX	44,225	24,862	39,153	5,072	1,576	126.1
37	Charlotte, NC	47,916	10,919	38,467	9,449	890	103.2
36	Las Vegas, NV	35,249	25,680	31,234	4,015	-6,343	103.1
39	Austin, TX	45,354	19,972	40,207	5,148	2,630	102.8
38	Nashville, TN	48,283	7,523	39,855	8,428	2,278	63.4
46	Oklahoma City, OK	53,785	4,177	38,757	15,029	1,180	62.8
22	Denver, CO	37,685	26,711	35,356	2,330	-2,221	62.2
26	Kansas City, MO	45,486	7,627	37,584	7,902	7	60.3
33	Norfolk, VA	42,412	13,047	37,880	4,532	303	59.1
34	Indianapolis, IN	45,138	7,459	37,454	7,683	-123	57.3
14	Phoenix, AZ	35,565	33,918	33,887	1,678	-3,690	56.9
24	Cincinnati, OH	46,174	7,182	38,713	7,461	1,136	53.6
16	Minneapolis, MN	38,930	11,487	35,195	3,735	-2,382	42.9
10	Boston, MA	37,831	29,852	36,503	1,328	-1,074	39.7
32	Providence, RI	40,250	10,957	37,019	3,231	-558	35.4
9	Detroit, MI	38,937	14,931	36,900	2,037	-677	30.4
40	Memphis, TN	44,874	5,512	39,506	5,368	1,929	29.6
30	Orlando, FL	38,656	21,157	37,271	1,385	-306	29.3
43	Hartford, CT	41,096	5,420	36,318	4,778	-1,259	25.9
45	Richmond, VA	43,332	7,259	39,859	3,473	2,282	25.2
20	Pittsburgh, PA	41,052	6,759	37,366	3,686	-211	24.9
42	Louisville, KY	46,570	3,703	39,876	6,694	2,299	24.8
23	Cleveland, OH	40,964	7,050	37,996	2,969	419	20.9
49	Salt Lake City, UT	35,132	7,032	32,294	2,837	-5,283	20
18	St Louis, MO	39,294	9,010	37,870	1,424	293	12.8
47	Birmingham, AL	44,738	2,858	41,112	3,627	3,535	10.4
29	San Antonio, TX	40,852	12,188	40,196	656	2,619	8
41	Buffalo, NY	38,538	3,346	36,880	1,659	-697	5.5
31	Columbus, OH	39,717	9,607	39,219	497	1,642	4.8
21	Tampa, FL	38,103	25,313	37,973	130	396	3.3
44	Jacksonville, FL	38,831	9,907	38,541	290	964	2.9
48	Rochester, NY	36,486	3,740	37,135	-649	-442	-2.4
27	Sacramento, CA	26,954	20,204	27,304	-350	-10,273	-7.1
35	Milwaukee, WI	32,964	5,485	35,561	-2,598	-2,016	-14.2
1	New York, NY	35,538	62,580	35,973	-435	-1,604	-27.2
25	Portland, OR	27,380	13,348	31,677	-4,297	-5,900	-57.4
19	Baltimore, MD	35,983	22,702	39,495	-3,512	1,918	-79.7
4	Philadelphia, PA	33,948	29,970	36,673	-2,726	-904	-81.7
15	Seattle, WA	28,130	23,688	33,150	-5,020	-4,427	-118.9
13	Riverside, CA	23,721	50,652	26,386	-2,665	-11,191	-135
3	Chicago, IL	33,240	36,288	37,256	-4,016	-321	-145.7
17	San Diego, CA	27,056	38,602	30,900	-3,844	-6,677	-148.4
28	San Jose, CA	23,628	26,082	29,500	-5,872	-8,077	-153.1
6	Miami, FL	32,810	41,102	37,260	-4,449	-317	-182.9
12	San Francisco, CA	25,123	57,726	30,190	-5,067	-7,387	-292.5
2	Los Angeles, CA	23,672	79,455	31,169	-7,497	-6,408	-595.7

Table 6: Representative In-Migration Cities 1992

Population Rank	MSA	MSA Carbon per HH	Total Inflow	In-Rep. Carbon per HH	MSA - In-Rep. Carbon per HH	Rep - National Avg Carbon	Total Rep Footprint
11	Atlanta, GA	45,992	29,387	36,075	9,917	-89	291.4
7	Washington, DC	41,133	41,416	34,910	6,223	-1,254	257.7
5	Dallas, TX	44,870	30,029	37,616	7,254	1,452	217.8
8	Houston, TX	43,611	26,508	37,822	5,789	1,658	153.5
22	Denver, CO	38,768	23,025	33,756	5,012	-2,408	115.4
39	Austin, TX	44,361	13,417	39,261	5,100	3,097	68.4
46	Oklahoma City, OK	52,115	4,478	37,375	14,740	1,211	66
38	Nashville, TN	47,354	6,665	37,841	9,513	1,677	63.4
37	Charlotte, NC	44,591	7,039	36,478	8,113	314	57.1
24	Cincinnati, OH	42,737	7,419	36,348	6,389	184	47.4
33	Norfolk, VA	39,559	14,068	36,346	3,213	182	45.2
36	Las Vegas, NV	33,493	17,260	31,006	2,487	-5,158	42.9
26	Kansas City, MO	42,434	6,726	36,223	6,211	59	41.8
16	Minneapolis, MN	38,208	10,233	34,172	4,036	-1,992	41.3
10	Boston, MA	37,557	20,499	35,548	2,009	-616	41.2
34	Indianapolis, IN	41,479	6,614	35,963	5,517	-201	36.5
40	Memphis, TN	43,314	5,532	37,547	5,767	1,383	31.9
32	Providence, RI	39,772	7,979	36,364	3,408	200	27.2
42	Louisville, KY	44,432	3,523	37,671	6,762	1,507	23.8
43	Hartford, CT	40,610	4,139	35,237	5,373	-927	22.2
14	Phoenix, AZ	33,507	25,012	32,680	828	-3,484	20.7
45	Richmond, VA	41,107	6,058	37,865	3,242	1,701	19.6
23	Cleveland, OH	37,987	7,001	35,622	2,366	-542	16.6
47	Birmingham, AL	43,758	3,071	39,058	4,699	2,894	14.4
9	Detroit, MI	36,265	11,934	35,069	1,196	-1.095	14.3
49	Salt Lake City, UT	33,587	6,814	31,513	2,074	-1,095	14.1
30	Orlando, FL	36,272	18,060	35,603	669	-561	12.1
20	Pittsburgh, PA	35,643	8,040	35,292	351	-872	2.8
18	St Louis, MO	36,502	8,219	36,207	295	43	2.4
41	Buffalo, NY	34,908	4,050	34,455	454	-1,709	1.8
31	Columbus, OH	37,131	8,183	37,024	107	860	0.9
21	Tampa, FL	35,903	24,655	35,950	-46	-214	-1.1
1	New York, NY	34,097	43,813	34,166	-69	-1,998	-3
27	Sacramento, CA	27,664	17,685	27,832	-169	-8,332	-3
48	Rochester, NY	33,453	4,074	35,108	-1,654	-1,056	-6.7
29	San Antonio, TX	38,151	11,277	39,034	-883	2,870	-10
44	Jacksonville, FL	35,516	9,623	36,865	-1,349	701	-13
35	Milwaukee, WI	31,387	5,337	34,258	-2,871	-1,906	-15.3
25	Portland, OR	27,693	11,464	30,433	-2,740	-5,731	-31.4
19	Baltimore, MD	34,354	19,850	38,139	-3,786	1,975	-75.1
13	Riverside, CA	24,907	53,015	26,481	-1,574	-9,683	-83.4
28	San Jose, CA	24,802	21,242	28,944	-4,142	-7,220	-88
15	Seattle, WA	28,307	24,973	31,948	-3,641	-4,216	-90.9
17	San Diego, CA	27,559	37,181	30,014	-2,455	-6,150	-91.3
3	Chicago, IL	32,103	32,733	35,338	-3,235	-826	-105.9
4	Philadelphia, PA	30,237	25,575	35,076	-4,839	-1,088	-123.8
12	San Francisco, CA	26,068	46,649	29,727	-3,659	-6,437	-170.7
6	Miami, FL	31,149	42,064	35,363	-4,214	-801	-177.2
2	Los Angeles, CA	24,625	76,364	30,946	-6,321	-5,218	-482.7

Table 7: Regression Results for Wharton Regulation Index

Variable	Estimate
Representative City Carbon Footprint (Millions of lbs)	-0.0014*** (0.00064)
Constant	0.2173 (0.10193)
Adj. R-Squared	0.0769
Observations	44

Figure 11: Washington D.C. In-Representative Carbon Differential, 2000

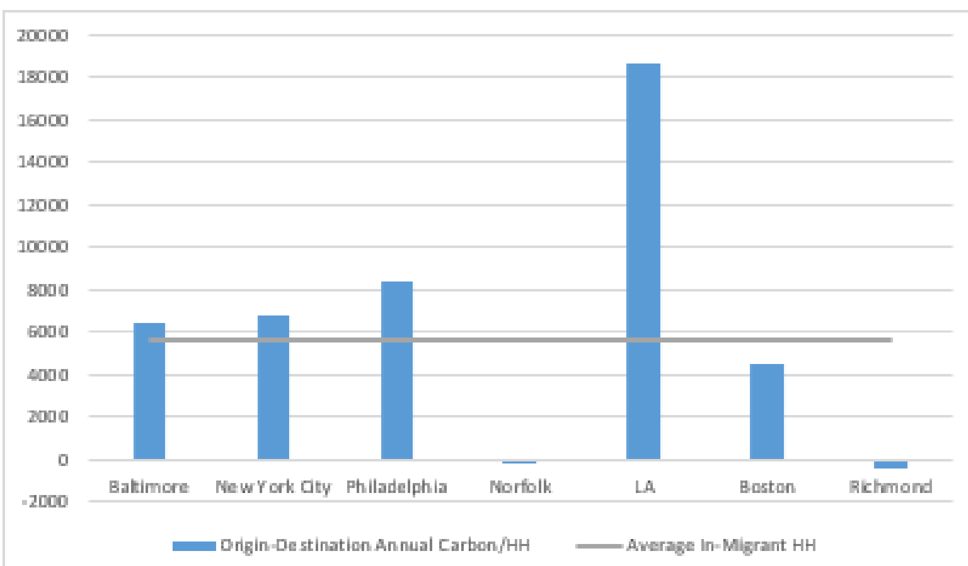


Table 8: Summary of Assigned Carbon Per Household

Population Rank	MSA	Carbon per HH 1992	Carbon per HH 2000	Carbon per HH 2008	Carbon per HH 2008-1992
1	New York, NY	34,097	35,538	30,158	-3,939
16	Minneapolis, MN	38,208	38,930	34,886	-3,322
48	Rochester, NY	33,453	36,486	30,444	-3,010
7	Washington, DC	41,133	42,371	38,376	-2,758
45	Richmond, VA	41,107	43,332	38,578	-2,529
41	Buffalo, NY	34,908	38,538	32,397	-2,511
22	Denver, CO	38,768	37,685	36,313	-2,455
43	Hartford, CT	40,610	41,096	38,308	-2,302
19	Baltimore, MD	34,354	35,983	32,227	-2,126
25	Portland, OR	27,693	27,380	25,706	-1,987
13	Riverside, CA	24,907	23,721	23,150	-1,757
32	Providence, RI	39,772	40,250	38,086	-1,686
28	San Jose, CA	24,802	23,628	23,117	-1,686
12	San Francisco, CA	26,068	25,123	24,405	-1,663
27	Sacramento, CA	27,664	26,954	26,098	-1,565
15	Seattle, WA	28,307	28,130	26,857	-1,449
10	Boston, MA	37,557	37,831	36,172	-1,385
5	Dallas, TX	44,870	45,104	43,557	-1,313
4	Philadelphia, PA	30,237	33,948	29,066	-1,171
17	San Diego, CA	27,559	27,056	26,449	-1,109
2	Los Angeles, CA	24,625	23,672	23,590	-1,035
20	Pittsburgh, PA	35,643	41,052	34,623	-1,020
38	Nashville, TN	47,354	48,283	46,342	-1,012
8	Houston, TX	43,611	44,225	42,705	-906
46	Oklahoma City, OK	52,115	53,785	51,257	-858
33	Norfolk, VA	39,559	42,412	38,759	-800
39	Austin, TX	44,361	45,354	43,617	-744
11	Atlanta, GA	45,992	48,299	45,556	-436
47	Birmingham, AL	43,758	44,738	43,932	175
49	Salt Lake City, UT	33,587	35,132	33,996	409
35	Milwaukee, WI	31,387	32,964	32,052	665
37	Charlotte, NC	44,591	47,916	45,263	672
40	Memphis, TN	43,314	44,874	44,112	798
3	Chicago, IL	32,103	33,240	33,031	928
42	Louisville, KY	44,432	46,570	45,956	1,523
21	Tampa, FL	35,903	38,103	37,470	1,567
29	San Antonio, TX	38,151	40,852	39,994	1,844
31	Columbus, OH	37,131	39,717	38,984	1,853
36	Las Vegas, NV	33,493	35,249	35,465	1,972
6	Miami, FL	31,149	32,810	33,161	2,012
14	Phoenix, AZ	33,507	35,565	35,725	2,218
30	Orlando, FL	36,272	38,656	38,929	2,657
18	St Louis, MO	36,502	39,294	39,487	2,985
9	Detroit, MI	36,265	38,937	39,388	3,123
26	Kansas City, MO	42,434	45,486	46,095	3,661
44	Jacksonville, FL	35,516	38,831	39,529	4,013
34	Indianapolis, IN	41,479	45,138	45,666	4,187
24	Cincinnati, OH	42,737	46,174	47,185	4,448
23	Cleveland, OH	37,987	40,964	42,533	4,545

Figure 12: Washington D.C. In-Representative Carbon Differential, 1992

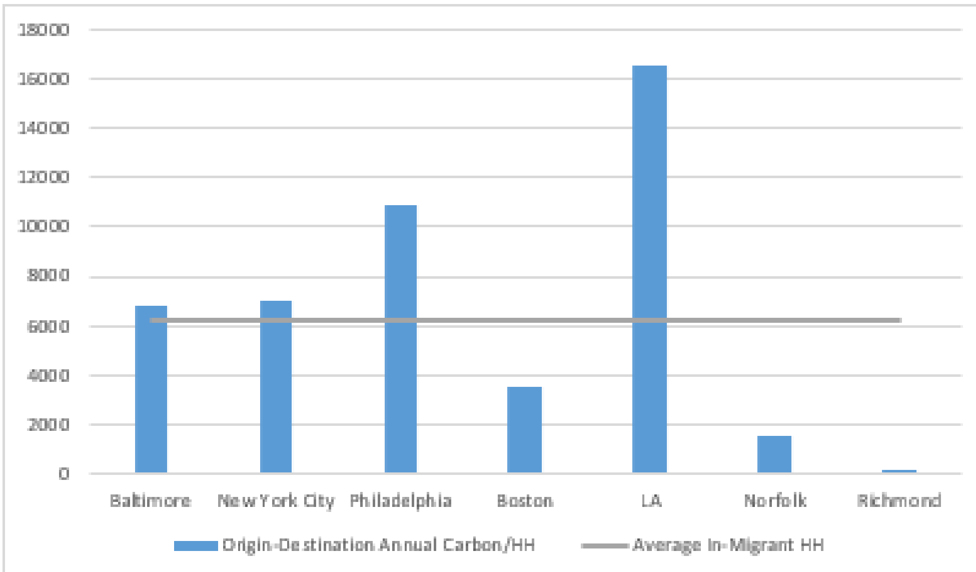


Figure 13: San Antonio In-Representative Carbon Differential, 2000

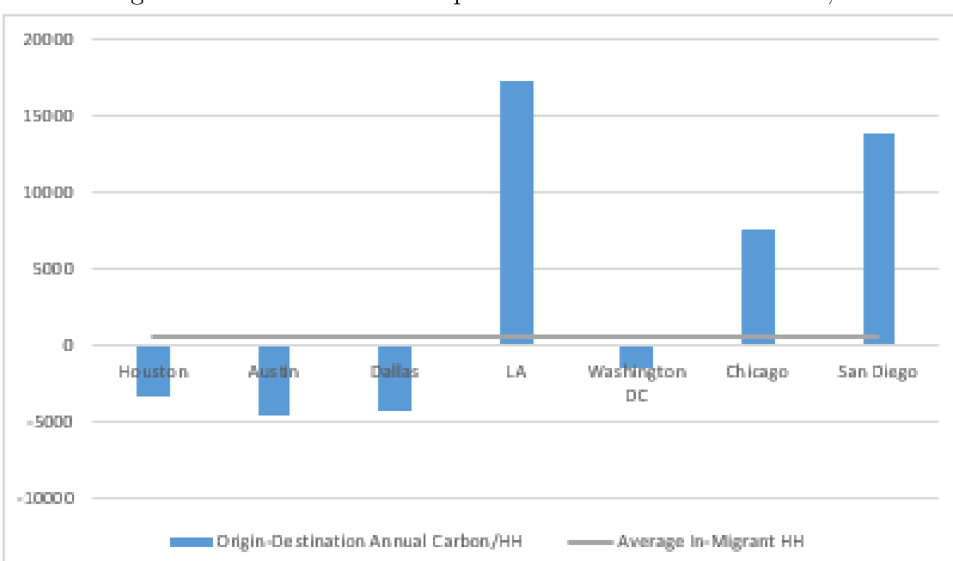


Figure 14: San Antonio In-Representative Carbon Differential, 1992

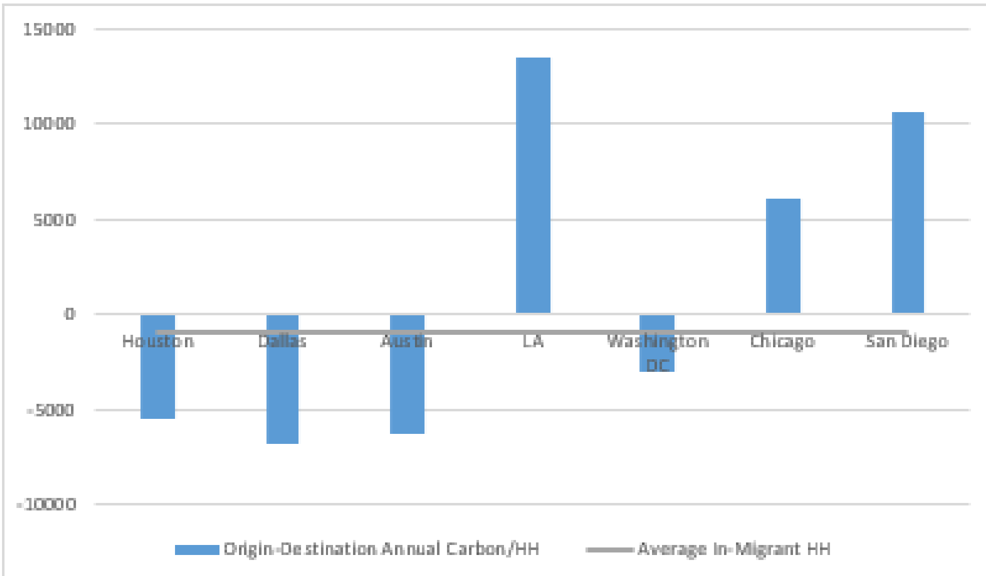


Figure 15: Atlanta In-Representative Carbon Differential, 2000

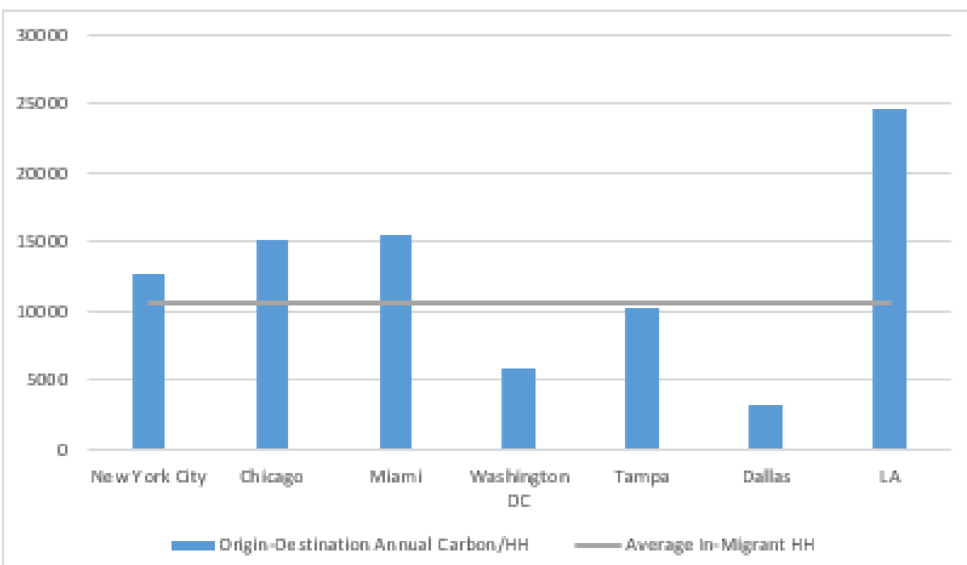
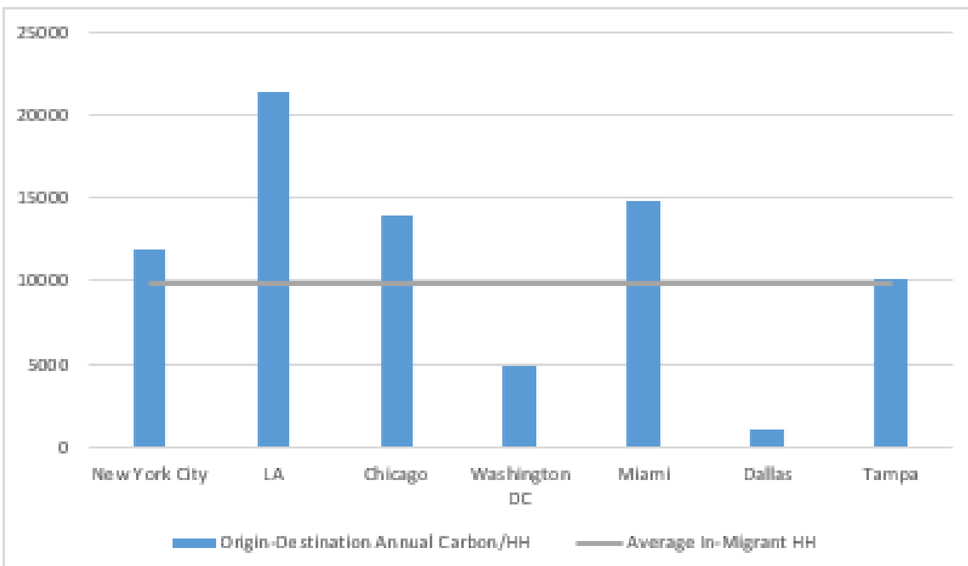


Figure 16: Atlanta In-Representative Carbon Differential, 1992



References

Bertaud, Alain, and Jan K. Brueckner. "Analyzing Building-Height Restrictions: Predicted Impacts and Welfare Costs." *Regional Science and Urban Economics*, vol. 35, no. 2, 2005, pp. 109–125., doi:10.1016/j.regsciurbeco.2004.02.004.

Borck, Rainald, and Jan K. Brueckner. "Optimal Energy Taxation in Cities" *Journal of the Association of Environmental and Resource Economists*, 2018 5:2, 481-516

Glaeser, Edward, and Matthew Kahn. "The Greenness of Cities: Carbon Dioxide Emissions and Urban Development." 2008, doi:10.3386/w14238.

Glaeser, Edward L., and Matthew E. Kahn. "The Greenness of Cities: Carbon Dioxide Emissions and Urban Development." *Journal of Urban Economics*, vol. 67, no. 3, 2010, pp. 404–418., doi:10.1016/j.jue.2009.11.006.

Gyourko, Joseph, et al. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index." *Urban Studies*, vol. 45, no. 3, 2008, pp. 693–729., doi:10.1177/0042098007087341.

Mangum, Kyle. "The Role of Housing in Carbon Emissions." *SSRN Electronic Journal*, 2017, doi:10.2139/ssrn.2957749.

Penn, Ivan. "California Will Require Solar Power For New Homes." *The New York Times* [New York City], May 10 2018: Pages B1

2. Accurate Assignment of School Quality to Properties

2.1 Introduction

One of the important factors a family considers when deciding to purchase a house is the quality of the local schools. These decisions by households can shape not only the composition of students at a school, but also the socio-demographic composition of a neighborhood. A recent study by Owens (2016) has shown that families with children account for the entire increase in income segregation from 1990-2010, showing that households will shape the places they choose to live in. As noted in Kornrich and Furstenberg (2012), there is a large "class gap" in investments in children among families. One such investment parents make in their children is purchasing houses in neighborhoods with high-quality schools, and this bidding-up of good-school housing will cause income segregation and ultimately lead to larger differences by family income of child education attainment. School quality is seen as a very important input from parents to children, and has been shown to be a key component of the intergenerational transmission of advantage as seen in Chetty et. al (2011). School quality has also been shown to have a large impact on earnings outcomes.

The standard procedure to measure the value of good schools is to measure how much more households pay to live in areas with these good schools. Studies of household "demand" for education often rely on property values to infer willingness to pay (Black 1999; Sieg et al. 2004; Figlio and Lucas 2004; Bayer, Ferreira, and McMillan 2007; Clapp, Nanda, and Ross 2008; Fack and Grenet 2010; Nguyen-Hoang and Yinger 2011; Brunner, Cho, and Reback 2012; Andreyeva and Patrick 2017; Mothorpe 2018; Banzhaf and Mangum

2018). However, obtaining data on exactly which school a given property is assigned to is not a simple problem, but one which is a prerequisite to drawing inferences about the importance of school quality. Accurate school assignment data is not widely available for many areas and many years, and therefore various methods are used to make this assignment. The most relevant consideration for the purposes of this paper is the method of matching used in prior work. Card and Krueger (1992) study education differences by making comparisons across states, aggregating many variables to the school district or state level and then assigning these means to properties. As we will see, districts are very different in size, number of schools and students, and other demographics. Sandra Black (1999) was the first major study which attempts to solve the problem of endogeniety of good schools being located in otherwise highly desirable areas. She uses attendance boundary data in Massachusetts and compares houses on either side of these attendance lines. Downes and Zabel (2002) use distance as the crow flies to assign properties to the closest school in the Chicago MSA. Fack and Grenet (2010) utilize attendance boundary data for secondary schools, but only for one european city (Paris).

The purpose of this paper is to compare methods of assigning school quality to a property. These methods include those from prior studies and ones introduced in this paper. Two new data sets, known as SABINS and SABS, allow for a much more accurate match and much larger sample size than any previous studies on school attendance. This data will be used as a comparison to test various matching methods used in prior research, and to determine the methods' accuracy in assigning the proper school quality to a particular property. The extent to which matching by these methods is accurate or inaccurate affects hedonic estimates of school quality on the price of homes. If prior studies used methods of matching which are largely inaccurate, then their estimates and policy implications could be flawed. I also test several new methods to see if they can improve the matching quality. Then I turn to the question of how significant the problem of miss-matching is. Finally, I explore the policy implications of failing to accurately value school quality. This paper contributes to the literature by directly testing methods used in previous research. It is an open question just how accurate some of these methods are, and if their accuracy is the same for different areas of the U.S. By extending these methods for a much larger sample, both by schools and areas of the U.S., this paper can test both the internal validity of the method within its sample and its external validity for other parts of the country. In the future, studies about the value of education will always be concerned with the accuracy of matching properties to schools, so this paper will be able to contribute to future studies as well.

In summary, assignment by distance and by aggregate district level means often leads to matching errors. These errors vary over different areas in the US. They also vary by the size of the district. Assignment of school quality such as student teacher ratio, free and reduced lunch proportion, and test scores show that the methods are mismatching the key components of school quality. The paper proceeds as follows. Section 2 discusses the various methods of school quality assignment and how they relate to prior research. Section 3 describes the data sets used throughout the paper. Section 4 provides analysis on the accuracy of the school quality assignment methods.

Second 5 provides insight on the magnitude of bias mismatching can have on studies which examine school quality. Section 6 concludes and outlines the plan for future research.

2.2 School Quality Assignment Methods

This section describes in detail methods used to assign school quality to a housing bundle. All assignments are done at the census block level so property data can be matched by its census block.

Attendance Boundary Administrative Data

The baseline way in which other matching will be compared is to match properties using GIS files of school attendance boundaries, provided by school administrators. These are attendance boundaries for individual schools within a school district. These data come from SABINS/SABS, and blocks are assigned to the school which corresponds to the attendance boundary they lie in. More details on SABINS and SABS are provided in section 3. Previous studies of local areas have used attendance boundary data and thus a direct matching from schools to blocks or properties. For example, Black (1999) does this for three areas in Massachusetts; Bayer, Ferreira and McMillian (2007) use attendance boundary data for some parts of the San Francisco area. However, until recently, such data was not available for large geographic scopes in the US.

Distance Proximity

One intuitive method for assigning schools to properties is to assign the geographically nearest school. Taking the latitude and longitude of both schools and properties (blocks), a distance is calculated and then the matching is made with the minimum distance(s) from the property (or block) to the school. A constraint can be added that the match must be within the school district, the boundaries of which are more easily observed. This is the method of school quality assignment that was used at the census tract scale by Downes and Zabel (2002).

Some concerns naturally arise when matching by distance. The matching will depend on the exclusion of atypical schools, such as charter schools or private schools. Distance "as the crow flies" may not represent the shortest travel time in some areas, such as those with rivers, lakes, and variance in elevation. School district boundaries are often explicitly set with travel time in mind, and thus any geographic features which cause travel time and straight line distance to differ will be a cause for concern (Hoxby, 2000.) Consequently, the distance matching method might produce more mismatching in areas which have more geographic features, such as rivers and mountains. A measure such as the geographic fragmentation index, which quantifies the likelihood that two people selected at random live at different elevations, could be a covariate of this matching efficiency. The U.S. Geological Survey contains data on small streams which could be another covariate of matching efficiency. Also, distance matching assumes fairly regular and logical planning of school locations and that these are consistent between different districts across the country. It also assumes new schools to have surrounding areas attend them according to distance. For the distance calculations in this paper, census blocks are matched to all of the schools within the school district and state containing the block, and schools are then ranked for each census block by distance. When merged with data from SABINS and SABS, this allows a comparison to be made for the match of nearest school, second nearest school, ect. Knowing this distribution of correct match by distance rank can be useful when the distance-weighted measure is considered.

District Means

Another method used in prior literature is assigning each block to its school *district* rather than its school. Equivalently, it assigns a school quality equal to an aggregated measure of all schools in that district. Examples of this approach include Card and Kruger (1992) and Sieg et al. (2004). This method has the lowest burden of school quality data, needing only district boundaries (which are more fixed than attendance zones) and aggregated school quality measures, which are widely available. However, it can only be used to make comparisons and measure capitalization between districts. School districts can vary drastically in size, population, and number of schools. Other policies and other neighborhood quality will also vary between different districts.

Rank Weighted Quality

The last method is to assign school quality by a convex combination of several nearby schools. Using results from a distance proximity matching method, it is possible to assign rank weighted school quality values to a given block. This is analogous to work commonly done with air quality monitors, using distance to monitors (in our case, distance to schools) to assign weighted pollution (school quality) measures.

In a universe where it is difficult to determine the exact school of attendance for a given block, this method could be a reasonable compromise in the attempt to accurately measure school quality. Moreover, in areas where attendance to the assigned school is not mandatory, this measure could actually be more representative of the choices the household faces. Certain programs, such as school vouchers, offer the household choice over school of attendance, and in this situation we would wish to assign the property a school quality value proportional to the bundle of schools nearby. There are cases in the data where attendance boundaries are not unique for this reason, because or programs involving school choice.

2.3 Data

This section describes the data used in this paper.

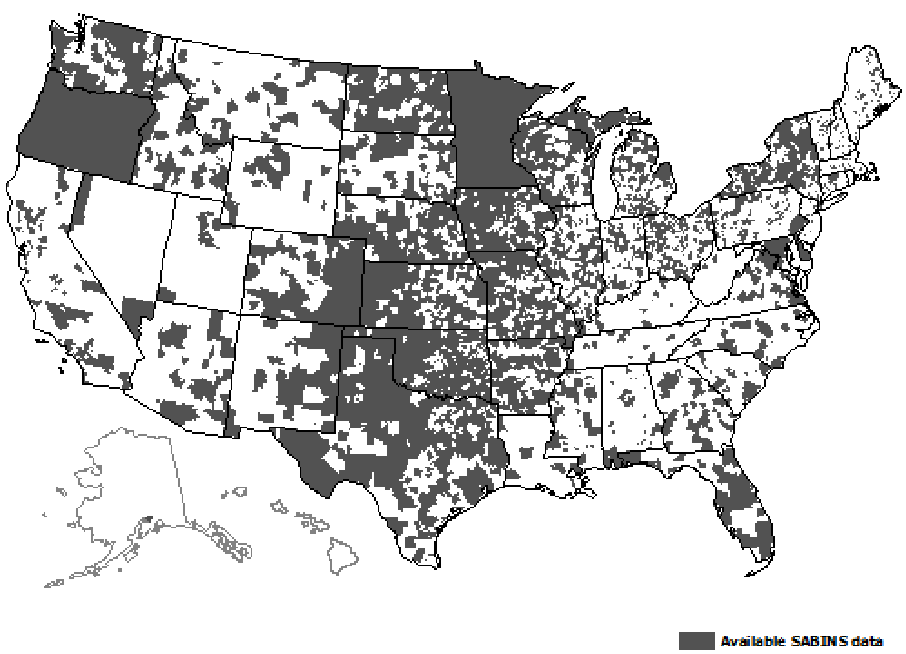
SABINS

The SABINS project (School Attendance Boundary Information System)[1] is the first comprehensive effort to assemble, unify, and disseminate GIS data on school attendance boundaries (The College of William and Mary and the Minnesota Population Center 2011). Funded by the NSF, the data was assembled for school years 2009-2010 until 2011-2012. The project attempts to determine the exact shapes of the school attendance boundaries. For this reason, this data is considered the standard to which the other methods will be tested for this paper. However, SABINS does not cover all areas of the US, and coverage varies significantly by state. SABINS uses administrative

[1] http://www.sabinsdata.org

information and converts this information to GIS maps of attendance boundaries for each school in its dataset (this includes all grades PK-12). These attendance boundaries are much finer than school districts, and show areas where properties are assigned to each school. SABINS is somewhat limited in scope, covering only 3 school years (with almost all of its data in 2009-10 year) and only certain areas of the US. For this research, data was taken from the SABINS website for academic year 2009-2010. This data includes GIS shapefiles for all elementary school attendance boundaries in the SABINS data. Properties are matched to their SABINS attendance boundary by GIS through their census block latitude and longitude.

Figure 17: SABINS Data Coverage



SABS

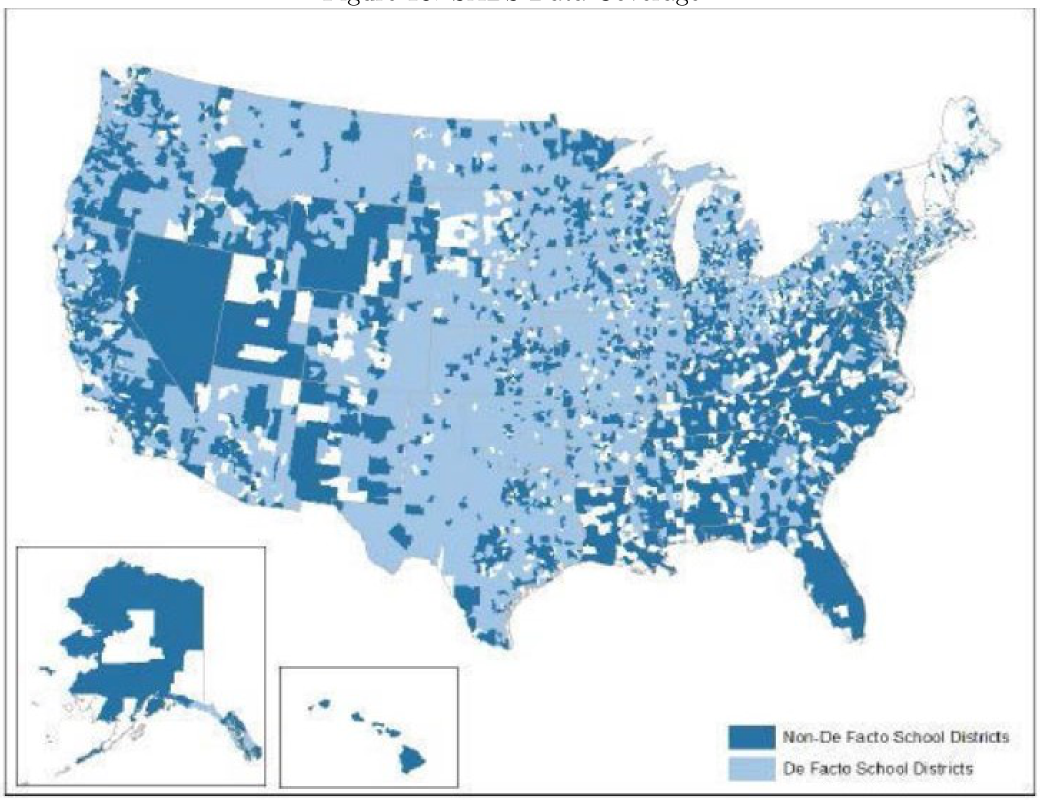
Very similar to SABINS, SABS data[2] is collected by the Department of Education's National Center for Education Statistics (NCES), and contains school attendance boundary information for much of the US at three school levels (primary, middle, high) (Tai Phan 2015). This data is surveyed from school district administrators. School districts have over 90 percent response rate in SABS. Data from 2013-14 school year were collected. This data is similar to SABINS but has better coverage and explicit labels for de facto school districts (see below) and open attendance policies.

The map above shows the coverage of SABS for the 2013-14 school year. These are the school districts which responded and sent data on their attendance boundaries to SABS. The distinction is made between de facto and non-de facto school districts, or districts which contain exactly one school accommodating each grade and districts containing more than one school for at least one grade, respectively. This distinction will be discussed at length in a following section. Below is a detailed look at the data, with schools represented as points and attendance boundaries represented as boxes.

Figure 19 illustrates schools and attendance boundaries. This is part of Los Angeles Unified, one of the largest school districts in the U.S. Many parts contain well behaved bijections of schools and attendance boundaries, with one dot (school) inside one shape (attendance boundary.) Some areas are missing in SABS, so while schools show up, boundaries do not (see Compton, circled, and Inglewood areas.) Some attendance boundaries have multiple

[2] https://nces.ed.gov/programs/edge/SABS

Figure 18: SABS Data Coverage



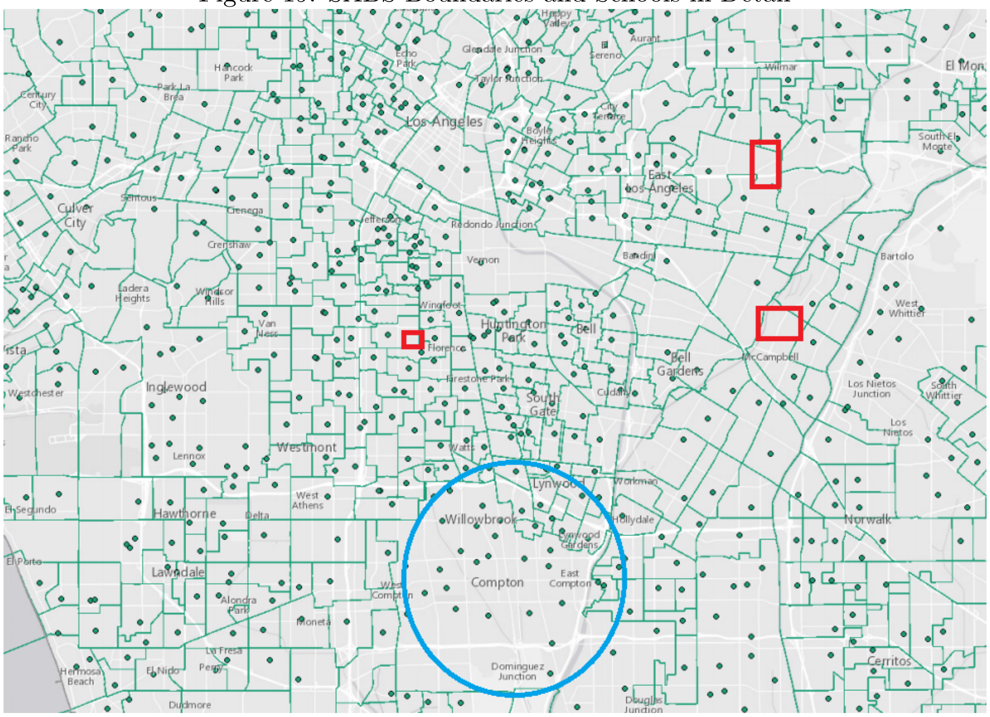
SOURCE: U.S. Department of Education, National Center for Education Statistics, School Attendance Boundary Survey (SABS), 2013-14.

schools, usually indication a charter or other open enrollment school. It is easy to find several areas inside one boundary with closer physical proximity to a school in another attendance boundary (squared areas in the figure), which demonstrates some of the concerns raised in the distance matching section.

School Districts

Data for census blocks and school districts were taken from TIGER/line shapefiles, provided by the US census. These data provide an exhaustive

Figure 19: SABS Boundaries and Schools in Detail



Notes: 1. This figure depicts attendance boundaries (green boxes) and schools (green dots) in Los Angeles 2. Some geographic areas are closer in proximity to a school which they are not assigned to (examples in red boxes.) 3. Some areas in the data are missing attendance boundaries (blue circle.)

map of all blocks and districts, as well as basic information about the areas. Blocks are then mapped to school districts via GIS.

School Quality

School information is taken from the Common Core Data, as well as a U.S. Department of Education test score release for all schools. These data contain the quality variables to be assigned, including pupil-teacher ratio, free and reduced-price lunch ratio, and math and reading test scores. Test scores are converted into Z scores within each state to be comparable between states. Table 9 contains summary statistics for schools and table 10 contains summary statistics for school districts, divided by de facto and non-defacto districts.

School data contains latitude and longitude coordinates, needed for the distance to be calculated. Distances from each census block to every school in the same school district are calculated, and then each block receives a list of its closest schools in distance order.

Table 9: Summary Statistics for School Quality Data, School Level

Variable	n	Mean	S.D.	Min	.25	Mdn	.75	Max
Total Students	21510	515.89	245.78	4.00	348.00	497.50	664.00	2837.00
Free-Reduced Lunch Students	20812	280.01	210.97	0.00	116.00	238.00	399.00	2546.00
Pupil-Teacher Ratio	21510	16.03	4.57	2.34	15.54	15.54	18.09	251.00
Math Pass Rate	21463	74.34	17.86	2.00	63.00	78.00	88.00	99.00
Reading Pass Rate	21465	72.26	18.80	2.50	60.00	77.00	87.00	99.00

Notes: 1. This table contains summary statistics for all schools. 2. Total students, free-reduced lunch students, and pupil-teacher ratio are taken from common core data. 3. Math and reading pass rates are taken from a national data release for standardized tests.

Table 10: SABINS 2009 Summary Statistics, School Districts

Variable	Description	De facto	Non-De facto
Count	Number of Districts	3845	1437
Rural	Defined by CCD	3132(81.5%)	421(30.3%)
Town		533(13.8%)	409(28.5%)
Suburb		177(4.6%)	393(27.3%)
City		3(0.08%)	214(14.9%)
Total Blocks	Mean(SD)	480.72(310.61)	2002.52(3460.36)
Mean Block-School Distance	Mean(SD)	8.05(4.89)	4.706(4.133)
First-Match Prop	Mean(SD)	0.9347(0.247)	.515(.230)
Total Teachers	For all grades	50.95(33.29)	817.48(1895.28)
Total Students	For all grades	691.1(502.3)	13738.4(33267.3)
Total Schools	Including Grade 4	1(0)	8.1161(19.1392)

Notes: 1. This table contains summary statistics for all school districts contained in Sabins 2009. 2. Districts are divided into de facto and non-defacto, de facto districts being those which contain exactly 1 school for elementary level. 3. Urban classification has 4 categories defined by Common Core Data. 4. Block school distance is the distance between a block and the nearest school. 5. First-match prop contains the proportion of blocks correctly matched from distance to Sabins.

2.4 Accuracy

This section describes the accuracy of the various matching methods by comparing them with the SABINS school. First is a test on matching vs

Table 11: Match Rates for Blocks to Schools

	SABINS Blocks	SABINS Coverage Proportion	Matches within 1 School	Matches within 2 Schools	Matches within 3 Schools	Matches within 4 Schools
Total	4732069	0.48	0.71	0.83	0.87	0.89
Low	2367295	0.47	0.63	0.79	0.84	0.87
High	2364774	0.49	0.78	0.89	0.90	0.91

Notes: 1. This table contains the proportion of blocks which are correctly matched to the school from Sabins data. 2. The table is divided into low and high match states. 3. Blocks with a match within the nearest 4 schools are added by column to obtain cumulative match with Sabins school.

mismatching accuracy for distance proximity; the tests for accuracy of school quality matches is done afterwords.

Distance Proximity Matching Accuracy

For all states, the distance proximity match is used and compared against SABINS matches. The top 4 closest schools are matched to each block, and then the total number of correct matches in each distance rank is divided by SABINS blocks to determine match proportion. In table 11, the results are summarized by total matches for all states (Total), as well as for the top half (high) and bottom half (low) of states by SABINS match. This is to contrast the differences in match distribution between the states with the best first-match and the states with the worst first-match.

The second column lists the number of blocks in each group which has a SABINS school match, meaning it is inside an attendance boundary in the

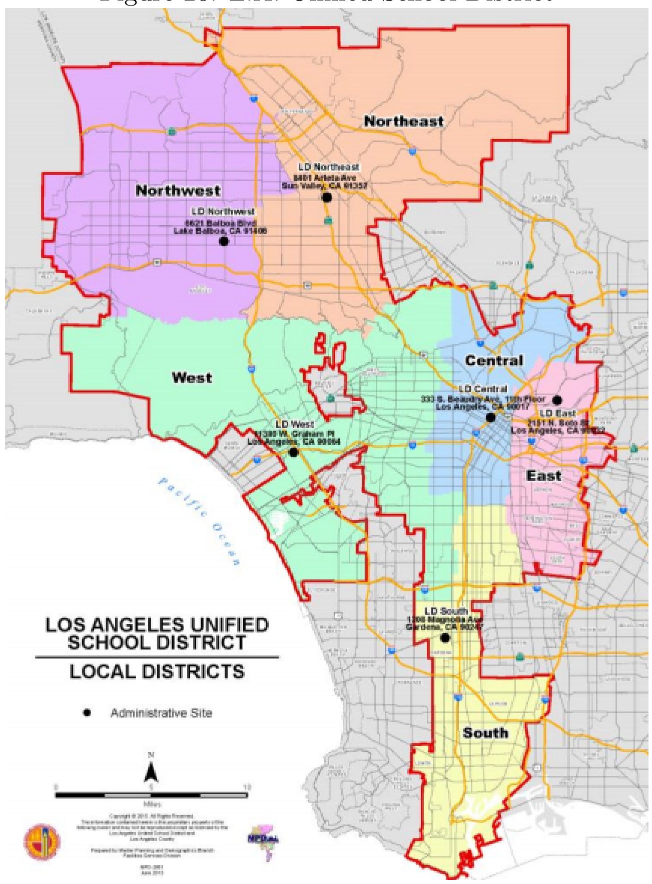
SABINS data. As seen in the third column, the proportion of total blocks covered by SABINS is a little less than half, and not dramatically different between the groups. The next column shows the proportion of blocks whose closet school is the same as it's SABINS school, where from here we are only counting blocks for which there is a SABINS match. The final columns track the distribution of matches, where the SABINS school is the same as the second, third, and fourth nearest school, respectively. The gap between high and low match states starts at 16 percentage points (about 20% of the high match rate) but decreases to only 4 pp (around 4%) at the fourth nearest school. Some states have better matches than others. Why this might be poses an interesting question for investigation. Straight line distance seems to better match blocks to schools for flat states than mountainous ones (KS, MO, NE among the best matched; NC, TN, CO among the worst.) Topography, as mentioned before, might play an important part in the usefulness of a distance match. Other proposed covariates of match are population growth over the last 10 years (tract level), date of school openings in district, and number of schools in the district. Each of these can relate to the number of schools in a district, the regularity of attendance boundary shapes and the number of changes to attendance boundaries.

One item to note is that SABINS coverage is not a strong predictor of first-school match. The proportion of blocks in each state which are included in the SABINS data is compared to the accuracy of the match in that state. For example, states in the low-match category have 47% of their blocks in SABINS data, compared to high-match states with 49%. The correlation between a state's SABINS coverage and that state's accurately matched blocks is almost exactly zero. It would be a cause for concern of external validity if SABINS appeared to be cherry-picking the easiest places to create their data, while omitting places which were difficult to map. This test seems to suggest that the states for which we have a lot of data from SABINS are not significantly different from those we have little data from SABINS, in terms of the accuracy of match to first school.

De Facto vs. Non-De Facto Districts

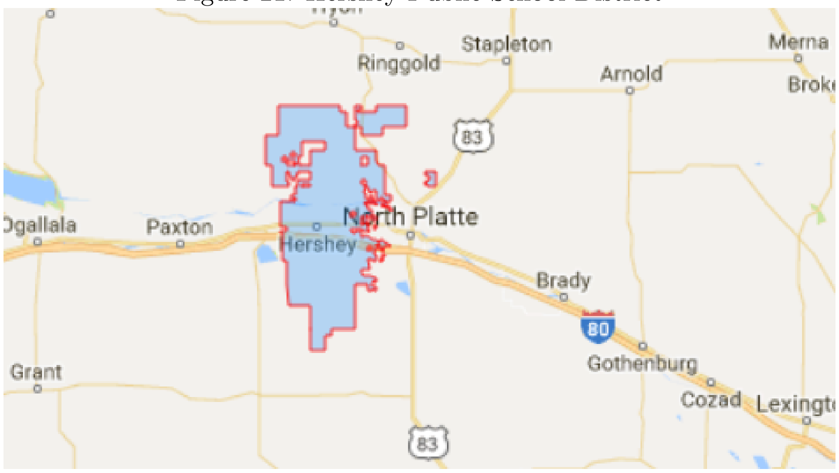
As mentioned previously, a de facto school district is one where the school attendance boundary is the same as the district boundary, because the district contains exactly one school of that type. Many districts are set up this way in the U.S., and the accuracy of distance matching is almost entirely dependent on a district being de facto or not. These schools/districts can be useful because it is very easy, even without attendance boundary data, to match properties to the correct school either by distance or through district-level data. De facto districts are typically much smaller in number of students and teachers, smaller in distance, and rural in urban code. To take two examples, consider first Los Angeles Unified School District, one of the largest school districts in the US.

Figure 20: L.A. Unified School District



L.A. Unified has an area of over 720 square miles. It enrolls over 640,000 students with over 900 public schools and over 180 charter schools. The area it covers has a population of over 4.5m in over 1.5m households. In contrast, consider Hershey Public School District in Nebraska.

Figure 21: Hershey Public School District



Hershey has an area of 351 square miles. It enrolls fewer than 500 students, and has just 3 public schools (one for each school level), making it a de facto district. It has a population around 2,200 with around 860 households. Clearly school districts in urban California and rural Nebraska will be very different. When using aggregate data at the district level, it's not clear why school districts should be the appropriate units of comparison. While families might compare at the district level when choosing between de facto districts, the comparison would be at a much finer level within the very large districts.

Table 10 shows summary statistics for the two types of district for SABINS 2009-10. The total number of districts are listed for each type, with de facto districts being much more numerous though smaller in area and students than non-de facto districts. Urban classification, as defined by the ccd, shows the proportion of districts in the rural/town/suburb/city categories. Nearly all of the de facto districts are rural or town classification, while non-defacto districts are much more evenly spread over the categories. On average, de facto districts contain fewer than 500 census blocks, while non-de facto districts contain over 2000 blocks. Block to school distance and the proportion of correct matches to the closest school are higher for de facto districts. Total teachers, and total students are much higher for non-defacto districts.

To explore patterns in where and when distance proximity will be a suitable proxy for the true school, I regress the match rate on geographic features, including dummies for rural and de facto district. Rural and de facto are separated into the possible combinations, since many rural distracts are also de facto. Table 12 contains the results. Other variables besides de facto have almost no explanatory power, though they significant, once de facto is accounted for. Note, in particular, that the R^2 drops from 0.39 to 0.06 once de facto districts are dropped. Variables include average distance in the district from blocks to closest school in miles, a rural dummy, number of students(thousand), 10 year population growth for the county that district is in (hundred thousand), and number of schools in the district.

Table 12: Regression Results for District Block Match

	Full Sample	Excluding De Facto
Mean School Block Distance	-0.021***	-0.007***
	(0.00073)	(0.00161)
Rural = 1, Defacto = 0	0.365***	0.050***
	(0.01243)	(0.01418)
Rural = 0, Defacto = 1	0.035**	
	(0.01438)	
Rural = 1, Defacto = 1	0.483***	
	(0.01007)	
Number of Students (Thousand)	-0.003***	-0.002**
	(0.00089)	(0.00084)
County Population Growth (Hundred Thousand)	0.031***	0.029***
	(0.00533)	(0.00517)
Number of Schools in District	0.004***	0.003***
	(0.00097)	(0.00092)
Constant	0.486***	0.506***
	(0.00928)	(0.01075)
Adj. R-Squared	0.3941	0.0625
Observations	5247	1435

Notes: 1. All regressions are split into a full sample (all school districts) and excluding de facto schools. 2. Regressions are at the school district level. 3. Mean school block distance is the average distance from each block to the nearest school. 4. Rural label is defined by the common core. Categories are made for rural non-de facto, non-rural de facto, and rural de facto. 5. County population growth is measured as the percent growth from 2000-2010.

School Quality Assignment Accuracy

Given the potential to match blocks to the wrong school, the next question to consider is whether these mismatches are important when assigning school quality. In terms of the relevant characteristics of the school, do the assignment methods assign significantly different values to properties? To answer, school quality measures are assigned to each census block covered by SABINS data.

Two samples are created for correlation of school quality. The first sample contains only blocks that have multiple schools assigned to them. This excludes blocks in de facto districts, which will have identical assignment from all sources with data on school quality. This leaves those blocks with the issue of school assignment only. The full sample contains all eligible blocks. The first sample will include de facto districts, which means many assignment measures are identical for these observations. Correlation tables for the first sample are presented with the results, and correlation tables for the second sample are presented in the appendix. The first sample contains 2,322,582 blocks; the second sample contains 4,357,816 blocks.

The baseline for comparison is the quality assignment from SABINS. The assignment methods compared are nearest distance matching, second nearest distance matching, distance rank matching, and district means assignment. The school quality measures assigned to blocks are student-faculty ratio, free and reduced lunch proportion of students, 4th grade math test scores and 4th grade reading test scores. The test score data uses a Z score for pass rate relative to other schools in the state. The correlation of interest is each assignment method's quality with SABINS quality. This is presented in column 1 of the correlation tables.

Overall, the rank weight distance match has the highest correlation across all measures of school quality in the first sample. Nearest distance match performs only slightly worse; district means lags significantly behind these two across all variables. Second nearest distance match has the lowest correlation, and is included mostly as a comparison to nearest distance for the importance of matching accuracy.

Student-Faculty Ratio

This variable measures the ratio of enrolled students of all grades to full time equivalent teachers, as defined by the common core. For blocks in the sample of interest, the nearest school has a correlation of 0.69 with SABINS assignment. Second nearest school has a poor correlation by itself (0.2001) but the rank-weight assignment improves the correlation from the nearest distance match, although only slightly. District mean assignment has a 0.44 correlation with SABINS quality. Results for the correlation for student-faculty ratio assignment are seen in Table 13.

Table 13: Correlation for Student-Faculty Ratio, Sample 1

Variable	1.	2.	3.	4.	5.
1. SABINS School	-				
2. Nearest Distance Match	0.6929	-			
3. Second-Nearest Distance Match	0.2001	0.1939	-		
4. Rank Weight Assignment	0.6964	0.9604	0.3338	-	
5. District Mean Assignment	0.4403	0.4687	0.4077	0.5029	-

Notes: 1. This table contains correlations for the assigned Student-Faculty Ratio. Each block has its quality assigned by all methods and then the assigned values tested for correlation. Sabins is the reference for proper school quality. 2. Second-Nearest match is included for reference, though it is not used outside of the rank weight assignment.

Free and Reduced Price Lunch Proportion

This variable measures the proportion of students enrolled eligible for free and reduced price lunch. This variable has higher correlations across

Table 14: Correlation for Free and Reduced Lunch Proportion, Sample 1

Variable	1.	2.	3.	4.	5.
1. SABINS School	-				
2. Nearest Distance Match	0.9113	-			
3. Second-Nearest Distance Match	0.6140	0.5935	-		
4. Rank Weight Assignment	0.9213	0.9768	0.6889	-	
5. District Mean Assignment	0.6987	0.6996	0.5479	0.7363	-

Notes: 1. This table contains correlations for the assigned free and reduced lunch proportion of students. Each block has its quality assigned by all methods and then the assigned values tested for correlation. Sabins is the reference for proper school quality. 2. Second-Nearest match is included for reference, though it is not used outside of the rank weight assignment.

all assignment methods with SABINS, especially with nearest distance and rank weight distance matching, with 0.91 and 0.92 correlation respectively. Rank weight again provides a slight improvement to distance matching, and the high correlations are likely due to the fact that free and reduced price lunch are measuring family income, which is highly spatially concentrated. All results for correlations of free and reduced price lunch proportion are found in Table 14.

Math and Reading Test Scores

These variables are taken from a national release for standardized tests. Pass rates for the tests are normalized for the average pass rates from all schools in the state. The two scores are similar in their correlation of assign-

Table 15: Correlation for Math Scores, Sample 1

Variable	1.	2.	3.	4.	5.
1. SABINS School	-				
2. Nearest Distance Match	0.8051	-			
3. Second-Nearest Distance Match	0.6027	0.5204	-		
4. Rank Weight Assignment	0.8328	0.9508	0.7479	-	
5. District Mean Assignment	0.5817	0.5642	0.5402	0.6277	-

Notes: 1. This table contains correlations for the assigned math scores. Each block has its quality assigned by all methods and then the assigned values tested for correlation. Sabins is the reference for proper school quality. 2. Math scores are standardized to a Z score within each state so they can be compared between states. 3. Second-Nearest match is included for reference, though it is not used outside of the rank weight assignment.

ment, with reading being slightly more correlated across all methods. Rank weight improves on nearest distance by nearly 0.03, and again outperforms the other assignment methods. District mean assignment has a correlation of less than 0.6 for both test scores. All results for math test scores are presented in table 15, and reading test scores in table 16.

Table 16: Correlation for Reading Scores, Sample 1

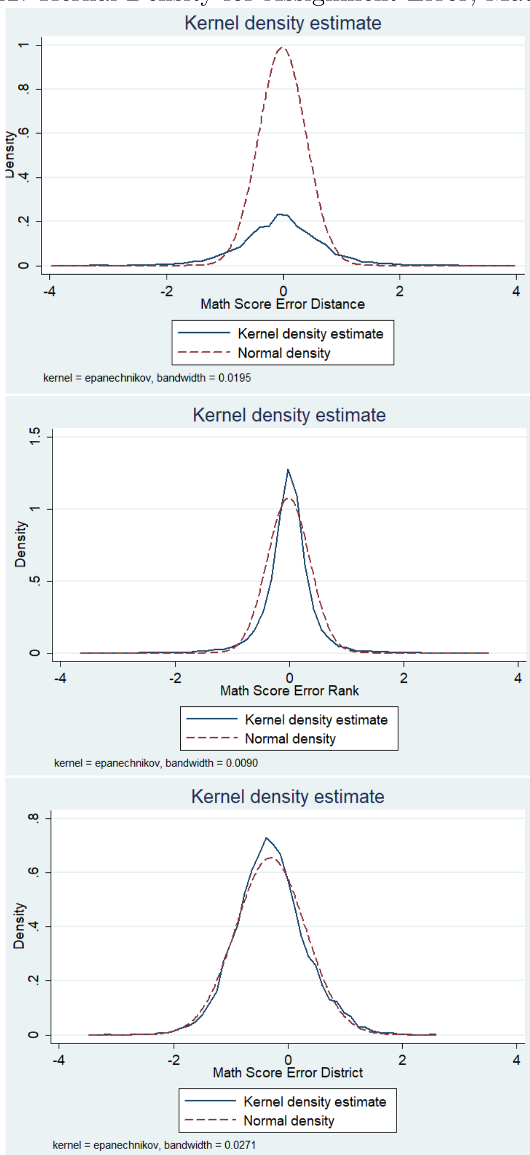
Variable	1.	2.	3.	4.	5.
1. SABINS School	-				
2. Nearest Distance Match	0.8237	-			
3. Second-Nearest Distance Match	0.6452	0.5721	-		
4. Rank Weight Assignment	0.8504	0.9558	0.7788	-	
5. District Mean Assignment	0.5960	0.5778	0.5575	0.6359	-

Notes: 1. This table contains correlations for the assigned reading scores. Each block has its quality assigned by all methods and then the assigned values tested for correlation. Sabins is the reference for proper school quality. 2. Reading scores are standardized to a Z score within each state so they can be compared between states. 3. Second-Nearest match is included for reference, though it is not used outside of the rank weight assignment.

School Quality Assignment Errors

When assigning school quality, assignment be distance, rank, and district all contain errors. The distribution of these errors is important for the validity of estimates based on such assignment. For all assignment methods and school quality variables, kernal density graphs are generated to show the distribution of assignment error. For assignment by distance (closest school), errors are not distributed normally. For assignment by rank, errors become closer to a normal distribution. Finally, assignment by district means has errors which are normally distributed. Figure 22 shows the kernal density graphs for math score assignment errors for distance, rank, and district assignment. Other school quality variables are shown in the appendix.

Figure 22: Kernal Density for Assignment Error, Math Scores



Notes: 1. Graph 1 shows the assignment error for closest school assignment, graph 2 for distance-rank assignment, and graph 3 for district mean assignment. 2. All graphs show math test score assignment. 3. All errors are calculated as assigned quality minus Sabins.

2.5 Conclusion

Matching properties to schools presents a unique and interesting data problem. Through the use of a near-ideal but geographically limited data set, SABINS, it is possible to test various methods to see how accurate they are. Distance Proximity is a fairly accurate approximation overall, but becomes inaccurate in the districts with difficult assignment (non-de facto). The accuracy of distance proximity matching depends on local factors such as topography, population growth, and new school openings. For school quality measures, rank weighted assignment improves on the nearest school assignment; District means assignment performs the worst among assignment methods.

In the absence of strong data such as SABINS/SABS, school assignment has been done using various matching methods. For the sample of all districts, closest proximity and rank weighted assignment do well at assigning the proper school quality, while district means do much worse. In the sample of districts with multiple schools, assignment is less accurate, with rank weighted and distance still outperforming district means assignment.

Because such school quality assignments are less than perfect, estimates of the value of school quality which depend on quality assignment are also less than perfect. Districts and states vary in their characteristics which strengthen or worsen these assignments. For large, non-de facto districts, reliable attendance boundary data becomes important for reliable quality assignment and value estimation.

Appendix

Table 17: Correlation for Student-Faculty Ratio, Sample 2

Variable	1.	2.	3.	4.
1. SABINS School	-			
2. Nearest Distance Match	0.8893	-		
3. Rank Weight Assignment	0.8938	0.9867	-	
4. District Mean Assignment	0.8339	0.8546	0.8739	-

Table 18: Correlation for Free and Reduced Lunch Proportion, Sample 2

Variable	1.	2.	3.	4.
1. SABINS School	-			
2. Nearest Distance Match	0.9379	-		
3. Rank Weight Assignment	0.9454	0.9839	-	
4. District Mean Assignment	0.8008	0.8011	0.8295	-

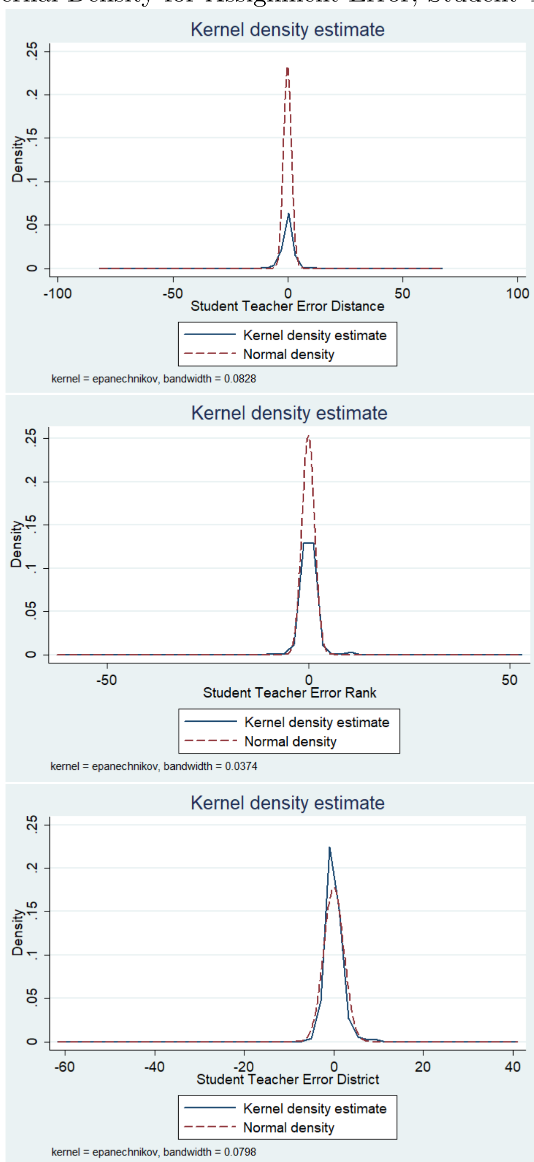
Table 19: Correlation for Math Scores, Sample 2

Variable	1.	2.	3.	4.
1. SABINS School	-			
2. Nearest Distance Match	0.8666	-		
3. Rank Weight Assignment	0.8846	0.9714	-	
4. District Mean Assignment	0.6342	0.6249	0.6629	-

Table 20: Correlation for Reading Scores, Sample 2

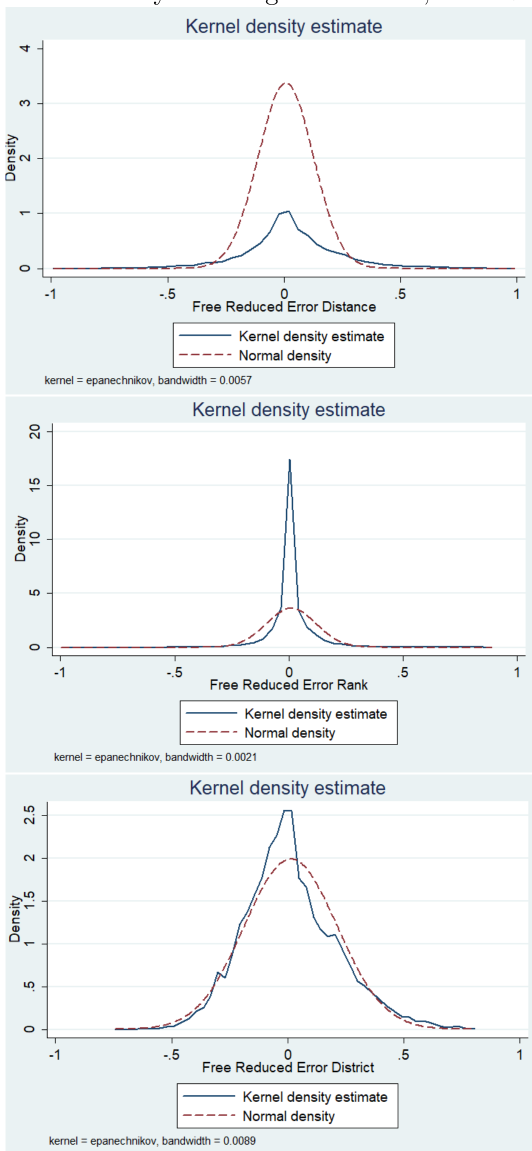
Variable	1.	2.	3.	4.
1. SABINS School	-			
2. Nearest Distance Match	0.8774	-		
3. Rank Weight Assignment	0.8958	0.9724	-	
4. District Mean Assignment	0.6447	0.6329	0.6707	-

Figure 23: Kernal Density for Assignment Error, Student-Teacher Ratio



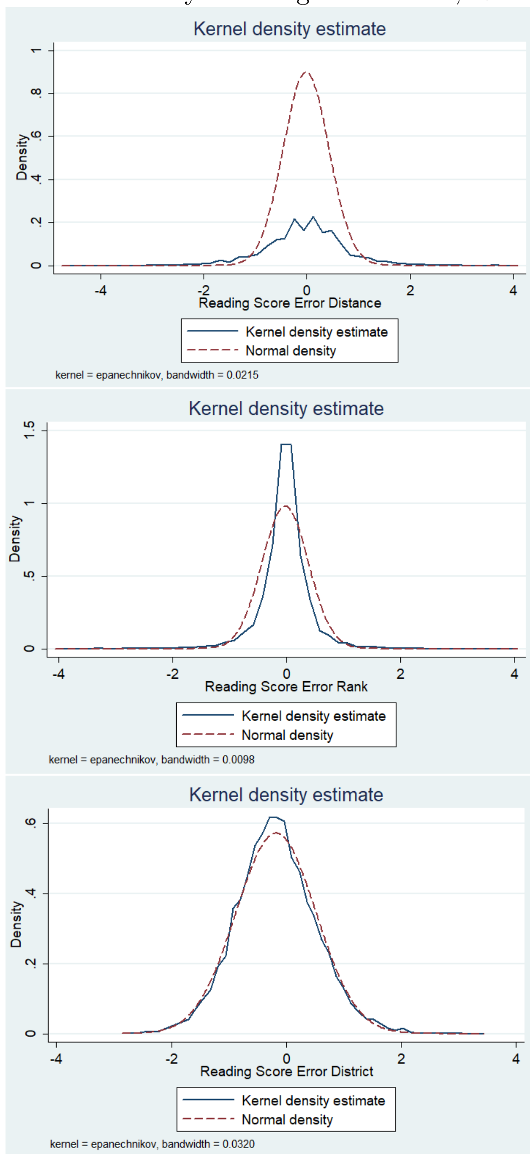
Notes: 1. Graph 1 shows the assignment error for closest school assignment, graph 2 for distance-rank assignment, and graph 3 for district mean assignment. 2. All graphs show student teacher ratio assignment. 3. All errors are calculated as assigned quality minus Sabins.

Figure 24: Kernal Density for Assignment Error, Free-Reduced Lunch



Notes: 1. Graph 1 shows the assignment error for closest school assignment, graph 2 for distance-rank assignment, and graph 3 for district mean assignment. 2. All graphs show free-reduced lunch assignment. 3. All errors are calculated as assigned quality minus Sabins.

Figure 25: Kernal Density for Assignment Error, Reading Scores



Notes: 1. Graph 1 shows the assignment error for closest school assignment, graph 2 for distance-rank assignment, and graph 3 for district mean assignment. 2. All graphs show reading test score assignment. 3. All errors are calculated as assigned quality minus Sabins.

References

Andreyeva, Elena and Patrick, Carlianne, 2017. "Paying for Priority in School Choice: Capitalization Effects of Charter School Admission Zones." *Journal of Urban Economics* 100: 19-32.

Banzhaf, H. Spencer, and Kyle Mangum. 2018. "Capitalization as a Two-Part Tariff: The Role of Zoning." Working Paper

Banzhaf, H. Spencer, and Randall P. Walsh. 2008. "Do People Vote with Their Feet? An Empirical Test of Environmental Gentrification." *American Economic Review* 98(3): 843-63

Bayer, Patrick, Ferreira, Fernando and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115(4): 588-638.

Betts, Julian R. 1995. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *The Review of Economics and Statistics* 77.2: 231

Black, S. E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114.2: 577-99.

Brunner, E. J., Cho, S., and Reback, R. 2012. "Mobility, housing markets, and schools: Estimating the effects of inter-district choice programs." *Journal of Public Economics*, 96(7-8), 604-614.

Card, David, and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*, 100.1: 1-40.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. "The longterm impacts of teachers: Teacher value-added and student outcomes in adulthood." No. w17699. National Bureau of Economic Research

Clapp, John M., Nanda, Anupam and Ross, Stephen L., 2008. "Which school attributes matter? The influence of school district performance and demographic composition on property values." *Journal of Urban Economics*, 63(2): 451-466

Downes, Thomas A., and Jeffrey E. Zabel. 2002."The Impact of School Characteristics on House Prices: Chicago 1987–1991." *Journal of Urban Economics* 52.1: 1-25.

Fack, G., and Grenet, J. 2010. "When do better schools raise housing prices? Evidence from Paris public and private schools." *Journal of Public Economics*, 94(1-2), 59-77.

Figlio, David, N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review*, 94(3): 591-604.

Kornrich, Sabino, and Frank Furstenberg. 2012. "Investing in Children: Changes in Parental Spending on Children, 1972-2007." *Demography* 50.1: 1-23.

Hoxby, Caroline M. 2000. "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90: 1209-238.

Mothorpe, Christopher. 2018. "The impact of uncertainty on school quality capitalization using the border method." *Regional Science and Urban Economics* 70: 127-141.

Nguyen-Hoang, P., and Yinger, J. 2011. "The Capitalization of School

Quality into House Values: A Review." SSRN Electronic Journal.

Owens, A. 2016. "Inequality in Childrens Contexts: Income Segregation of Households with and without Children." *American Sociological Review* 81.3

Phan, T. 2015. "Documentation for the School Attendance Boundary Survey: School Year 2013-14 (NCES 2015-118)" U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Reardon, Sean F., and Ann Owens. 2014. "60 Years After Brown : Trends and Consequences of School Segregation." *Annual Review of Sociology* 40: 199-218.

Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randall Walsh. 2004. "Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods." *International Economic Review* 45: 1047-77.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64: 416-24.

The College of William and Mary and the Minnesota Population Center. 2011. School Attendance Boundary Information System (SABINS): Version 1.0. Minneapolis, MN: University of Minnesota

Vita

Personal

Cody Karl Reinhardt

Born December 23, 1989 in Pensacola, FL

Address: 4600 Creighton Rd

Pensacola, FL 32504

Education

Ph.D Economics, Georgia State University (2019)

B.S. Mathematics and Economics, University of West Florida (2012)

Research

"U.S. Internal Migration Networks, Energy Use, and Emissions" Working Paper

"Accuracte Assignments of Properties to School of Attendance" Working Paper

"A New Look at Urban Centrality" Working Paper, with Alicia Morgan Plemmons

Work

Visiting Instructor, University of West Florida, 2018-