Using urban migration flows for non-market amenity valuation*

Adam Theising Daniel J. Phaneuf

September 2020

Abstract

Economists often use a household's residential location as a revealed preference for place. This paper studies what a household's previous location — their residence during the migration decision — can add to our understanding of tastes for regionally-varying environmental amenities. We show evidence of heterogeneous migration propensities across space, and find a correlation between the "stickiness" of a place and its quality of life. This motivates our development and estimation of a generalized national-level sorting model that accommodates heterogeneity in migration costs across origins. Our demand framework uses structure similar to gravity models of migration. We leverage this structure to identify our key parameters from variation in spatial differences of migration flows across origins and destinations. As a result, this model exists in a single temporal cross-section; notably, our flow approach produces credible estimates without relying on temporal variation. In our empirical application, we estimate our model on a national sample of US households who sort amongst metropolitan statistical areas, and report marginal willingness to pay values for climate amenities and air quality.

Keywords: migration, residential sorting, nonmarket valuation, air quality, climate, gravity
JEL Codes: Q51, R23, R21, Q53

*Authors at the University of Wisconsin-Madison in the Department of Ag & Applied Economics. We thank seminar and conference participants at Hamburg, Kiel, Leipzig, UW-Madison, 2019 Heartland ERE Conference, 2019 Workshop on Nonmarket Valuation, 2020 LSE-Imperial Workshop in Environmental Economics, and 2020 EAERE Conference for valuable feedback. We also thank Josh Deutschmann, Tim Hamilton, Laura Schechter, and Chris Timmins for helpful comments. This material is based on work supported in part by the National Institute of Food and Agriculture, United States Department of Agriculture, AFRI project 1002303. Correspondence to: theising@wisc.edu.

1 Introduction

Do migration flows tell a richer story about amenity preferences than population stocks? Economists often use a household's residential location as a revealed preference for place. If a large stock of people reside in a location with low wages and high housing costs, local amenities must be plentiful or migration costs high. We study how these migration costs vary over space, and ask whether information on a household's previous location can add to our understanding of tastes for regionally-varying environmental amenities.

To answer this question, we construct a national equilibrium sorting model. This type of model has become a popular methodological alternative to the national hedonic framework (Rosen, 1979; Roback, 1982; Albouy et al, 2016). In the conventional hedonic approach, household preferences are implicitly captured by equilibrium wages and housing prices. In the sorting framework, households' trade-offs between wages, housing, and amenities are made explicit: households select their residential location from a discrete menu of metropolitan statistical areas (MSAs) based on the locale's characteristics and their MSA-specific budget constraint. Each location decision reveals household preferences, from which we can infer marginal willingness to pay (MWTP) for an amenity based on Tiebout's (1956) logic. By modeling the household location choice directly, this approach can relax assumptions of free mobility, thereby allowing for the incomplete capitalization of amenities into wages and rents.

We build from this standard sorting framework, but innovate by modeling the role that a household's pre-choice MSA location (henceforth, a household's *origin*) plays in its migration decision. This distinction proves economically intuitive in a world where migration costs may vary by origin: some *places* are harder to leave than others. This notion underlies well-known federal relocation programs like Moving to Opportunity, but certainly applies at wider geographic scales. Recent labor and urban economics research explaining differences in domestic US migration patterns has underlined these systematically-varying rates of "stickiness" or "churn" across cities. We highlight evidence of these spatially-heterogeneous migration frictions in our data and argue they may be important when inferring non-market amenity values.

More broadly, our modification of the standard sorting model yields an estimation structure similar to that of a gravity model, which we emphasize as a convenient link between demand estimation and classical economic models of migration. This structure is econometrically substantive: for a given destination MSA, we observe meaningful variation in the size of incoming migration flows from different origin MSAs. Our reliance on this alternative source of preference revelation — in essence, a repeated cross-section of observed population shares selecting a given destination — provides novel identification of our structural parameters of interest. In the era of big, real-time data, this turn to spatial differences in lieu of temporal differences is meant to be provocative: can we credibly measure MWTP using a single temporal cross-section?

Through the proposed reliance on origin and destination for preference revelation, we revisit the estimation of MWTP for two well-studied regional environmental amenities: climate and air quality. In spatial equilibrium, under the common assumption of linear-in-variables utility, we derive MWTP for an amenity as a familiar ratio of parameterized marginal utilities. We take the model to data, relying on a publicly-available, cross-sectional database constructed from 2012-2016 Census migration flow estimates, American Community Survey (ACS) microdata, and various MSA-level measures of natural and anthropogenic amenities.

As an illustrative example of our model's divergence from the conventional national sorting literature, our empirical analysis begins with a parsimonious representative agent model that only uses aggregated migration flows and location characteristics. Leveraging econometric tools from the trade and migration literatures, our gravity-like estimates produce two indices of structural interest: MSAs' propensity for migration ("push" measures) and overall attractiveness or quality of life ("pull" measures). We highlight these rankings, and document substantial heterogeneity in migration propensities across space. Smaller Rust Belt cities and retiree-friendly cities in Florida have the stickiest populations, while college towns, military bases, and smaller cities in the Western states have the most annual population churn. Our quality of life index proves sensible and foreshadows our main valuation findings: many Sun Belt, non-coastal cities rank amongst the most attractive, while smaller cities in the Rust Belt and Appalachia fall among the least attractive.

Next, we proceed to our main empirical exercise, extending this aggregate model to the household-level by incorporating a 5-year sample of ACS microdata to model individual households' location decisions. We link the gravity estimation structure of our aggregate analysis to the two-stage, microdata-driven approaches used in previous random-utility-based national sorting models. This link allows us to bifurcate households' deterministic utility into a mean origin-destination specific component and a household-specific component. As a result, our household-level model captures rich economic heterogeneity: household income and housing expenditures are modeled destination-by-destination, while tastes for amenities and the household's costs of migration vary based on observable demographics.

We estimate first-stage, household-level logit parameters by maximum likelihood, nesting a contraction mapping in the procedure to match moments: the micro-model's predicted migration shares across space are equalized to the observed shares in the Census Bureau's aggregated MSA to MSA migration flow matrix. In the second stage of the model, we run instrumental variable regressions of average origin-destination indirect utility on destination amenities, while controlling for origin-conditions with fixed effects. This two stage approach allows us to account for both spatially-heterogeneous migration frictions and endogenous amenities in a tractable fashion.

In our preferred IV results, we find that a household earning close to the median national income level has an average annual MWTP of \$1,615 for a 1μg reduction in daily PM2.5 exposure, \$2,507 for a one degree (Celsius) increase in average winter temperature, and \$1,287 for a one degree decrease in average summer temperature.[1] Notably, this specification models utility from income using a semi-parametric linear spline; our results suggest the marginal utility of income diminishes more slowly

[1] In our estimation's microdata sample of non-rural households, these are respectively: 2.8%, 4.4%, and 1.3% of median annual income net of housing expenses. Among US MSAs, the sample standard deviations of these environmental amenities are roughly 2μg (PM2.5), 6°C (winter temperature), and 3°C (summer temperature). See Table 1 and Figure 4 for more detailed information on this variation.

than a conventional log specification would imply. We also study preference heterogeneity in several dimensions. Younger and more educated households exhibit relatively stronger tastes for climate amenities. Preferences also differ by origin MSA: results suggest households from the most temperate origins dislike summer heat, while those those from cities in the middle of the air quality distribution are most averse to particulate pollution.

Contribution to literature

This paper's approach to nonmarket valuation augments practitioners' methodological toolkit when data on residential migration flows are available. Thus, our principal contribution is a fresh marriage of revealed preference logic to a rich literature on migration that studies population flows across space.[2] Our model builds on static spatial equilibrium notions, but relaxes several key assumptions inherent to the continuous hedonic approach (Cropper et al, 1993; Sinha et al, 2019). The cost of doing so is the imposition of more conspicuous structure on household preferences.

Our work adopts much from national sorting models that value regionally-varying environmental amenities. Cragg and Kahn (1997) use US census data to model households' choice of state; they then infer MWTP for climate amenities, noting that their estimates are substantially larger than those from hedonic models. In now-seminal work, Bayer et al (2009) show theoretically and empirically that moving costs help explain the wedge between the true MWTP for MSA-level air quality and the hedonic value capitalized into national housing prices and wages. Their identification strategy relies on long-run first-differences and an IV for air pollution. Sinha et al (2018) incorporate random coefficients into a purely cross-sectional version of the Bayer et al model, methodologically updating the earlier climate findings by Cragg and Kahn. In our empirical application, we use an alternative sorting framework with a novel source of identifying variation to reexamine the amenity values of both climate and air quality in a more recent time period.[3]

Our paper also adds to a growing body of work that methodologically extends the sorting framework for amenity valuation. Bishop (2007) estimates a dynamic national sorting model, noting that estimates of MWTP may be biased if households are forward-looking and make location decisions based on expectations of future amenities, wages, and prices. Hamilton and Phaneuf (2015) estimate a national sorting model that allows for a two-stage household choice of city and neighborhood. By jointly modeling between- and within-MSA sorting, they account for potential substitution patterns at the local level while inferring non-market value at the MSA level. Lee (2017) finds that household-level heterogeneity in moving costs is important in explaining national patterns of environmental injustice. In a hypothetical migration setting, Kosar et al (2020) use stated preference methods to infer MWTP

[2] There is a wealth of classic environmental/regional/urban economics work - likely beginning with Lowry (1966) - that explores US-centric migration determinants using behavioral models with underlying similarities to ours. None make the step to nonmarket valuation, as far as we know. Greenwood (1997) provides an overview of this research.

[3] While our main contributions are methodological, it merits note that the aforementioned papers all rely on Census data from the year 2000 or earlier. Air quality in American cities has continued its drastic improvement over the past 20 years - an average reduction of more than 45% for PM2.5 - while domestic migration has continued its secular decline (Molloy et al, 2011).

for several amenities of common interest in the urban and environmental literatures. The results from their hypothetical choice experiments imply substantial and systemic household-level heterogeneity in moving costs. The joint message from this growing set of papers is clear: overlooking behavioral complexities related to moving costs can yield misstated estimates of MWTP for air quality, climate, or other amenities.[4] Our work delivers a similar message. Household origin can play a meaningful role in explaining migration patterns, and therefore merits inclusion in the discrete choice framework.

Finally, we contribute to the large social science literature that explains migration patterns through the use of gravity-like models.[5] Our work builds from random-utility micro-foundations that have recently reestablished the gravity model as a structural workhorse in the study of international migration patterns. Beine et al (2016) summarize this resurgence and describe best practices for estimating gravity models of migration with aggregate flow data. Our approach to including micro-data into this gravity-like migration model varies from previous work. Hunt and Mueller (2004), Bertoli et al (2013), and Plantinga et al (2013) estimate household-level migration models using a nested logit framework to capture a stay/go decision hierarchy: a household's origin location falls in one nest and all other potential destinations reside in another. None of these papers value local amenities. In contrast, the sorting model we propose defines each origin MSA as a separate "market," thus serving as a more direct micro-analogue to Beine et al's baseline gravity framework. In a similar spirit to work by Druckenmiller and Hsiang (2019), we view this repeated cross-section of location choices as an opportunity to test the credibility of parameter estimates derived from spatial differences.

Organization

Section 2 of this paper describes stylized facts that motivate our model, while highlighting the publicly-available data on which the following analyses rely. Section 3 lays out theoretical foundations, underlining key differences from previous national sorting work. Section 4 discusses estimation and stylized results from an illustrative aggregate-data model. Section 5 discusses the estimation and identification strategies for our full model, incorporating household microdata. Section 6 discusses empirical results from our nonmarket valuation application. Section 7 concludes.

2 Context and data

A household's decision to move reflects the overcoming of locational inertia. This proves to be of crucial import when estimating the determinants of location choice. In a world without moving costs,

[4]Diamond (2016) and Bieri et al (2019) are also concerned with complexities in national sorting behavior, though are not focused on the valuation of specific amenities. The former fully endogenizes MSA-level amenity indices in a general equilibrium sorting model, while the latter builds from the model's key pieces to carefully measure *aggregate* US amenity expenditures.

[5]Applications of gravity-like models appear in several fields beyond migration where flow data are common: trade, transport, network analysis, and beyond. Though some in the migration and transport literatures have pivoted away from gravity models for purely predictive work - see, for instance, the radiation model - the gravity model's parametric structure remains highly tractable for our goal of amenity valuation. For more discussion of these applications, see Anderson, 2011; Sen and Smith, 2012; Simini et al, 2012.

the spatial distribution of housing prices and wages tell a rich story about the monetary value of a place's characteristics (Roback, 1982). In a less stylized world, the psychological and familial frictions of human ties to a specific place render these stories inapplicable; households stay, despite economic opportunity and amenity elsewhere. The unobserved costs of moving may overwhelm the observable benefits, and researchers therefore struggle to learn how households value spatial characteristics of other localities.

The national sorting approach adapts to this behavioral phenomenon and models it accordingly.[6] Nearly all previous discrete choice measures of household WTP for regional amenities have conditioned on demographic-varying measures of moving costs *away from birthplace*. This assumes that the cost of leaving *any* birthplace is homogeneous. The logic is simple: it is costly to choose a location further from your birthplace, perhaps more so if your family includes children or multiple full time wage earners. But these measures may still miss the subtlety of place: some cities are "stickier" than others. For example, several lines of research suggest that the psychological and monetary costs of leaving a former industrial MSA in the Rust Belt may differ from the costs of leaving a Sun Belt MSA.[7] If this is the case, a population-stock-based description of locational determinants will miss details that could be captured with flow data.

Figures 1 and 2 illustrate patterns in national data to support this assertion. Figure 1 shows the near-linear scatterplot of annual average out-migration and in-migration rates across the MSAs that we study from 2012-2016. A study of population stocks over time can tell us about population growth or decline rates over this period, but it would mask the fact that there is rich heterogeneity in inflow *and* outflow rates. Places that attract a relatively large number of people are also losing a lot of people. Some of this is mechanical: both city size and the characteristics of the city play a role. Figure 2 plots each MSA's annual churn ratio[8] from 2012-2016 against the percentage of initial population born in-state. If we differentiate by whether the MSA has a large university-related or military-related population (both of which may breed higher turnover naturally), the negative correlation between the share of population born nearby and the population's dynamism is striking. Stickier places do not appear to be drawing people from elsewhere, and places with a lot of churn seem less likely to have a population with high moving costs.

Urban and labor economists have shown recent interest in why locations become "sticky." Papers like Glaeser and Gyourko (2005), Gyuorko et al (2013) and Ganong and Shoag (2017) tell a market-driven story of the phenomenon. The durability of housing makes urban decline gradual in depressed cities: a large available stock of housing pushes prices down while the local economy slows. Local residents stay and wages stagnate. Meanwhile, land supply restrictions are pervasive in fast-growing, high-wage cities in the US: housing costs explode as local demand exceeds the constrained supply.

[6] To be clear, an MSA consists of an urban center and any surrounding suburban counties. We strictly study migration between these regional areas, not local migration between cities and their suburbs.

[7] See, as discussed in more detail below, Coate and Mangum (2019) or Zabek (2019). More broadly, results from Bartik (2018) and Morten and Oliviera (2018) also support such an assertion.

[8] (Inflows + Outflows)/Stayers

Together, these market mechanisms conspire to make migration costly and unattractive to households in depressed areas, especially those with less-skilled workers. But this does not necessarily explain the low migration from sticky places to alternative non-"superstar" cities, nor the low in-migration rates to sticky places.

Coate and Mangum (2019) approach this shortcoming by formalizing the notion of population rootedness: their motivating example describes how a Boston-native is likely to have deeper family history in Boston than an LA-native would in LA. Their paper documents how rooted preferences for home have affected spatial propensities to migrate over time, including the propensity to return "home" after years spent elsewhere. Zabek (2019) builds on this idea in general equilibrium: strong ties to a place keep locals there in spite of better economic opportunity elsewhere. In turn, this lowers local wages, making the location even less attractive to outsiders who don't have ties to the place. Dynamic cities naturally have more population churn: local ties are naturally weaker there and residents are more willing to pursue opportunity and amenities elsewhere.

In our valuation context, we remain agnostic on the specific mechanisms behind locational stickiness.[9] We simply note that if low amenity places are indeed burdened with higher moving costs (and high amenity places are endowed with lower moving costs), a conventional stock-based model with homogeneous birthplace-derived moving costs may inaccurately characterize preferences for amenities. Our approach accounts for heterogeneous spatial ties by varying migration costs with a household's origin location. These additional modeling demands require a data environment that contains the origin and destination of household moves, as well as detailed demographic information on the household itself. We construct a database that meets these requirements and provides broad coverage of household residential movements between 344 metropolitan areas in the continental USA during the period 2012-2016.[10]

The first piece of this database is an aggregated MSA-to-MSA migration flow matrix. This Census Bureau product provides an estimate of gross annual migration flows between MSAs based on 2012-2016 American Community Survey (ACS) responses. While the estimate is based on five years worth of data, it is generated from the ACS's annual survey question on where the household was located one year ago. We treat these aggregate flow estimates as a cross-sectional snapshot of migration in the US during the 2012-2016 period.[11] About 60% of the origin-destination pairs in our sample

[9] An additional driver of stickiness may be place-specific human capital. Becker (1962) and Neal (1995) discuss firm- and industry-specific human capital in labor markets. Workers with human capital in spatially-concentrated industries will earn systematically higher wages in those places. Krupka (2009) also notes that local levels of amenities can endogenously motivate place-specific human capital investments in the domains of home production and leisure/recreation. We do not directly study human capital in this work, but do control for Roy-type sorting when predicting wage counterfactuals.

[10] There are actually 363 MSAs covered in our sample, but due to spatial boundaries in the microdata, we merged a subset of 35 MSAs into 16 super-MSAs. See below, and Appendix 2 for details.

[11] An alternative data product from the Internal Revenue Service (IRS) also measures population flows over space. They annually release county-to-county migration matrices based on counts of previous- and current-year tax filings. We rely on the Census measure for two reasons. First, the IRS data only counts tax-filing households. This may be problematic since a non-trivial percentage of US households do not file federal taxes. Second, the IRS data is highly censored: county-to-county flows of less than 10 households are not observable in the data.

represent zero migration flows. Looking merely at this migration flow data, we can observe gravity-like aggregate behavior from migrants in Figure 3. Larger destinations (in terms of population) and shorter migration distances result in greater flows, thus motivating the estimation approach described in Section 4.1.

Next, we have collected a range of amenity data measured at the MSA-level. Our empirical application is focused on climate and air quality. Figure 4 maps out the spatial distribution of these key amenities. Our measures of climate amenities are straightforward: 30-year-averages (1980-2010) of summer temperature, winter temperature, and annual precipitation.[12] For air quality, we use average annual PM2.5 levels, a measure that encompasses the general, background air quality that people are likely to associate with a given city. Since these air quality and climate variables are generated from remote sensing data, we spatially weight our MSA-level measures using census tract populations.[13] Summary statistics and data sources for these variables and other key MSA-level measures are shown in Table 1.

We have also collected annual measures for a variety of other MSA-level amenities that have been shown to be important to quality-of-life. These include geographic, economic, transportation, and urban characteristics of a given MSA, and are described in Table 2. Given the obvious concerns about omitted variable bias in cross-sectional modeling of the determinants of location choice, we include this rich set of covariates to serve as controls in our main model specifications.

Finally, household-level data was obtained from public-use microdata (PUMS) created by the Census Bureau. PUMS are a 1%, nationally-representative sample of US households, taken annually. Microdata of this type has been used in many previous sorting papers; the data's principle advantage is its provision of broad information on households' demographics, employment, and location. In order to temporally match aggregate measures of migration flows across MSAs, we have collected samples of this microdata from the 2012-2016 period. Crucially, for the vast majority of MSAs, we matched surveyed households to their current MSA as well as their previous MSA (had they moved in the past year).[14] Because we need to predict counterfactual incomes across MSAs in order to operationalize our empirical model, the microdata sample is cleaned by excluding households without an individual who earned more than $5000 in wage income, households with only self-employed earners, and households with primary earners actively employed by the military. Summary statistics for the resulting sample of around 2.5 million households are shown in columns 1-2 of Table 3. To speed computation time for models relying on this microdata, all results in Section 6 are based on a 10% subsample, described in the middle columns of Table 3.

[12] Summer temperature averages are over the months June, July and August. Winter temperature averages are over the months December, January, and February. Heating/cooling degree days have sometimes been used to model a temperature "bliss" point of 65°F. In our study sample of 344 MSAs, only one has an average winter temperature above 65°F, and 16 have an average summer temperature below that measure. In an appeal to parsimony and ease comparison with the previous sorting literature, we elect to proceed with our average seasonal measures.

[13] Spatial resolution on the climate variables' grid is 4km; for the PM2.5 data it is roughly 1km.

[14] See Appendix Table A2.1 for a list of omitted MSAs. Note that this matching exercise may result in small assignment errors for some MSAs, as geographic indicator boundaries in the microdata do not perfectly align with MSA boundaries. We discuss this in more detail in Appendix 2.

All together, our database includes complete amenity data, migration flows, and population microdata for 344 metropolitan areas in the US. These 344 MSAs account for over 97% of the 2012 non-rural population in the continental US; we deem this broadly representative of urban (and suburban) America. MSAs included in our sample can be seen in Figure 4. In the analyses that follow, we assume migration between these MSAs occurs in a closed-system. Households can select to live only in these areas, and households originate only in these areas. When calculating migration shares from a given origin, the implication of this assumption is that the summation of flows across these 344 potential destinations serves as the denominator of the measure.

3 Conceptual framework

Our static MSA-level migration model builds from the canonical random utility approach commonly adopted in the context of location choice. Households make a one-time decision on where to locate, based on the costs and benefits of each potential destination. Given a choice set of J_k location options, let household i, originating in MSA k and moving to MSA j, receive utility

$$U_{ij}^k = U(Y_{ij}, h_{ij}, X_j, M_{ij}^k, C_i, \xi_j). \tag{1}$$

A household generates utility from its consumption of a tradable, numeraire good, Y_{ij}; a non-tradable local housing good h_{ij} for which utility is separable; and a vector of MSA-level amenities, X_j. A household's well-being is net of any psychological or physical moving costs, M_{ij}^k, incurred from relocation. Preference heterogeneity arises from variation in observable household demographics, C_i. Finally, ξ_j denotes destination-specific amenity characteristics that are unobserved by the econometrician.

Conditional on destination, j, a household maximizes utility by choosing consumption levels of the housing good and numeraire good subject to destination-varying budget constraints

$$W_{ij} = Y_{ij} + p_j(h_{ij}). \tag{2}$$

In Equation (2), W_{ij} denotes the household's location-specific income and $p_j(h_{ij})$ is the local expenditure level needed to purchase housing of a given quality h_{ij}. Income is the sum of wage earnings for all household members; we abstract away from non-wage income. The household determines its optimal location by selecting the destination with the highest location-conditional utility.

To solidify ideas, we envisage amenity levels (X_j) as common to all households who reside in MSA j. These average MSA-level characteristics - climate, background air quality, etc. - are our primary objects of interest for nonmarket valuation. Undoubtedly, access to some amenities varies spatially *within* MSAs, and should be correlated with intra-MSA land prices and wages (Cropper, 1981). Conceptually, we view highly localized amenities like school quality or proximity to public transportation as characteristics embodied in housing quality, of which households can purchase more or less for local prices.[15]

Adding structure, we study indirect utility functions of the general form

$$V_{ij}^k = f(W_{ij} - H_{ij}; \alpha) + v(X_j, M_{ij}^k, C_i, \xi_j; \beta) + \varepsilon_{ij}^k, \; j = 1, \ldots, J_k \tag{3}$$

where $v(\cdot)$ and $f(\cdot)$ reflect standard neoclassical microeconomic behavior. Location-optimal housing expenditures are denoted by $H_{ij} \equiv p_j(h_{ij})$, for each household. β is a vector of utility function parameters, α parameterizes the marginal utility of numeraire consumption, and ε_{ij}^k is a separable, unobserved, and idiosyncratic component of utility. In our empirical work, we will let $v(\cdot)$ be linear in parameters but allow for the interaction of its components, primarily to model preference heterogeneity based on observable household characteristics.

Two noteworthy details follow from the setup. First, recall our primary focus is measuring a household's marginal willingness to pay for a given amenity, X^a. Taking the total differential of indirect utility, we have

$$dV_j^k = dX_j \frac{\partial v(\cdot)}{\partial X_j} + dM_{ij}^k \frac{\partial v(\cdot)}{\partial M_{ij}^k} + dY_{ij} \frac{\partial f(W_{ij} - H_{ij})}{\partial Y_{ij}}. \tag{4}$$

Setting all differentials to zero except for (i) our amenity of interest X_j^a and (ii) the numeraire good serving as our money metric, Y_{ij},

$$0 = v_{X^a} dX_j^a + f_Y dY_{ij} \Rightarrow \frac{dY_{ij}}{dX_j^a} = -\frac{v_{X^a}(\cdot)}{f_Y}. \tag{5}$$

which is our simply-derived measure of marginal willingness to pay for a local amenity improvement.

Second, when ε_{ij}^k is distributed i.i.d. type I extreme value (EV1), the probability of household i settling in location j, conditional on originating in location k, can be conveniently expressed as

$$\Pr[V_{ij}^k \geq V_{il}^k, \forall l \neq j] \equiv \Pr_{ij}^k = \frac{\exp\Big(f(Y_{ij}; \alpha) + v(X_j, MC_{ij}^k, \xi_j, C_i; \beta)\Big)}{\sum_{J_k} \exp\big(f(Y_{il}; \alpha) + v(X_l, MC_{il}^k, \xi_l, C_i; \beta)\big)} \tag{6}$$

The mechanical, logit-derived relationship between observed choice probabilities in the data and utility as a function of origin and destination characteristics proves valuable for inferring the value of a location's amenities in a gravity-like fashion. We elaborate on this subject next while discussing our estimation approaches.

[15] See Hamilton and Phaneuf (2015) for a framework that more formally integrates inter- and intra-MSA substitution patterns.

4 Using gravity to understand varying migration propensities

We now take the model to our data environment, viewing the study period as a single cross-section of location decisions. As described in Section 2, the data contains counts of average annual migration flows, where we observe population estimates of individuals relocating from an origin to destination location. We also work with a sample of household microdata that provides demographic information and residency locations over the past year. We refer to these as the aggregate and micro-level data, respectively, and again use $i = 1, \ldots, I$ to denote elements (i.e. households) of the micro-level data.

Denote by $j = 1, \ldots, J$ the universe of locations under analysis. Consider moves that originate at origin $k \in J$, and let $J_k \subseteq J$ denote the destination alternatives that are in the choice set for origin k households. Decompose indirect utility for a household i starting in location k and then migrating to destination j as

$$V_{ij}^k = \delta_j^k + \mu_{ij}^k,\ j = 1, \ldots, J_k \tag{7}$$

where δ_j^k is the deterministic average utility of moving to location j among households starting in place k, and μ_{ij}^k collects the household-specific components of utility that are both deterministic and random. In what follows, we leverage this delineation of indirect utility in two ways. First, we describe the model without household-level heterogeneity, and highlight what its gravity-like structure can tell us about a location's average attractiveness and migration propensity. Then we develop the full model for amenity valuation, incorporating microdata to study richer household-level preferences; our two-stage estimation method builds from the decomposition above to generate parameter estimates.

4.1 Stylized model with aggregate data

To develop an aggregate migration flow model, we assume μ_{ij}^k is i.i.d. type I extreme value. Thus, household deviations from origin-destination average utility are treated as random, meaning we study a representative, average household. Suppose a simple, linear-in-parameters average utility for households moving to j from k:

$$\delta_j^k = \alpha \bar{Y}_j + \beta X_j + \gamma M_j^k + \xi_j \tag{8}$$

where X_j is a vector of (observed) destination-specific amenity characteristics, M_j^k are (observed) migration cost factors with dyadic variation, and $\bar{Y}_j \equiv \bar{W}_j - \bar{H}_j$ is the average income of households in j, net of average housing expenditures.

Given the assumed extreme value assumption on μ_{ij}^k and the nature of δ_j^k, we can write the parameterized migration probability of a $k \rightarrow j$ move as:

$$\Pr_j^k = \frac{\exp\left(\alpha \bar{Y}_j + \beta X_j + \gamma M_j^k + \xi_j\right)}{\sum_{l \in J_k} \exp\left(\alpha \bar{Y}_l + \beta X_l + \gamma M_l^k + \xi_l\right)}. \tag{9}$$

As pointed out by Beine et al (2016), and earlier by Anas (1983), the random utility structure behind this choice probability yields a gravity-like equation when taken to the data. To see this, define S^k as the stock of households originating in origin k, and rearrange:

$$N_j^k \equiv \Pr_j^k \times S^k = \frac{\overbrace{\left(\exp\left(\alpha\bar{Y}_j + \beta X_j + \xi_j\right)\right.}^{\phi_j}\overbrace{\exp\left(\gamma M_j^k\right)}^{\tau_j^k}}{\underbrace{\sum_{J_k}\exp\left(\alpha\bar{Y}_l + \beta X_l + \gamma M_l^k + \xi_l\right)}_{\Omega_k}}(S^k) = \tau_j^k \frac{\phi_j}{\Omega_k} S^k \tag{10}$$

The left-hand side variable, N_j^k is the aggregate gross flow of migrants from k to j, as observed in the data. Algebraically, this is equivalent to the predicted proportion of $k \to j$ movers multiplied by the population stock that began in k. On the equation's right hand side are the several structural components of gravity. The *push* factors pertain strictly to the origin location. Larger population flows originate from MSAs with high initial population stocks (S^k) or low levels of expected utility for prospective migrants (Ω_k), taking into account an origin's other potential destinations.[16]

τ_j^k captures the *friction* of movement from k to j. For $\gamma < 0$, τ_j^k must take a value less than or equal to 1, which implies smaller flows occur where the costs of migration between origin and destination are high. In this aggregate model, we specify distance-based migration costs in order to capture the inertial forces that keeps people from leaving their origin, and the psychological costs of moving longer distances from family and friends:

$$M_j^k = \sum_n \mathbf{1}\{\text{move distance} \in b(n)\} \tag{11}$$

This specification categorizes distances between origin and destination into the following seven bins, $b(n)$: 0km (staying in origin MSA), 0-250km, 250-500km, 500-1000km, 1000-2000km, 2000-3000km, 3000+km. We set 0km as our baseline bin, implying that the remaining bins' γ values should all be negative, and decease as the distance between j and k increases.

And finally, the model's *pull* factor (ϕ_j) describes the benefits of moving to destination j - the size of the observed migration flow should increase with the attractiveness of the destination. Here, a destination MSA's average net income and level of amenities both serve primarily as pull factors.[17]

[16] As Beine et al (2016) point out, $\frac{\partial\Omega_k}{\partial\tau_l^k} > 0$ and $\frac{\partial\Omega_k}{\partial\phi_l} > 0$. In essence, this term captures how the characteristics of alternative potential destinations, l, affect the magnitude of an origin's population flows.

[17] One characteristic of an MSA that we treat as a pull factor is population. When $ln(\text{population}_j)$ enters as an element of X_j, we have the classic $\frac{P_1 P_2}{D}$ gravity structure described by Zipf (1946). Our inclusion of population as a pull factor is a notable advantage of the framework. Equilibrium sorting models generally aim to predict population shares in each destination, typically rendering inappropriate the direct inclusion of population as an attractive feature. Since we predict origin-specific *migration shares*, we are able to control for this important regional characteristic.

4.2 Estimation and results

Our goal with the aggregate model is to present stylized evidence on the attractiveness and migration propensities of different MSAs. The former is crucial in a revealed preference sense: overall quality of life in our model is the combination of net income and local amenities from which households derive utility. From this measure, we can break down trade-offs between income and amenities that allow us to calculate marginal willingness to pay for local characteristics. Migration propensities are also important: as we show below, population dynamics are heterogeneously fluid across space. If we do not control for this as a migration cost, we may infer a preference for amenities rather than an unrelated preference for place.

To recover these measures, we turn to econometric tools commonly used in the international trade and migration literatures. Our main estimation equation is an empirical specification of Equation (10), written as:

$$N_j^k = \exp\left(\phi_j + \theta_k + \gamma M_j^k\right) \upsilon_j^k \tag{12}$$

The inclusion of a destination fixed effect, ϕ_j, indexes the attractiveness or quality of life in j. Ortega and Peri (2013) note that the inclusion of origin fixed effects, θ_k, captures multilateral resistances to migration that potentially exist in the origin dimension of Equation (10), $\frac{S^k}{\Omega_k}$. Put differently, a vector of origin fixed effects serves as an index of potential push factors and heterogeneity in the propensity to migrate across space. A well-behaved error term, υ_j^k, brings the model to data.

Since our dependent variable is always a non-negative integer, a count model is appropriate here. We estimate equation (12) by Poisson Pseudo Maximum Likelihood (PPML). This approach allows us to include origin-destination pairs with zero migration flows in our estimation sample, thus fixing $J_k = J, \forall k$ in the aggregate model. As demonstrated by Santos Silva and Tenreyro (2006) and Guimaraes et al (2003), this nonlinear estimator is appropriate in our context and remains consistent even when there is a high frequency of LHS zeros in the data or heteroskedasticity in the error term.

To accurately characterize heterogeneous migration propensities, we are most interested in getting at a measure of Ω_k. Since the set of estimated origin fixed effects resulting from equation (12) necessarily contains a population stock component, S^k, we derive our migration propensity index as the residuals of an additional linear regression:

$$\widehat{\theta_k} = \beta \ln(S^k) + \varepsilon_k. \tag{13}$$

Thus, we denote our migration propensity index as $\widehat{\varepsilon_k}$.

Tables 4 and 5 give a sense of the indices (ϕ_j, $\widehat{\varepsilon_k}$) that result from this empirical exercise. Table 4 displays the 25 least attractive destinations and 25 most retentive origins per the gravity framework. The destination MSAs populating this list are unsurprising - almost entirely smaller Rust Belt and Appalachian cities that have struggled economically over the past generation. The list of origin MSAs with low rates of population churn is more nuanced. Interspersed with declining and “sticky” Rust

Belt cities are Floridian MSAs with high levels of less-mobile retiree populations and a large Sun Belt city (Houston) that seems to be retaining its residents exceptionally well. Despite these examples, the fact that many cities are jointly included on these lists provides suggestive evidence that low quality of life areas may have higher associated moving costs.[18]

Table 5 tells the other side of the story. The most attractive MSAs, generally speaking, are large cities in the Sun Belt and Pacific/Mountain western states. Our index, derived fully from migration flow patterns, naturally mirrors the common growth narrative around these more affordable MSAs. Also in the top 25 are the amenity-rich and expensive coastal cities. In the right-hand columns, the 25 highest scoring MSAs on the migration propensity index tell a story that is less apparent on first glance. Most of these cities host military bases or universities, institutions that drive population churn organically (Watertown-Fort Drum, NY; Ames, IA; Lawrence, KS). Others are smaller Mountain/Western towns (Great Falls, MT; Missoula, MT; Medford, OR; Cheyenne, WY). And even controlling for population, our index finds that New York and San Diego both churn out migrants at a far higher rate than average.

Overall, these results tell a compelling and sensible story. In particular, they highlight heterogeneity in migration propensities across origins: a key result that motivates our empirical strategy in the next section. One additional detail to mention, however, is the potential to use this aggregate-flow-only framework to recover estimates of α and β in equation (10). Though the estimation results above lump these marginal values together in the quality-of-life index, a different approach could provide basic insights on average marginal valuations of individual amenities with very limited data requirements. To give a sense of this potential, we replace ϕ_j with its observable components $\alpha\bar{Y_j} + \beta X_j$ in equation (12), and re-estimate the model. These results are available in appendix Table A1.2. While these estimates are only suggestive, we find both parameter signs and magnitudes to be in the range of reason.

5 Empirical strategy for valuation: Incorporating microdata

Recall that our data environment also includes household microdata on a subset of migrants. We return to Equation (7), where we defined a household's indirect utility generally, and seek to make use of the additional information available in the microdata.

[18]There is room for further subtlety here. One can ask: why is a place sticky and how do its amenities influence this? Though Parkerburg-Vienna, WV and Sebastian-Vero Beach, FL have the lowest estimated migration propensities, they are likely sticky for very different reasons. Conventional wisdom would suggest that low-migration-propensity MSAs in Florida are likely to have amenity characteristics that are quite valuable to highly immobile populations (retirees), while sticky cities in the Rust Belt may have populations that lack the monetary or human capital resources to leave. We undertake an exploratory analysis of this in appendix Table A1.1 We regress $\widehat{\epsilon_k}$ on MSA-level demographic and economic characteristics, thus decomposing the aggregate migration propensity into some potential underlying drivers. Results in columns [1] and [2] do suggest there is a life-cycle element at play in our aggregate findings, while column [3] finds little clear evidence that "stickiness" is driven by income constraints. This line of thought is not our main focus here, but we suspect that working through how moving costs interact with amenity valuation over the life-cycle is a subject meriting future research.

In order to model the household's decision directly, we again divide indirect utility into an origin-destination-pair average, δ_j^k, and household-specific component, μ_{ij}^k:

$$\begin{aligned} V_{ij}^k &= \delta_j^k + \mu_{ij}^k \\ \delta_j^k &= \beta X_j + \gamma M_j^k + \theta_k + \rho_j^k \quad (14) \\ \mu_{ij}^k &= f(Y_{ij}) + \kappa M_{ij}^k C_i + \phi X_j C_i + \varepsilon_{ij}^k \quad (15) \end{aligned}$$

Here, ρ_j^k is the unobserved component of average origin-destination utility, which we will decompose and further discuss in section 5.3. There are several distinctions that result from our use of additional data. One is that income net of housing expenditures is now embedded in the household-specific component of utility. We specify $f(\cdot)$ as a piecewise linear spline with knots at the net income quintile cutoffs, $q_{(\cdot)}$, in our data's empirical distribution:

$$f(\cdot) = \begin{cases} \alpha_0 Y_{ij} & \text{if } Y_{ij} \leq q_{20} \\ \alpha_0 q_{20} + \alpha_1 (Y_{ij} - q_{20}) & \text{if } q_{20} < Y_{ij} \leq q_{40} \\ \quad\vdots & \quad\vdots \\ \alpha_0 q_{20} + \alpha_1 (q_{40} - q_{20}) + \ldots + \alpha_4 (Y_{ij} - q_{80}) & \text{if } Y_{ij} > q_{80}. \end{cases}$$

This semi-parametric specification allows for well-established patterns of diminishing marginal utility from income and consumption (Layard et al, 2008), without imposing it directly on the data through the use of common log or quadratic functional forms.[19]

Households have location-specific income and housing expenditures across potential destinations; we describe our counterfactual prediction procedures below. Another distinction is that preferences for amenities are heterogeneous; deviations from an amenity's average marginal utility are captured in the model's household-specific component by interacting a potential destination's amenity levels with observable characteristics of households. Finally, we allow migration costs to depend on household characteristics:

$$M_{ij}^k = \gamma M_j^k + \kappa_1 M_{ij} + \kappa_2 C_i (\mathbf{1}\{\text{move distance} > 0\}) \quad (16)$$

In addition to the binned origin-to-destination-distance measure from our aggregate model (M_j^k), we also allow for varying migration costs to certain destinations based on a household head's birthplace. M_{ij} is a vector of three dummies taking on a value of zero if the potential destination is in the head of household's birth state, Census sub-region, or Census region. In the case of households "coming back home", this control reflects the reduced psychological costs of return migration.[20] Our model also allows for demographic heterogeneity in migration costs through the interaction of a move/stay

[19]Morey et al (2003) discuss the stepwise marginal utility functions that result from this specification and how to perform welfare analysis in light of the resulting discontinuities.

[20]Indeed, "negative" migration costs for a return to home are found in papers like Coate and Mangum (2019) or Kennan and Walker (2011).

dummy with household characteristics. In particular, leaving one's current location may be more costly for households with children or multiple full-time earners.

5.1 Predicting counterfactual household variables

In the microdata, wages and housing expenditures are observable only in a household's chosen location. To estimate the model, counterfactual wages and expenditures are necessary for each potential destination. For this, we begin by estimating standard hedonic wage regressions using individual earners as the level of observation:

$$ln(wage_{ij}) = \alpha_j^w + \gamma_j^w C_i + \psi_{1,jk}\omega_{jk} + \psi_{2,jk}\omega_{jk}^2 + \varepsilon_{ij} \tag{17}$$

These regressions are run separately to obtain MSA-specific intercepts and slope parameters for the individual-level characteristics listed in Table A1.3. Using the microdata, we keep part- and full-time workers in the estimation sample, but have dropped all observations where individuals earned wages of less than \$5000 in the previous year. Fixed effects are included for industry and occupation. To correct for potential prediction bias resulting from nonrandom, Roy-type sorting, we follow Bayer et al (2009) in using Dahl's (2002) semi-parametric control-function correction. ω_{jk} denotes the worker-type-specific probability of regional migration. We create 16 distinct types by differentiating workers: they are binned by marital status, presence of children in the household, and which of four education levels they have obtained (less than high school diploma, HS or some college, college degree, graduate degree). We then calculate the empirical frequency of each type moving from one of nine census divisions to the others. Each worker in the estimation sample is assigned the resultant probability that matches their type, origin MSA, and destination MSA. The second-order polynomial of these worker-specific migration probabilities acts as a control function; the result is consistent estimation of the other hedonic parameters.

A summary of estimation results is presented in appendix Table A1.3. Across MSAs, the results are largely in line with the long literature studying the hedonic decomposition of wages: experience, education, gender, race, and ethnicity are generally found to influence wage levels, with some variation in magnitudes across labor markets. Likewise, the wage effect of working in a given industry or occupation is found to be place-dependent. We use the fitted values resulting from these parameter estimates as predictions of each individual's income in other locations. The simplifying assumption behind this modeling decision is that individuals maintain similar work and hours upon migrating. Then for each household, we sum individuals' annual wage earnings, yielding our counterfactual household wage income, $\widehat{W_{ij}}$.

The other counterfactual component of indirect utility is expenditures on housing. We follow Sinha et al (2018) and assume households select an identical housing bundle across space.[21] We motivate

[21] The standard alternative to this strong assumption would be a reliance on Cobb-Douglas preferences, as in Bayer et al (2009). This specification results in households spending a fixed percentage of income on housing, with the percentage being designated by the analyst.

this assumption in two ways. The first is practical: it transparently eases computation. When optimal housing quality is assumed uniform across choice options, it will only appear in household's indirect utility function via consumption of the numeraire. Therefore, we can simply predict housing expenditures as

$$H_{ij}^{*} = p_j\left(h_i^{*}\right) \tag{18}$$

where local implicit prices (p_j) for a given housing quality vary across potential destinations.[22] Below, we detail the estimation of these implicit prices.

Second, this assumption avoids a structurally-imposed relationship between household income and housing expenditures. National data shows that poorer households often spend a substantially larger fraction of their income on housing, especially in expensive markets (Quigley and Raphael, 2004). Our simple approach accommodates this phenomenon, which in turn should improve the quality of our estimates for the marginal utility of (net) income at the lower and higher ends of the income distribution.

We proceed by estimating hedonic housing price regressions, taking Equation (18) to the data. As with the wage regressions described above, these functions are allowed to vary by MSA, capturing regional differences in implicit prices for housing quality.

$$ln(cost_{ij}) = \alpha_j^c + \gamma_j^c h_i + \varepsilon_{ij} \tag{19}$$

The dependent variable in these regressions is annual housing costs, measured as the sum of monthly mortgage or rent payments, plus associated utility bills, taxes, and insurance. Independent variables include the usual medley of housing characteristics (h_i) that allow us to measure quality. A summary of regression results is presented in Table A1.4. As expected, newer and bigger houses generally cost more, while homeowners face lower expenses than renters in many cities, all else equal.

Per Equation (18), we construct the fitted values resulting from these regressions to predict households' annual housing expenditures across each MSA. With household incomes and housing expenditures predicted for each household $i = 1, \ldots, I$ in each potential destination $j = 1, \ldots, J$, we ultimately obtain our variable of interest: $\widehat{Y_{ij}} = \widehat{W_{ij}} - \widehat{H_{ij}}$ is the counterfactual net income that serves as our measure of numeraire consumption.

5.2 Maximum likelihood estimation

With all the relevant components of indirect utility now characterized, we proceed to describe our estimation of the model. Assuming ε_{ij}^{k} is distributed EV1, we return to the general logit structure of Equation (6) and adapt it to the specification laid out above. This is estimated by maximum

[22] Our approach also avoids the supposition of a national hedonic housing equilibrium. Under Cobb-Douglas structure, this is necessary to obtain the city-level housing price index that operationalizes econometric estimation.

likelihood, with the likelihood function given by

$$\mathcal{L}(\alpha, \kappa, \phi, \delta) = \prod_i \prod_{j=1}^{J_k} \left[\frac{\exp\left(f(\widehat{Y_{ij}}) + \kappa M_j^k C_i + \phi X_j C_i + \delta_j^k\right)}{\sum_{J_k} \exp\left(f(\widehat{Y_{il}}) + \kappa M_l^k C_i + \phi X_l C_i + \delta_l^k\right)} \right]^{\chi_{ij}} \tag{20}$$

where χ_{ij} is a dummy taking on the value of one if household i migrates to j.

We obtain the parameters in Equation (20) using methods from industrial organization literature. The origin-destination-pair average utilities, δ_j^k, are estimated as alternative-specific constants (ASCs) in a first-stage logit model, then decomposed in a second stage linear regression (Equation (14)) that recovers parameter estimates for β and γ. In our context, each origin serves as a "market"; thus, a flow from a given origin to destination is a *migration share*. Since the resulting parameter count is on the order of 10^4, to speed computation, we rely on a contraction mapping nested inside the first-stage maximum likelihood algorithm (Berry et al, 1995). As is well known, the inclusion of a complete set of ASCs in a logit model results in perfect in-sample prediction of observed market shares due to the distribution's mean-fitting property (Klaiber and von Haefen, 2019). We use this moment condition to aid in our numerical parameter search. Denoting σ_j^k as a migration share observed in the aggregate data and $\widehat{\sigma_j^k}$ the maximum likelihood estimation's predicted migration share, each iteration (r) of the contraction proceeds as

$$(\delta_j^k)^{r+1} = (\delta_j^k)^r + ln(\underbrace{\sigma_j^k}_{\frac{N_j^k}{S^k}}) - ln(\underbrace{\widehat{\sigma_j^k}}_{\frac{\sum_i Pr_{ij}^k}{I^k}})^r. \tag{21}$$

The key implication of this contraction approach is that δ_j^k is derived from the aggregate-level migration flow data, conditional on first-stage parameters which rely on variation coming from the microdata. A second point to note is our inability to estimate δ_j^k for origin-destination pairs with zero flows. When an empirical migration share falls to zero, the second term on the right side of Equation (21) becomes undefined. To deal with this shortcoming, we restrict choice sets by origin, imposing that households only consider migration to destinations where we see actual population flows occurring in the data. Therefore, the choice set J_k varies by origin.[23] A third and final point: since we can identify only $J_k - 1$ of the ASCs in each vector δ^k, we elect to normalize $\delta_k^k = 0, \forall k$. To account for this normalization in the second stage, we include origin fixed effects, θ_k, and thus rely on within-origin variation in all specifications.

[23]In essence, a zero flow is a symptom of a fairly classic selection problem (Carson and Louviere, 2014). In this context, we are suggesting that this selection is largely a consequence of distance from the origin and population size, rather than a systematic selection based on (observed or unobserved) quality of a destination's amenities. In their equilibrium framework for measuring total amenity expenditures, Bieri et al (2019) allow for similar consideration sets. Though their framework differs in key ways, they find little evidence that restricting the choice set in this way affects their results qualitatively. Future work - with more appropriate data - should more carefully study richer consideration patterns at the household level.

5.3 Identification

In the demand estimation literature surrounding this class of models, attention paid to identification issues typically focus on the second-stage regression. That is the case here as well: our model involves using a linear regression to decompose average utility into its observable and unobservable elements. From equation (14), we have

$$\begin{aligned}\delta_j^k &= \beta X_j + \gamma M_j^k + \theta_k + \rho_j^k \\ &= \beta X_j + \gamma M_j^k + \theta_k + \underbrace{\xi_j + \nu_j^k}_{\rho_j^k},\ k = 1, \ldots, J_k,\end{aligned} \tag{22}$$

where each δ_j^k is recovered in the first stage maximum likelihood routine. We emphasize that our framework has (1) provided us a second stage panel structure without any assumptions about holding preferences constant over time, and (2) provided substantially more second stage observations than we would obtain in the standard sorting framework. Together, these points naturally suggest estimating β and γ using a panel fixed effects model, where origins are the cross-sectional units and each origin contributes J_k observations. Thus, we identify the elements of β from spatial variation in X_j across the landscape, within origin variation in aggregate migration flows to destinations in J_k, and variation in the size and composition of J_k across origins.

A challenge for consistent estimation is that the unobserved drivers of migration flows arise in part from unobserved amenities at the destinations. This is apparent when ρ_j^k is rewritten to reflect a destination-fixed component ξ_j and an idiosyncratic component, as in the second line of equation (22). Consistent estimation by least squares requires $corr(X_j, \xi_j) = 0$, which may not hold due to the natural tendency of correlation between observed and unobserved amenities across the landscape. This suggests an instrument is needed for a potentially endogenous attribute like air quality.

We explore possible instruments for air quality by adapting strategies from the industrial organization literature, whereby attributes of competitors' products are used to construct an instrument for product price. In our case, pollution concentrations are correlated across space due to geography and the location of emitters in the landscape. For example, if X_j and X_m denote air quality at neighboring locations j and m, it is likely that $corr(X_j, X_m) \neq 0$. This remains the case when there is a moderate distance between places j and m due to the strong effects of wind patterns and geography. At the same time, it is plausible that $corr(X_m, \xi_j) = 0$ for large enough distances between the locations. Based on this logic we employ an instrumental variables strategy using

$$\tilde{X}_j = \sum_{m=1}^{J} \mathbf{1}(\text{dist}_{jm} \in T) \left(\frac{1}{\text{dist}_{jm}^2} \right) X_m \tag{23}$$

as the instrument for X_j. $\mathbf{1}(\cdot)$ is an indicator function equal to 1 if the distance between j and m is in some threshold window, T. This is multiplied by the pollution level in m, then inversely weighted by the squared distance between j and m. All together, our instrument is an inverse distance-weighted sum of air quality. In constructing the instrument, we explore the robustness of results to different distance thresholds, T, and to summing over the full set of locations J versus the subset J_{-k}, defined as the set of destinations not in the choice set for origin k. Summing over the subset of MSAs not in an origin's choice set is potentially compelling here: it provides instrument variation in the origin-destination pair dimension and relies on air quality measures in places excluded from household consideration by our model's construction. The trade-off of this approach is that it necessarily creates a weaker instrument, as fewer neighboring cities contribute their pollution.

Finally, our preferred instrument also incorporates wind patterns, a substantial component of localized pollution that is plausibly exogenous. Taking the logic and data from Deryugina et al (2019), we leverage the North American Regional Reanalysis (NARR) daily satellite wind direction data to calculate the percentage of days from 2012-2016 that wind blows from a neighboring MSA, m, into the destination MSA, j. More formally, we sum the number of days that wind in MSA j came from a directional angle in the 45° window around the compass bearing between m and j, and then divide by the count of days over the 2012-2016 period.[24] This measure, w_{jm}, is then interacted with the instrument measure above, generating the new sum:

$$\tilde{X}_j = \sum_{m=1}^{J} \mathbf{1}(\text{dist}_{jm} \in T) w_{jm} \left(\frac{1}{\text{dist}_{jm}^2} \right) X_m. \tag{24}$$

This preferred instrument captures the decaying nature of air pollution over distance and increasing effect of wind direction, summing these weighted values over all MSAs in the threshold band.[25]

We illustrate two versions of these instruments on a map in Figure 5. Panel A displays a case where all MSAs located at a distance within 100 to 1000km of the destination (Madison, WI) contribute to the instrument. These distances are used as our baseline thresholds. Winds from the south and west dominate in Madison, so pollution from neighboring MSAs in that direction contribute relatively more to the instrument. Panel B displays a case where only MSAs outside the choice set of the origin (Tucson, AZ) contribute to the instrument.

These instruments' validity hinge on the assumption that $corr(X_m, \xi_j) = 0$, which we believe is plausible in our migration flow context. Note in particular that ρ_j^k in equation (22) arises from origin-

[24]Since the satellite data is on $2.5° \times 2.5°$ grid, our daily measure of an MSA's wind direction is the spatial average of relevant grid cells, weighted over the MSA's entire land area.

[25]For completeness, we note that our instrument does have inherent similarities to those previously used in national sorting models. Bayer et al (2009) use an EPA point source-receptor pollution matrix to calculate the amount of an MSA's particulate matter that was generated by emitters sufficiently far away the city. This measure serves as their instrument. Freeman et al (2019) use instrumental variables in the China context that fall somewhere between the Bayer et al instrument and ours: theirs are constructed from a combination of prevailing wind directions, the location of thermal power plants outside the city, and the amount of coal used by those plants. Though our preferred instrument relies on similar correlations between non-local emissions and local pollution, we find it advantageous in several dimensions. Its construction is transparent and easily replicable across different pollutants, it has some ability to capture inbound air pollution from both point and non-point sources, and its source of exogenous variation (inter-MSA wind patterns) is plausible.

specific migration shares, rather than the overall share of the population that resides in location j in spatial equilibrium. This is an important distinction between our approach based on flows, and models that use the stock of residents at a location to estimate preferences. In our case, the main endogeneity problem is omitted correlated amenities, which we address with a rich set of controls and our instrument. In applications using aggregate population shares, however, the main endogeneity problem is simultaneous determination: amenities X_j and ξ_j are joint endogenous outcomes in the spatial equilibrium, making identification more challenging. Thus, the distinct sources of variation used in flow versus stock models should condition how we evaluate the validity of identification assumptions in each context.

6 Results

We start by discussing our full model's baseline results, which focus on mean MWTP estimates for our amenities of interest. The key parameter estimates for this model are in Table 6. The first-stage, household level parameters - estimated by maximum likelihood and shown in Panel A - are in accordance with our priors. The (piecewise) marginal utility of income is shown to be diminishing and moving costs specific to the household are sizable and of the correct sign.

Panel B contains the parameter estimates for the second-stage, linear regression. Recall that we are regressing the average origin-destination utility level on a set of destination characteristics, an origin fixed effect, and a vector of distance bins that proxy for distance-related moving costs. The four sets of results here are from separate models; column (1) is a standard least squares estimator, while columns (2)-(4) use different versions of our instrumental variable to address endogeneity concerns about the air quality parameter. Because we are interested in a nationally representative estimate of mean MWTP, we use generalized least squares in all specifications, weighting observations by their origin's population divided by the size of the origin's choice set.

The economic interpretation of our key parameters holds across specification, though the magnitudes vary slightly with the introduction of the instrument. As we expected, households prefer drier climates, warmer winter temperatures, cooler summer temperatures, and cleaner air. As a point of comparison, the model in column (1) finds that a household earning close to the median national income has an annual MWTP of \$702 for a one unit reduction in PM2.5.

In column (2), we instrument for PM2.5 using a distance- and wind-weighted sum of pollution from **all** neighboring MSAs that fall between 100 and 1000 km from the destination. This is our preferred instrument. Our comparison, median-income household's implied MWTP for cleaner air more than doubles to \$1,615. For the results in column (3), we use an instrument that does not weight by wind direction, but rather sums distance-weighted pollution from all neighboring MSAs to the destination city. With this instrument, the magnitude of MWTP for cleaner air increases to over \$2,700. Lastly, in column (4), we rely on the imposed structure of our model to develop an instrument that includes only pollution from places not under consideration. We sum distance- and wind-weighted pollution from neighboring MSAs that are not in the origin's choice set. The magnitude of the parameter on PM2.5 is dampened significantly relative to columns (2) and (3): our comparison household has a MWTP of $1,116 for cleaner air.

Across different flavors of the instrument, we find our results to be statistically significant and generally sensible.[26] Our preferred instrument in column [2] balances a plausible exclusion restriction with a strong first stage. For this preferred set of results, we also find that our comparison household has a MWTP of $2,507 for a one degree (Celsius) warmer winter temperature, $1,287 for a one degree cooler summer temperature, and approximately $202 for 1 cm less annual precipitation. Stepping back for a broader view, Figure 6 traces out the average MWTP curve for each of our four environmental amenities of interest. These stepwise curves illustrate how MWTP for these amenities increases with household income levels.

6.1 Benchmarking against stock-based models

Next, we contrast our model's results against those from a conventional national sorting model. In sorting models powered by the distribution of population stocks over space, first stage estimation is nearly identical to our framework, but ASCs are included only for destinations. In other words, there are 344 ASCs in the conventional model versus over 40,000 in our migration flows model. Origin does not play an essential role in stock-based settings; the model's second stage simply regresses each destination's average utility on its set of amenities.

Results in Table 7 underscore some of the advantages of our approach. To ensure full comparability, in all three specifications shown, we use an identical IV (see column 2 of Table 6) and the same sample of data. Column (1) shares parameter estimates for a standard stock-based model. As the first-stage parameters look broadly similar to those in our flow model, our ultimate interest is primarily in the second stage estimates. Here, the parameters for climate amenities are correctly signed and of sensible magnitude. However, we find a perverse sign on PM2.5, *despite* use of an IV. This is a fairly regular phenomenon in this literature, and indeed motivates Bayer et al's (2009) use of temporal first differences to absorb MSAs' time-invariant unobservable characteristics. Additionaly, due to the comparative paucity of observations ($N = 344$) in the second stage of the conventional model, parameter estimates are imprecise.

We next turn to an adaptation of this stock-based model in column (2). The two most apparent differences between conventional stock models and our flow model are the dimensionality of ASCs and the inclusion of distance-based migration costs. To isolate the role of these differences, we run a stock-based model that includes corresponding origin-to-destination distance bins in the first stage. The resulting second stage estimates from this specification are noisy: the sign on air quality flips, but its magnitude is implausibly large and the estimate is imprecise. Likewise, the magnitude of the summer temperature grows substantially and suggests households greatly prefer warmer summer

[26] In appendix Table A1.5, we test robustness to alternative distance thresholds for our instrument. The parameter on PM2.5 remains negative and maintains strong statistical significance across alternative distance bandwidths.

temperatures on average.

Column (3) re-displays flow-based results from Table 6 that rely on an instrument identical to the one used in columns (1) and (2). Structurally, this model is identical to column (2) except it uses ASCs derived from flows instead of stocks and therefore includes origin fixed effects in the second stage. This flow model is the most directly comparable to the stock models we've presented here. Looking across columns, our flow approach appears to deliver the strongest results. Our take-away from this set of estimates is that our flow model's key innovation is the inclusion of both origin and destination in our ASCs. The dimensionality and richer spatial variable available in our approach's second stage produces estimates that are plainly preferable to the conventional sorting model. Of course, as mentioned, we have avoided inclusion of a temporal dimension here - a tactic that would surely improve the quality of estimates from the stock-based models.[27] Nonetheless, we note that our model manages to consistently sign the parameter on air quality correctly without this temporal dimension, and even without use of an IV.

6.2 Robustness

We also consider our preferred baseline model's robustness to specification and sample selection. We first re-examine our utility function's treatment of net income; our model relies on a separable piecewise linear spline. Using this semi-parametric technique in Table 6, we found point estimates that met our priors - marginal utility from income is diminishing. Nonetheless, the convention in the literature has been to allow income to enter the utility function linearly or logarithmically. We estimate models using these more restrictive assumptions and show the results in Table 8.

Column (1) shows point estimates when income enters the utility function linearly; as is well known, this assumption fixes the marginal utility of income as constant. In monetary terms, we find the resulting mean national MWTP for a one unit reduction in PM2.5 to be on the order of \$2,130. This is more than 30% larger than the value implied by our preferred model for a household earning the median income. Indeed, this value is closer to the implied air quality MWTP (\$2,250) for households in the top income quintile under our preferred income specification.

Similarly, we find larger MWTP estimates when income enters logarithmically. In this specification, marginal utility is restricted to fall proportionally with the household's income level. Parameter estimates are shown in column (2). For a household earning close to the median value of income net of housing (\$56,000), the implied MWTP for a one unit reduction in PM2.5 is around \$2,587. In imposing the logarithmic functional form to capture diminishing marginal utility in the standard way, we've forced the marginal utility of income to diminish faster than shown by our piecewise linear

[27] To give some sense of comparison, in appendix Table A1.6, we collect inflation-adjusted and measure-normalized estimates of mean MWTP for air quality and climate from several national sorting models that depend on a stock-based structure. Based on parameter values taken from these prior papers, we find that our MWTP estimates for climate amenities are of a similar magnitude, though they do tell a qualitatively different story. On the surface, our estimated mean MWTP for air quality is large compared to those derived from a time of higher overall air pollution. Interestingly, however, if we use the structural parameters in Bayer et al (2009) and Hamilton and Phaneuf (2015) to calculate their implied MWTP at *modern* air quality levels, we find that our point estimate falls in between the two papers.

estimates. Conditional on the rest of our model's structure, this logarithmic assumption results in MWTP estimates that we find implausibly large for much of the income distribution.

As a final point before moving on to sample robustness, we emphasize that in most applications of discrete choice sorting models, measures of MWTP are a ratio of only two parameters. The functional form chosen for the marginal utility of income defines one half of this equation; ensuring the credibility of its estimate is of utmost import if we are to believe the values of MWTP implied by our model. We feel that our semi-parametric approach to estimating the marginal utility of income is a novel step forward, but encourage further exploration of this matter in future work.

Moving along, Column (3) of Table 8 shares parameter estimates when our model is estimated on a movers-only subsample of the national population. More specifically, the sample includes only households who had moved from their previous residence to a new one in the past 12 months.[28] Previous work has suggested that these households might be closer to "in equilibrium" than households who stay in place, and that this can contribute to the gap between MWTP estimates from hedonic and sorting models. Indeed, Wong (2018) intuitively demonstrates under strong but tractable conditions that inframarginal households have indifference curves that are unlikely to fall tangent to the hedonic price function.

This sample of households differs substantially from our national sample (see Table 3 for summary statistics), but the parameter estimates imply mean MWTP values roughly in line with those from the complete national sample. A household at the median net income level has a MWTP for lower PM2.5 of $1,036. Interestingly, the estimated marginal utility of income for recent-movers earning in the upper income quintiles is noticably larger than that estimated for the population as a whole; the resultant MWTP curves are a bit flatter for recent movers than those illustrated for the entire population in Figure 6. As discussed above, this accords with economic intuition - households who move are perhaps more likely to relocate for a wage premium or housing discount than the population's average household.

6.3 Heterogeneity in preferences

Next, we use variation in observable characteristics to study heterogeneity in household MWTP for environmental amenities. As in previous sorting work, we begin by interacting demographic characteristics with amenity levels at the destination; from this first-stage interaction, we can see how preferences vary across subgroups of the population. Here, we study how tastes vary for households with older (> 59 years), younger (< 30 years), or more educated (college grad) heads, as well as those with children present.

Parameter estimates are shared in Table 9; note that reported values in this table are from a single estimated model. Panel A contains the heterogeneity-related parameters, while Panels B and C hold the the homogeneous parameters familiar from previous specifications. Several findings common to

[28] We also adjust the aggregate MSA-to-MSA migration flow matrices to include only population counts who moved in the past year.

the literature emerge from these results: households with younger and college educated heads are substantially more sensitive to environmental amenities than the population at whole, household with children are less sensitive to climate and air quality (dis)amenities than their childless peers, and there seems to be relatively limited heterogeneity in distaste for PM2.5. Surprisingly, we do not find that older households have strong tastes for climate amenities — perhaps because our model only includes households who remain active in the labor force.

As there are many dimensions of heterogeneity occurring here simultaneously, we focus on a single set of subgroups in order to illustrate some richer MWTP patterns. In Figure 7, we trace out MWTP curves separately for four types of households: childless without college, childless with college, children-having without college, and children-having with college. As expected, some notable variation in tastes is masked by simply looking at mean MWTP estimates. Note that these MWTP estimates rely on homogeneous marginal utility of income across subgroups; an interesting but computationally-costly potential avenue for future work would be the incorporation of demographic-driven heterogeneity in this dimension as well.[29]

Next, we study how preferences may vary *by origin*. While others in the sorting literature have studied preference heterogeneity conditional on destination via random coefficient models, the structure of our model lends itself to the study of how tastes vary by origin. To illustrate this, we use the ASCs recovered from our preferred baseline model in Table 6 and estimate a second-stage regression where destination amenity levels are interacted with dummy variables indicating the quintile of the origin's amenity level. For example, a very cold city - say Fargo, ND - would fall into the lowest quintile (Q1) for average winter temperature. Through the use of these interaction terms, we can study how preferences for climate or air quality differ conditional on a household's origin baseline amenity level.

Results from this second stage regression are shown in Table 10. Some findings suggest a degree of previous sorting on environmental preferences: households in drier origins (largely in Western and Plains states) have the strongest average distaste for precipitation, and households in the coldest origin cities (namely, the Pacific northwest) strongly prefer cooler summer temperature. On the other hand, we surprisingly find that the strongest preferences for warmer winter temperatures comes from households in colder origin cities. One explanation for this might be that moving costs from the these colder origins in the upper Midwest and East coast are larger than average. Despite stronger preferences for warmer winter temperatures, many households from these origins remain close to home. Those that do leave, leave for warmer places. And lastly, we find that households with the strongest distaste for PM2.5 are from origins falling in the middle of PM2.5 distribution. Notably, those from origins with the highest levels of particulate pollution appear not to value cleaner air at all.

Finally, we highlight an additional dimension of preference heterogeneity offered by our framework that could be of interest in other applications. It would straightforward but computationally expensive

[29]In an unpublished working paper, Lee (2017) emphasizes this very point.

to incorporate household-by-origin preference heterogeneity in the first stage of our model or estimate separate ASCs by demographic group. Despite these computation costs, we can imagine the interplay of demographic characteristics and place as particularly interesting in the context of the environmental justice literature; perhaps this flavor of sorting framework could shed new light on nuisance-driven residential mobility (Depro et al, 2015).

7 Conclusion

Bayer et al (2009) showed the importance of moving costs for measuring the marginal willingness to pay for regionally varying environmental amenities using residential location choices. These moving costs drive a wedge between amenity values and gradients estimated using wage and home price hedonic regressions. In the national sorting model literature these moving costs are proxied using a household head's birth location. Recent literature in urban and labor economics, however, has documented that there is place-based heterogeneity in moving costs. Some places are "stickier" while other locations display a lot of "churn". If low quality of life locations are also places that are costlier to leave – an empirical regularity that we document – we may incorrectly conclude from the spatial distribution of population stocks that local amenities are not important to households. Our main contribution in this paper has been to develop and estimate a generalization of the familiar national-level sorting model that accommodates heterogeneity in migration costs across different origins by exploiting information on migration flows.

We find that estimates of marginal willingness to pay for air quality and climate amenities from our migration flow model are more plausible and better identified than corresponding estimates from a conventional cross-sectional sorting model. Observation of both origin and destination in our aggregate and household-level migration data allows us to: (a) include origin fixed effects in our model; (b) treat each origin as a separate market in our discrete choice model; and (c), identify average marginal utilities using within origin variation in average preferences for destinations. Relative to using population stocks, these innovations provide more precision – our second stage regression is an unbalanced panel in origin MSAs rather than a single cross section of MSAs – and opportunities for identification based on spatial differences. Second stage estimates of marginal utilities for air quality and climate amenities from our migration flow model are intuitive and qualitatively similar in sign and significance across OLS and IV estimators; in contrast, second stage IV estimates from the conventional population stock model are imprecise and in some cases perversely signed. This provides at least suggestive evidence that it is possible to consistently estimate preferences using a standard IO-literature instrument and without temporal first differences when migration flow data are available.

Specifically, our IV estimator is motivated by the older empirical IO literature that uses attributes of competing products as instruments for endogenous attributes, and recent literature in environmental economics that uses wind direction as an exogenous source of variation in pollution. Since we calibrate average utilities for destinations using migration flows from specific origins, rather than equilibrium population stocks, endogeneity problems for attributes like air pollution are more likely to arise from correlated omitted attributes and measurement error, rather than simultaneous determination in equilibrium. As a result, the exclusion restrictions needed for consistency of our IV estimator are more credible in our migration flow model than they are for a population stock model, and our estimates from the two models show this in practice.

In addition to our main innovation of using migration flow data, we also investigate an alternative functional form for how net income enters utility, which provides a flexible representation of the marginal utility of income across income levels. We find using our piece-wise linear representation that the marginal utility of income falls off at a slower rate than is implied by the commonly used log transformation. This has important ramifications for marginal willingness to pay estimates, and our results suggest this is an important topic for further research.

We close by noting that our migration flow model could be useful for other applications in environmental economics. Our examination of heterogeneity in this application has been largely illustrative, but our ability to differentiate marginal willingness to pay across income groups, household types, and geography could be useful for understanding the distributional consequences of environmental policy. Estimating a generalized version of this model that differentiates moving costs based on household types could be useful for environmental justice type questions. We leave these ideas for further research.

References

[1] Albouy, D., Graf, W., Kellogg, R., and H. Wolff, (2016). "Climate amenities, climate change, and American quality of life." *Journal of the Association of Environmental and Resource Economists*, 3(1): 205-246.

[2] Anas, A., (1983). "Discrete choice theory, information theory, and the multinomial logit and gravity models." *Transportation Research Part B: Methodological*, 13-23.

[3] Anderson, J., (2011). "The gravity model." *Annual Review of Economics*, 3(1): 133-160.

[4] Bartik, A., (2018). "Moving costs and worker adjustment to changes in labor demand: evidence from longitudinal census data." Working Paper.

[5] Bayer, P., Keohane, N. and C. Timmins, (2009). "Migration and hedonic valuation: the case of air quality." *Journal of Environmental Economics and Management*, 58: 1-14.

[6] Becker, G., (1962). "Investment in human capital: a theoretical analysis." *Journal of Political Economy*, 70(5): 9-49.

[7] Beine, M., Bertoli, S. and J. Fernandez-Huertas, (2016). "A practitioners' guide to gravity models of international migration." *The World Economy*, 39(4): 496-512.

[8] Bertoli, S., Fernandez-Huertas Moraga, J. and F. Ortega, (2013). "Crossing the border: self-selection, earnings and individual migration decisions." *Journal of Development Economics*, 101: 75-91.

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

[10] Bieri, D., Kuminoff, N. and J. Pope, (2019). "National expenditures on local amenities." Working paper.

[11] Bishop, K., (2007). "A Dynamic Model of Location Choice and Hedonic Valuation." Working paper.

[12] Carson, R., and J. Louviere, (2014). "Statistical properties of consideration sets." *Journal of Choice Modelling*, 13: 37-48.

[13] Coate, P. and K. Mangum (2019). "Fast locations and slowing labor mobility." Working Paper.

[14] Cropper, M., (1981). "The value of urban amenities." *Journal of Regional Science*, 21(3): 359-374.

[15] Cropper, M., Deck, L., Kishor, N., and K. McConnell, (1993). "Valuing product attributes using single market data: a comparison of hedonic and discrete choice approaches." *Review of Economics and Statistics*, 75(2): 225-232.

[16] Depro, B., Timmins, C., and M. O'Neil, (2015). "White flight and coming to the nuisance: can residential mobility explain environmental injustice?" *Journal of the Association of Environmental and Resource Economists*, 2(3): 439-468.

[17] Deryugina, T., Heutel, G., Miller, N.H., Molitor, D., and J. Reif, (2019). "The mortality and medical costs of air pollution: evidence from changes in wind direction." *American Economic Review*, 109(12): 4178-4219.

[18] Diamond, R., (2016). "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *American Economic Review*, 106(3): 479-524.

[19] Druckenmiller, H. and S. Hsiang, (2019). "Accounting for unobservable heterogeneity in cross section using spatial first differences." Working Paper.

[20] Freeman, R., Liang, W., Song, R. and C. Timmins, (2019). "Willingness to pay for clean air in China." *Journal of Environmental Economics and Management*, 94: 188-216.

[21] Ganong, P. and D. Shoag, (2017). "Why has regional income convergence in the US declined?" *Journal of Urban Economics*, 102: 76-90.

[22] Glaeser, E. and J. Gyuorko, (2005). "Urban decline and durable housing." *Journal of Political Economy*, 113(2): 345-375.

[23] Greenwood, M. J, (1997). "Internal migration in developed countries." *Handbook of Population and Family Economics*, 1: 647-720.

[24] Guimaraes, P., Figueirdo, O., and D. Woodward, (2003). "A tractable approach to the firm location decision problem." *Review of Economics and Statistics*, 85(1): 201-204.

[25] Gyuorko, J., Mayer, C. and T. Sinai, (2013). "Superstar cities." *American Economic Journal: Economic Policy*, 5(4): 167-199.

[26] Hamilton, T., and D. Phaneuf, (2015). "An integrated model of regional and local residential sorting with application to air quality." *Journal of Environmental Economics and Management*, 74: 71-93.

[27] Hunt, G., and R. Mueller, (2004). "North American migration: returns to skill, border effects, and mobility costs." *Review of Economics and Statistics*, 86(4): 988-1007.

[28] Kennan, J., and J. Walker, (2011). "The effect of expected income on individual migration decisions." *Econometrica*, 79(1): 211-251.

[29] Klaiber, H.A., and R. von Haefen, (2019). “Do random coefficients and alternative specific constants improve policy analysis? An empirical investigation of model fit and prediction”. *Environmental and Resource Economics*, 73: 75–91.

[30] Krupka, D., (2009). “Location-specific human capital, location choice and amenity demand.” *Journal of Regional Science*, 49(5): 833-854.

[31] Kosar, G., Ransom, T., and W. van der Klaauw, (2020). “Understanding migration aversion using elicited counterfactual choice probabilities.” *Journal of Econometrics*, forthcoming.

[32] Layard, R., Mayraz, G., and S. Nickell, (2008). “The marginal utility of income.” *Journal of Public Economics*, 92(8-9): 1846-1857.

[33] Lee, S., (2017). “A novel explanation for environmental injustice: household sorting and moving costs.” Working Paper.

[34] Lowry, I., (1966). *Migration and metropolitan growth: two analytical models.* Chandler Publishing Company.

[35] Molloy, R., Smith, C., and A. Wozniak, (2011). “Internal migration in the United States.” *Journal of Economic Perspectives*, 25(3): 173-96.

[36] Morey, E., Sharma, V., and A. Karlstrom, (2003). “A simple method of incorporating income effects into logit and nested-logit models: theory and application.” *American Journal of Agricultural Economics*, 85(1): 248-253.

[37] Morten, M., and J. Oliveira, (2018). “The effects of roads on trade and migration: evidence from a planned capital city.” Working Paper.

[38] Neal, D., (1995). “Industry-specific human capital: evidence from displaced workers.” *Journal of Labor Economics*, 13(4): 653-677.

[39] Ortega, F., and G. Peri, (2013). “The effect of income and immigration policies on international migration.” *Migration Studies*, 1(1): 47-74.

[40] Quigley, J., and S. Raphael, (2004). “Is housing unaffordable? Why isn’t it more affordable?” *Journal of Economic Perspectives*, 18(1): 191-214.

[41] Plantinga, A., Detang-Dessendre, C., Hunt, G. and V. Piguet, (2013). “Housing prices and inter-urban migration.” *Regional Science and Urban Economics*, 43: 296-306.

[42] Roback, J., (1982). “Wages, rents, and the quality of life.” *Journal of Political Economy*, 90(6): 1257-1278.

[43] Rosen, S., (1979). "Wage-based indexes of urban quality of life." In *Current Issues in Urban Economics, vol. 2*, edited by Mieszkowski, P. and M. Strassheim. Baltimore, Maryland: Johns Hopkins University Press, 74–104

[44] Santos Silva, J.M.C. and S. Tenreyro, (2006). "The log of gravity." *Review of Economics and Statistics*, 88(4): 641-658.

[45] Sen, A., and T. Smith, (2012). *Gravity models of spatial interaction behavior*. Springer Science & Business Media.

[46] Sinha, P, Caulkins, M. and M. Cropper, (2018). "Household location decisions and the value of climate amenities." *Journal of Environmental Economics and Management*, 92: 608-637.

[47] Sinha, P, Caulkins, M. and M. Cropper, (2019). "Do discrete choice approaches to valuing climate amenities yield different results than hedonic models?" Working Paper.

[48] Simini, F., González, M., Maritan, A., and A. Barabási, (2012). "A universal model for mobility and migration patterns." *Nature*, 484: 96-100.

[49] Tiebout, C.M., (1956). "A pure theory of local expenditures." *Journal of Political Economy*, 64(5): 416-424.

[50] Wong, M., (2018). "A tractable approach to compare the hedonic and discrete choice frameworks." *Journal of Housing Economics*, 41: 135-141.

[51] Zabek, M., (2019). "Local ties in spatial equilibrium". Working Paper.

[52] Zipf, G., (1946). "The $P_1 P_2/D$ hypothesis: on the inter-city movement of persons." *American Sociological Review*, 11: 677-86.

Figures and Tables

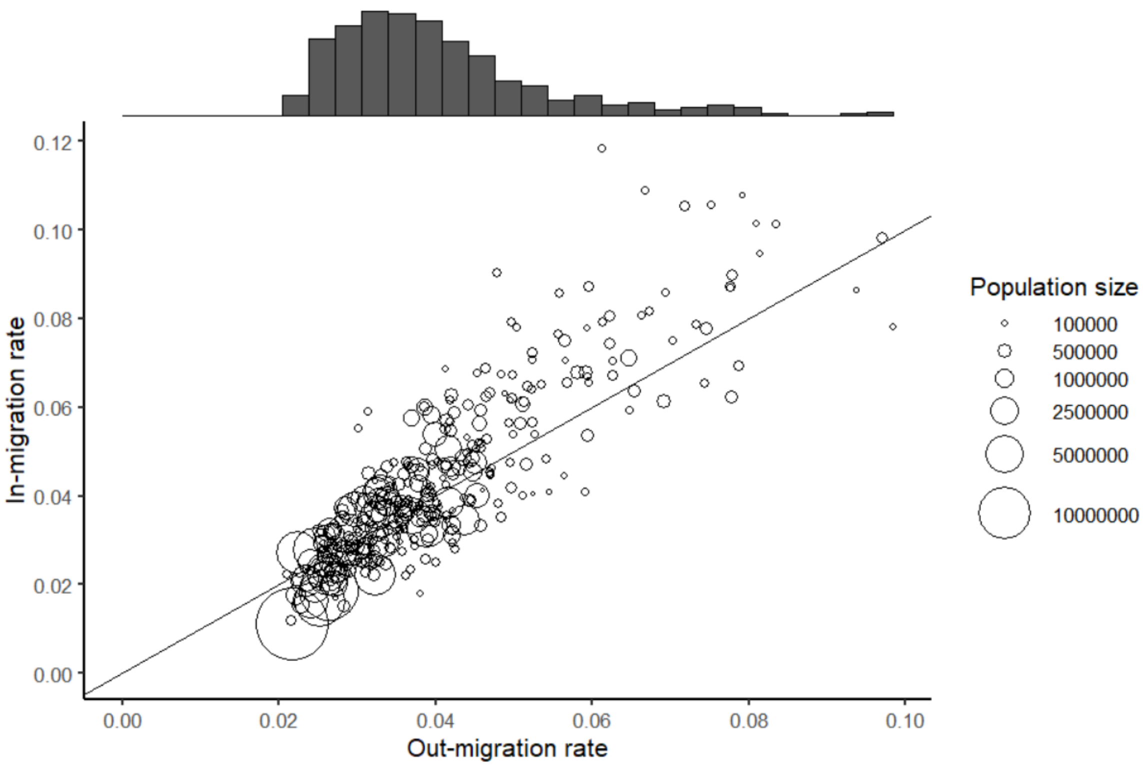


Figure 1: Net migration flows provide a fuller picture of migration patterns than population stocks. Each point represents a MSA included in our study sample (N=344). Points' size denotes the MSA's 2016 population size. Histogram of MSA out-migration rates, a measure of interest to us, shown above scatterplot; this illustrates rich heterogeneity across MSAs. Population size underlies this heterogeneity. Due to scale, percentage in- and outflows are smaller in large cities. 45-degree line maps equal replacement; deviation from this line denotes population growth or decline in an MSA. Data: US Census Bureau, 2012-2016.

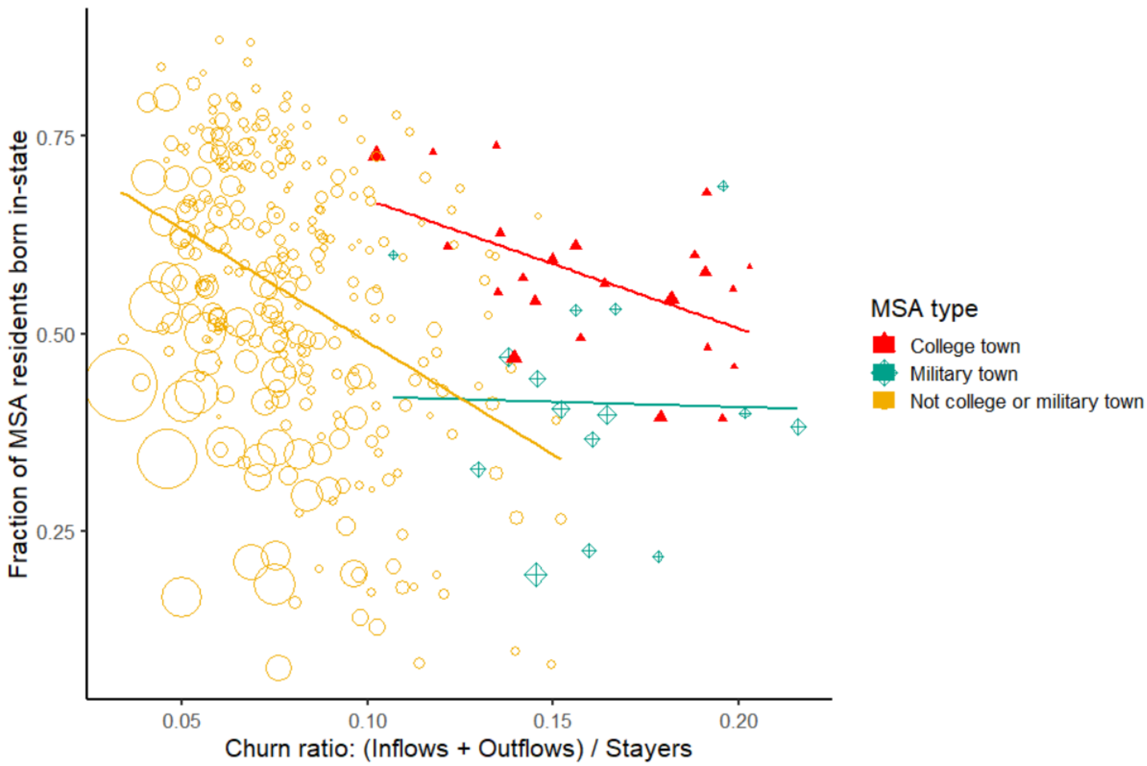


Figure 2: Sticky MSAs are those with less overall churn. Each point represents a MSA included in our study sample (N=344). Churn here is defined as total in- and out-migration counts divided by the count of people staying in place. The correlation between churn and rates of birth-state residency is partially mechanical: more churn may bring in more out-of-birth-state individuals. Nonetheless, research designs relying on distance from birthplace as a proxy for household moving costs miss important heterogeneity in leaving a *certain* place. Note: "College towns" and "military towns" are differentiated here as they inherently create churn. The former are defined as MSAs with > 15% of residents enrolled in college (data: Census Bureau, 2012-2016). The latter are defined as MSAs where > 8% of the labor force is employed by the military (data: Bureau of Economic Analysis, 2016).

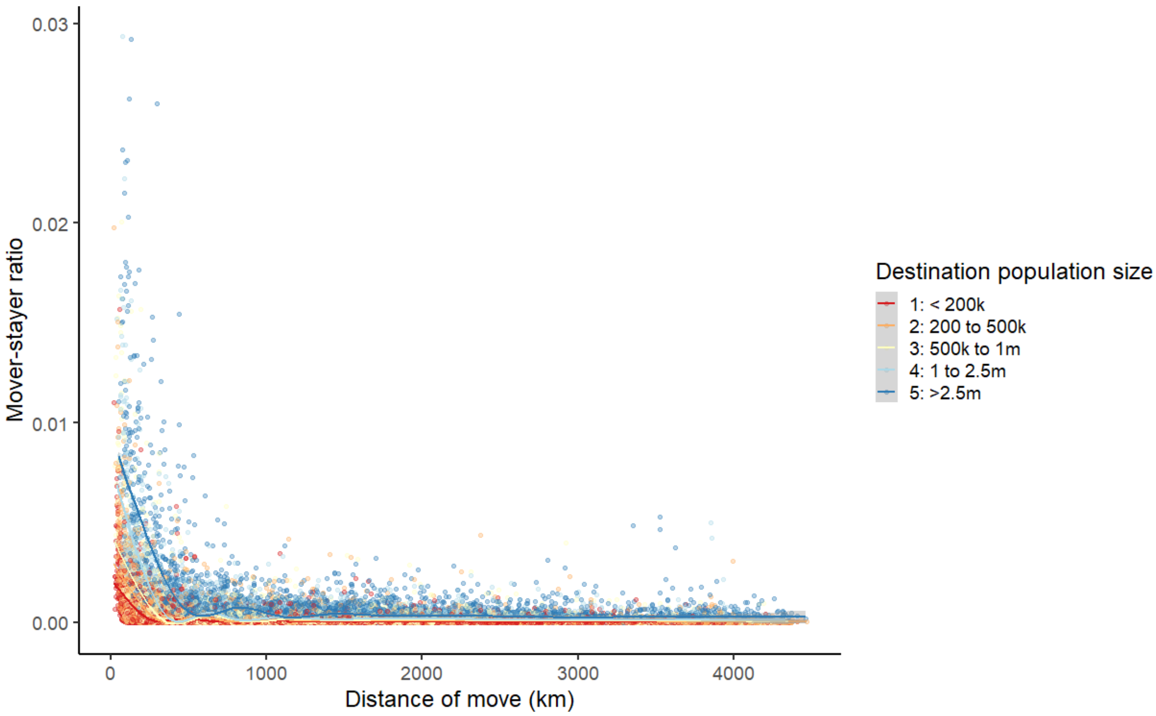
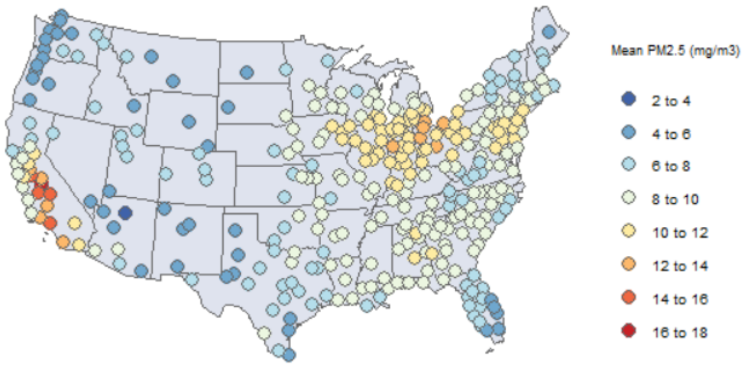
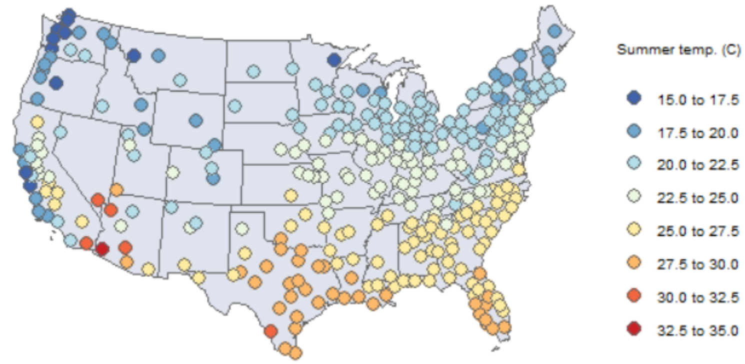


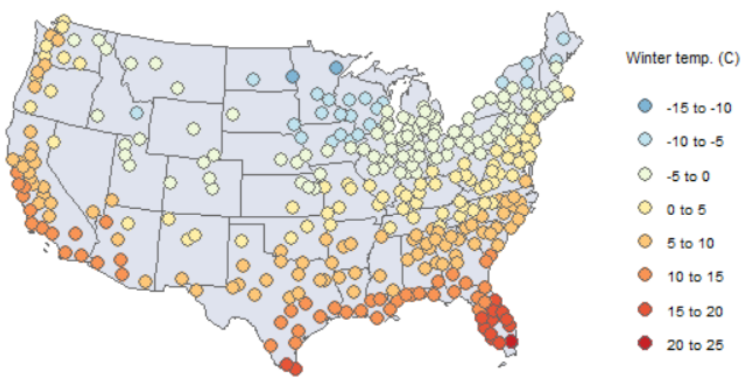
Figure 3: Gravity illustrated. Each point denotes an origin-destination-pair observation (N=118,336). Mover-stayer ratio is the population moving to an observation's destination divided/normalized by the population staying in the observation's origin. Two takeaways are evident and familiar: distance and destination size are key determinants of migration. Plotted curves are cubic splines over distance for destinations of different population sizes. Data: US Census Bureau, 2012-2016.



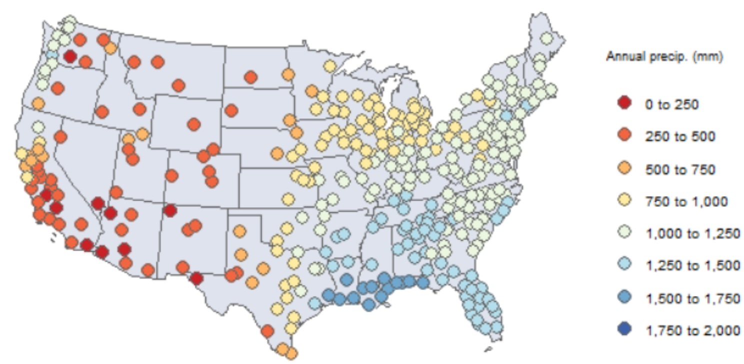
(a) Between-MSA variation in average PM2.5 ($\mu g/m^3$)



(b) Between-MSA variation in average summer temperature (Celsius)

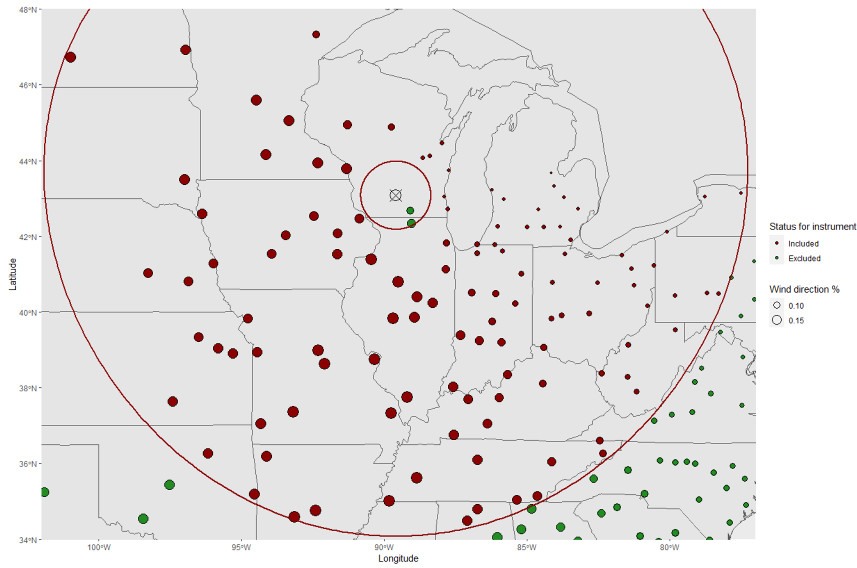


(c) Between-MSA variation in average winter temperature (Celsius)

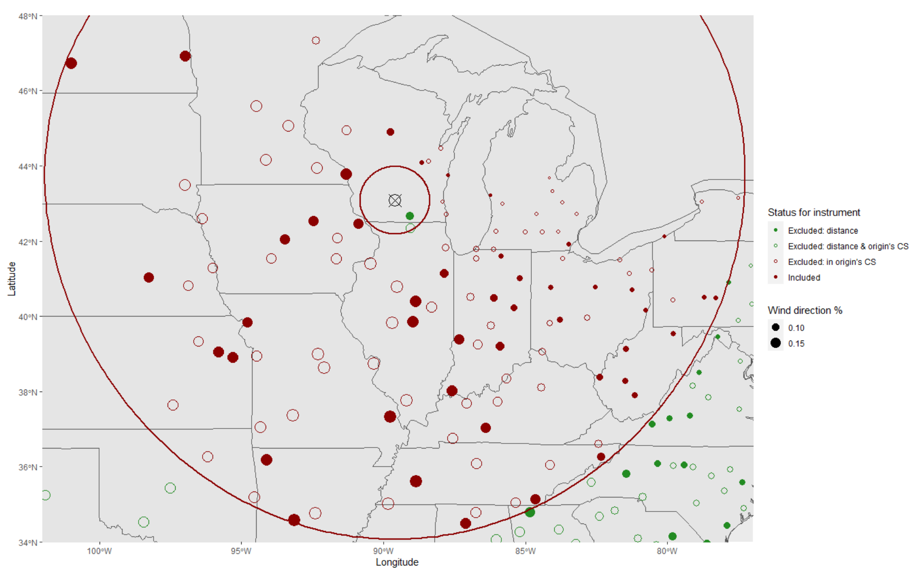


(d) Between-MSA variation in annual precipitation

Figure 4: Visualizing levels across space for environmental amenities of interest.



(a) IV: distance- and wind-weighted sum of incoming pollution from neighboring MSAs, not within 100km of the destination.



(b) IV: distance- and wind-weighted sum of incoming air pollution from neighboring MSAs that are not in the origin's choice set and not within 100km of the destination.

Figure 5: Visualization of logic behind our instrumental variable strategy. In both panels, the destination city is Madison, WI; the origin city (out of frame) is Tucson, AZ. Our instrument sums distance-weighted (and in some cases, wind-weighted) PM2.5 levels from cities that neighbor Madison. Across panels, MSAs less than 100km from Madison are excluded, as are those greater than 1000km away. See Table 6 for comparison of estimation results using different flavors of this IV.

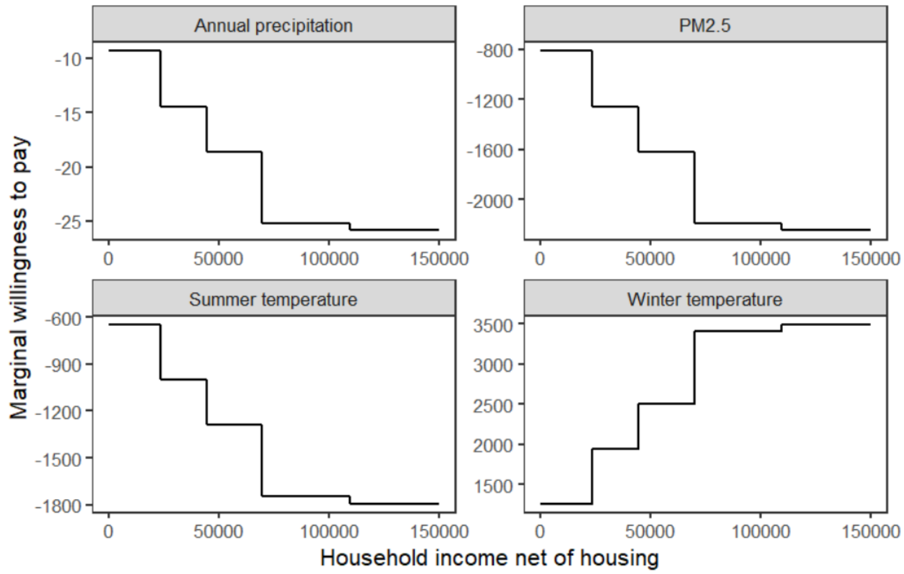


Figure 6: Tracing out the average marginal willingness to pay curve, by net income level and amenity. Parameters estimates taken from Table 6, column 4.

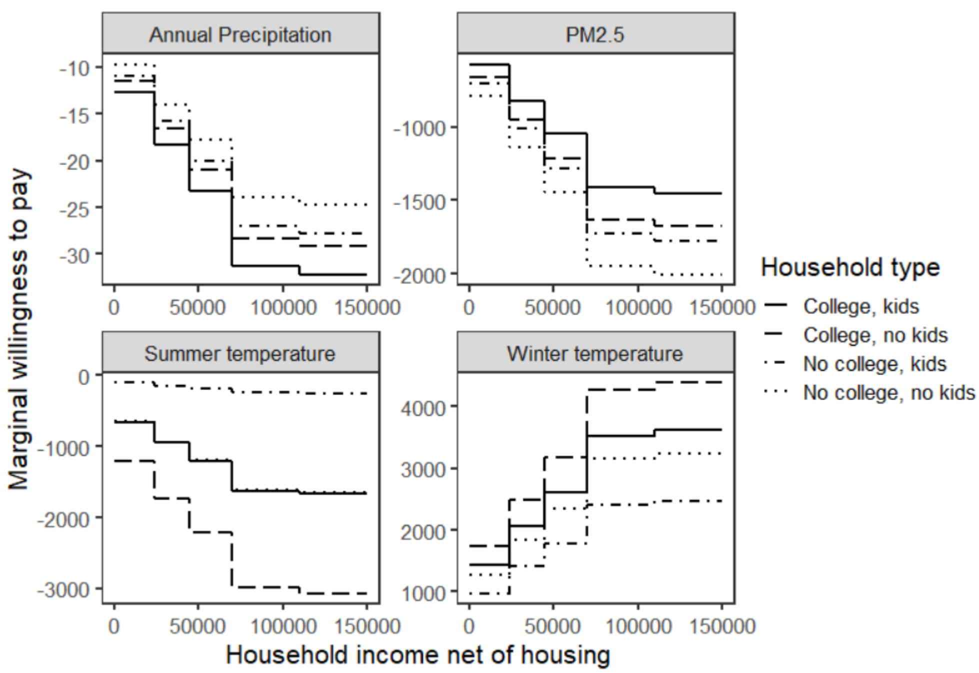


Figure 7: Tracing out marginal willingness to pay when preferences vary by observable demographics. Here, MWTP varies with the household head's education level and whether children are present in the household. Parameter estimates taken from Table 9.

	Mean	St.Dev	10th percen.	90th percen.	Source
Mean PM2.5 ($\mu g/m^3$)	8.649	2.088	5.691	11.230	Van Donkelaar et al (2012-16)
Annual Precipitation (mm)	974.4	353.2	390.5	1360.1	PRISM (1980-2010)
Mean Winter Temp ($^{\circ}$C)	3.311	6.421	-4.083	11.59	PRISM (1980-2010)
Mean Summer Temp ($^{\circ}$C)	23.52	3.251	19.69	27.57	PRISM (1980-2010)
Mean household income ($)	69,879	12,408	57,862	84,020	Census Bureau (2012-16)
Median household rent ($)	11,037	2,735	8,344	14,664	Census Bureau (2012-16)
Population	763,002	1,680,456	114,482	1,842,424	Census Bureau (2012-16)
MSA count:	344				

Table 1: Summary statistics for MSA-level variables of primary interest. All annual MSA average measures generated from remote sensing sources (climate, PM2.5) are weighted by population at the census tract level.

	Variable	Source
Geographic		
	Average Elevation	Authors' calculations
	Ruggedness (TRI)	Riley et al (1999)
	Coastal (dummy)	Authors' calculations
	Total Land Area (km^2)	IPUMS-NHGIS (2011)
	% Water Area	IPUMS-NHGIS (2011)
	% Developed Area	IPUMS-NHGIS (2011)
	% Urban Open Space	IPUMS-NHGIS (2011)
	% Forest Cover	IPUMS-NHGIS (2011)
Economic		
	Local Gov't Spending Index	Stansel (2013)
	Local Tax Index	Stansel (2013)
	Local Labor Index	Stansel (2013)
	Local unemployment rate	Census Bureau (2012-2016)
	% employed in the arts	Census Bureau (2012-2016)
	% employed in manufacturing	Census Bureau (2012-2016)
Transport		
	Average Commute Time	Census Bureau (2012-16)
	Public Transport Usage (% of pop)	Census Bureau (2012-16)
Urban Amenities		
	Libraries/km^2	IMLS (2016)
	Museums/km^2	IMLS (2016)
	Recreation Facilites/km^2	USDA (2015)
	Farmers' Markets/km^2	USDA (2015)
	Full-service Restaurants/km^2	USDA (2015)
	Fast-food Restaurants/km^2	USDA (2015)
	Malls & Supercenters/km^2	USDA (2015)
	Grocery Stores/km^2	USDA (2015)
Urban Disamenities		
	Annual Violent Crimes (per 100k)	FBI (2012-16)
	Annual Property Crimes (per 100k)	FBI (2012-16)
	Brownfields/km^2	EPA (2012)
	Superfund/km^2	EPA (2012)
	Annual Extreme Weather Events	NOAA (1996-2011)

Table 2: List of additional urban characteristics included as MSA-level amenities.

	Full Sample		Estimation Sample		Movers-only Sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Net HH income	75,958	81,897	76,019	81,754	62,172	71,035
Multiple earners in HH	0.393	0.488	0.394	0.489	0.354	0.478
Child in HH	0.484	0.500	0.482	0.500	0.375	0.484
HH head: Age	46.448	12.379	46.466	12.386	37.726	12.160
HH head: Female	0.436	0.496	0.435	0.496	0.457	0.498
HH head: Married	0.579	0.494	0.579	0.494	0.395	0.489
HH head: College grad	0.463	0.499	0.464	0.499	0.496	0.500
HH head: Black	0.094	0.292	0.094	0.292	0.104	0.305
HH head: Other non-white	0.122	0.328	0.123	0.328	0.147	0.354
HH head: Hispanic	0.131	0.337	0.131	0.337	0.143	0.350
HH gead: Reside in birth state	0.470	0.499	0.469	0.499	0.429	0.495
Mover: Out of state	0.017	0.130	0.017	0.129	0.145	0.352
Mover: Out of MSA	0.026	0.160	0.026	0.159	0.210	0.407
	N = 2,484,088		N = 248,409		N= 294,306	

Table 3: Key descriptive statistics: household microdata. Full sample includes all households with at least one member earning wages of at least $5000, and whose current (and previous, if applicable) residence falls in one of our study MSAs. Estimation sample is a 10% random draw of households from the full sample; this sample is used in most estimations involving microdata. Movers-only includes all households from the full sample who moved primary residence in the past year.

Least attractive MSAs	θ_j	MSAs with lowest churn	$\widehat{\epsilon_k}$
Parkersburg-Vienna, WV	-2.400	Parkersburg-Vienna, WV	-0.456
Columbus, IN	-2.235	Sebastian-Vero Beach, FL	-0.344
Kokomo, IN	-2.167	Kankakee, IL	-0.332
Kankakee, IL	-2.121	Monroe, MI	-0.332
Altoona, PA	-2.089	Dalton, GA	-0.326
Beckley, WV	-2.047	California-Lexington Park, MD	-0.306
Lebanon, PA	-2.034	Houston-The Woodlands-Sugar Land, TX	-0.303
Pittsfield, MA	-2.027	Altoona, PA	-0.294
East Stroudsburg, PA	-2.007	Johnstown, PA	-0.293
Johnstown, PA	-2.003	Kokomo, IN	-0.285
California-Lexington Park, MD	-1.971	Port St. Lucie, FL	-0.284
Williamsport, PA	-1.969	Deltona-Daytona Beach-Ormond Beach, FL	-0.274
Lima, OH	-1.969	Jackson, TN	-0.273
Sheboygan, WI	-1.945	North Port-Sarasota-Bradenton, FL	-0.271
Monroe, MI	-1.936	Cape Coral-Fort Myers, FL	-0.269
Kingston, NY	-1.858	Gainesville, GA	-0.263
Glens Falls, NY	-1.844	Punta Gorda, FL	-0.254
Dubuque, IA	-1.828	Michigan City-La Porte, IN	-0.252
Staunton-Waynesboro, VA	-1.824	Weirton-Steubenville-Wheeling, WV-OH	-0.240
Grand Island, NE	-1.821	Orlando-Kissimmee-Sanford-The Villages, FL	-0.240
Sumter, SC	-1.803	Lima, OH	-0.239
Madera, CA	-1.790	Hickory-Lenoir-Morganton, NC	-0.233
Gadsden, AL	-1.785	Lebanon, PA	-0.233
Jackson, TN	-1.784	Columbus, IN	-0.232
Mansfield, OH	-1.782	East Stroudsburg, PA	-0.232

Table 4: Stylized results from aggregate model (I). Left column lists the 25 MSAs with the smallest destination fixed effects (θ_j) resulting from our PPML gravity model estimation. Locations in Appalachia and the eastern Midwest abound. The right column lists the 25 MSAs with the smallest origin fixed effect, netting out the role of population ($\widehat{\varepsilon_k}$). Many names on the list look familiar, as unattractive places also tend to have low rates of out-migration. Additionally, we do note that many cities in Florida fall on this list - perhaps due to their heavy retirement populations.

Most attractive MSAs	θ_j	MSAs with highest churn	$\widehat{\epsilon_k}$
Houston-The Woodlands-Sugar Land, TX	1.026	Watertown-Fort Drum, NY	0.429
Dallas-Fort Worth-Arlington, TX	1.023	Ames, IA	0.378
Phoenix-Mesa-Scottsdale, AZ	0.952	Lawrence, KS	0.367
Atlanta-Sandy Springs-Roswell, GA	0.815	Amarillo, TX	0.333
Denver-...-Greeley, CO	0.800	Columbia, MO	0.308
Seattle-Tacoma-Bellevue, WA	0.761	Logan, UT-ID	0.299
Riverside-San Bernardino-Ontario, CA	0.686	Manhattan, KS	0.290
Miami-...-West Palm Beach, FL	0.635	San Diego-Carlsbad, CA	0.283
Los Angeles-Long Beach-Anaheim, CA	0.628	Clarksville, TN-KY	0.281
Chicago-Naperville-Elgin, IL-IN-WI	0.565	El Paso, TX	0.279
Tampa-St. Petersburg-Clearwater, FL	0.555	Great Falls, MT	0.272
Austin-Round Rock, TX	0.549	Missoula, MT	0.271
Orlando-...-The Villages, FL	0.548	Iowa City, IA	0.258
Las Vegas-Henderson-Paradise, NV	0.527	Lawton, OK	0.249
Washington-...-Alexandria, DC-VA-MD-WV	0.504	Las Cruces, NM	0.243
San Francisco-Oakland-Hayward, CA	0.494	Medford, OR	0.242
Portland-Vancouver-Hillsboro, OR-WA	0.465	Jacksonville-New Bern, NC	0.241
New York-Newark-Jersey City, NY-NJ-PA	0.463	Sioux Falls, SD	0.241
San Antonio-New Braunfels, TX	0.453	El Centro, CA	0.236
Boston-...-Nashua, MA-CT-NH	0.448	Lubbock, TX	0.230
Charlotte-Concord-Gastonia, NC-SC	0.368	Fayetteville, NC	0.225
Sacramento–Roseville–Arden-Arcade, CA	0.359	Ann Arbor, MI	0.218
Minneapolis-St. Paul-Bloomington, MN-WI	0.347	Cheyenne, WY	0.211
Philadelphia-...-Bridgeton, PA-NJ	0.341	Burlington-South Burlington, VT	0.211
Detroit-Warren-Dearborn, MI	0.253	New York-Newark-Jersey City, NY-NJ-PA	0.206

Table 5: Stylized results from aggregate model (II). Left column lists the 25 MSAs with the largest destination fixed effects (θ_j) resulting from our PPML gravity model estimation. Many Sun Belt cities top the list, but high-population coastal cities also rank high. The right column lists the 25 MSAs with the largest origin fixed effect, netting out the role of population ($\widehat{\varepsilon_k}$). Atop the list are a number of cities with a large university and military presence. Interspersed are a number of smaller cities in Mountain and Western states.

Panel A: 1st stage parameters	Estimate	S.E.
Income: Q1 (<23.7k)	0.3934	(0.0973)
Income: Q2 (<44.7k)	0.2550	(0.0569)
Income: Q3 (<69.9k)	0.198	(0.0398)
Income: Q4 (<109.7k)	0.1459	(0.0267)
Income: Q5	0.1422	(0.0147)
MC: birth state	-0.8202	(0.0473)
MC: birth div	-0.4216	(0.056)
MC: birth region	-0.2301	(0.0486)
MC: kids	-0.7328	(0.0333)
MC: dual earners	-0.213	(0.0333)
Observations	248409	
Log-likelihood	-53975.6	

	[1]: OLS		[2]: IV, preferred		[3]: IV, alternative 1		[4]: IV, alternative 2	
Panel B: 2nd stage parameters	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Annual precipitation (mm)	-0.0003	(0.0001)	-0.0004	(0.0001)	-0.0004	(0.0001)	-0.0004	(0.0001)
Average winter temp. ($^\circ$C)	0.0519	(0.0042)	0.0496	(0.0044)	0.0467	(0.0044)	0.0509	(0.005)
Average summer temp. ($^\circ$C)	-0.0296	(0.0065)	-0.0255	(0.007)	-0.0203	(0.0069)	-0.0277	(0.0081)
Average PM2.5 ($\mu g/m^3$)	-0.0139	(0.0075)	-0.032	(0.0146)	-0.0553	(0.0134)	-0.0221	(0.0211)
Origin FEs	Yes		Yes		Yes		Yes	
Additional amenity controls	Yes		Yes		Yes		Yes	
Moving costs: distance	Yes		Yes		Yes		Yes	
Observations	44090		44090		44090		44090	
IV F-stat			13.07		45.55		1.727	

Table 6: Household-level location choice model: parameterizing the mean MWTP for amenities of interest. First stage parameters (Panel A) estimated by maximum likelihood. Income scaled by $10k in first stage. Second stage parameters (Panel B) estimated by generalized least squares, with observations weighted by origin population divided by size of origin's choice set. Standard errors are het-robust. Column 1 of Panel B does not instrument for PM2.5. Column 2 uses an instrument derived from distance- and wind-weighted summation of PM2.5 from neighboring MSAs in the destination's distance bandwidth. Column 3 uses an instrument derived from distance-weighted-only summation of PM2.5 from neighboring MSAs in the destination's distance bandwidth and NOT in the origin's choice set. Column 4 uses and instrument derived from distance- and wind-weighted summation of PM2.5 from neighboring MSAs in the destination's distance bandwidth and NOT in the origin's choice set.

	[1] Stock Model		[2] Stock Model: distance		[3] Flow Model	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
1st stage parameter estimates						
Income: Q1 (<$23.7k)	0.3577	(0.0297)	0.2889	(0.0161)	0.3934	(0.0973)
Income: Q2 (<$44.7k)	0.3360	(0.0165)	0.2624	(0.0384)	0.2550	(0.0569)
Income: Q3 (<$69.9k)	0.2967	(0.0131)	0.2435	(0.0047)	0.198	(0.0398)
Income: Q4 (<$109.7k)	0.2454	(0.0109)	0.2269	(0.0061)	0.1459	(0.0267)
Income: Q5	0.1299	(0.0096)	0.1989	(0.0076)	0.1422	(0.0147)
2nd stage parameter estimates						
Annual precipitation (mm)	-0.0003	(0.0003)	-0.0007	(0.0003)	-0.0004	(0.0001)
Average winter temp. (°C)	0.0154	(0.0205)	0.0116	(0.0262)	0.0496	(0.044)
Average summer temp. (°C)	0.0154	(0.0337)	0.0819	(0.0432)	-0.0255	(0.007)
Average PM2.5 ($\mu g/m^3$)	0.0047	(0.0702)	-0.086	(0.0899)	-0.032	(0.0146)
ASCs identified from:	Population share		Population share		Origin migration share	
Additional amenity controls	Yes		Yes		Yes	
Moving Costs: birthplace	Yes		Yes		Yes	
Moving Costs: demographic	Yes		Yes		Yes	
Moving Costs: distance	No		Yes		Yes	
Origin FEs	No		No		Yes	
Log-likelihood (1st stage)	-391008.6		-78030.4		-53975.6	
Observations (1st stage)	248409		248409		248409	
Observations (2nd stage)	344		344		44090	

Table 7: Comparing the flow model to stock-based sorting models. First stage parameters estimated by maximum likelihood. Income scaled by $10k in first stage. Second stage parameters estimated by generalized least squares, with observations weighted by origin population divided by size of origin's choice set. In this table, we instrument for PM2.5 with our preferred IV across specifications (see column 2 of Table 6). Standard errors are het-robust. Column [1] is a purely stock-based model in the spirit of the literature; ASCs are estimated for each destination, origin does not play a direct role. Column [2] includes distance-binned migration costs, but still estimates ASCs for only destinations. Column [3] is our flow model - identical to column [2] of Table 6.

	[1]		[2]		[3]	
	Linear MU of income		Log MU of income		Sample: movers only	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
1st stage parameter estimates						
Net income (linear)	0.1554	(0.013)				
log(Net income)			0.7837	(0.0899)		
Income: Q1 ($<$ 18.6k)					0.3487	(0.0339)
Income: Q2 ($<$ 34.9k)					0.2627	(0.0201)
Income: Q3 ($<$ 55.4k)					0.2151	(0.0156)
Income: Q4 ($<$ 89.2k)					0.1871	(0.0113)
Income: Q5					0.1756	(0.0063)
2nd stage parameter estimates						
Annual precipitation (mm)	-0.0003	(0.0001)	-0.0004	(0.0001)	-0.0003	(0.0001)
Average winter temp. (°C)	0.0489	(0.0044)	0.0502	(0.0044)	0.0481	(0.0044)
Average summer temp. (°C)	-0.0248	(0.007)	-0.0267	(0.007)	-0.0262	(0.007)
Average PM2.5 ($\mu g/m^3$)	-0.0331	(0.0146)	-0.0362	(0.0146)	-0.0223	(0.0147)
Additional amenity controls	Yes		Yes		Yes	
Moving Costs: birthplace	Yes		Yes		Yes	
Moving Costs: demographic	Yes		Yes		Yes	
Moving Costs: distance	Yes		Yes		Yes	
Origin FEs	Yes		Yes		Yes	
Log-likelihood	-53982.3		-54008.9		-392518.1	
Observations (1st stage)	248409		248409		294306	
Observations (2nd stage)	44090		44090		44090	

Table 8: Robustness to alternative specifications and data. First stage parameters estimated by maximum likelihood. Income scaled by $10k in first stage in columns [1] and [3]. Second stage parameters estimated by generalized least squares, with observations weighted by origin population divided by size of origin's choice set. Instrument for average PM2.5 is preferred IV (see column [4] of Table 6). Column [1] imposes linear marginal utility (MU) of (net) income. Column [2] imposes log MU of income. Our preferred MU is a piecewise linear function by quintile. Column [3] studies estimates on a sample of only households that moved (either within- or between-MSAs) in the past year, using our preferred specification.

Panel A: HH preference heterogeneity	[1] Winter Temp Estimate	[1] Winter Temp S.E	[2] Summer Temp Estimate	[2] Summer Temp S.E	[3] Precipitation Estimate	[3] Precipitation S.E	[4] PM2.5 Estimate	[4] PM2.5 S.E
HH head college grad x ...	0.0162	(0.0079)	-0.0199	(0.0147)	-0.0063	(0.0127)	0.0046	(0.0179)
HH head older than 60 x ...	-0.0068	(0.0048)	0.0098	(0.009)	0.0069	(0.008)	-0.0187	(0.0106)
HH head younger than 30 x ...	0.0555	(0.0177)	-0.052	(0.0391)	0.0082	(0.0355)	-0.0199	(0.0507)
Children in HH x ...	-0.0107	(0.0065)	0.0186	(0.0159)	-0.0043	(0.0111)	0.0032	(0.0149)

Panel B: Homogenous HH parameters	Estimate	S.E.
Income: Q1 (<22.2k)	0.3564	(0.0871)
Income: Q2 (<43.5k)	0.2476	(0.0544)
Income: Q3 (<68.8k)	0.1948	(0.0386)
Income: Q4 (<108.6k)	0.1447	(0.0262)
Income: Q5	0.1406	(0.0147)

Panel C: Mean amenity preferences	Estimate	S.E.
Annual precipitation (mm)	-0.0003	(5.5e-5)
Average winter temp. (°C)	0.0455	(0.0044)
Average summer temp. (°C)	-0.0232	(0.007)
Average PM2.5 ($\mu g/m^3$)	-0.0282	(0.0146)

Additional amenity controls	Yes
Moving Costs, birthplace	Yes
Moving Costs, demographic	Yes
Moving Costs, distance	Yes
Origin FEs	Yes
Log-likelihood	-55518.10
1st stage observations	251474
2nd stage observations	44090

Table 9: HH heterogeneity in amenity preferences: note that this table shows results from a single model estimation. Panel A shows 16 parameter estimates for interactions of HH demographic characteristics and our environmental amenities of interest. Panel B shows parameter estimates for the marginal utility of income; this is held uniform across household types. Panel C shows parameter estimates for the baseline demographic group: a household with no children and a household head in the 30-59 age range without a college degree. First stage (panels A and B) estimated by maximum likelihood; second stage (panel C) estimated by generalized least squares with observations weighted by origin population divided by size of origin's choice set and using preferred IV for PM2.5. All standard errors het-robust.

	Quintile (at origin)	Estimate	S.E.	Implied MWTP*
	Q1	-0.0004	(0.0001)	-22
	Q2	-0.0004	(0.0001)	-19
Annual Precipitation	Q3	-0.0004	(0.0001)	-21
	Q4	-0.0003	(0.0001)	-16
	Q5	-0.0002	(0.0001)	-8
	Q1	0.0598	(0.0087)	3,006
	Q2	0.0661	(0.0083)	3,319
Average Winter Temperature	Q3	0.0635	(0.0131)	3,190
	Q4	0.0381	(0.0064)	1,913
	Q5	0.0355	(0.0074)	1,786
	Q1	-0.0548	(0.0115)	-2,755
	Q2	-0.0196	(0.0083)	-985
Average Summer Temperature	Q3	-0.0272	(0.0086)	-1,368
	Q4	-0.0218	(0.0113)	-1,096
	Q5	-0.0292	(0.013)	-1,469
	Q1	-0.0065	(0.0106)	-326
	Q2	-0.0358	(0.0107)	-1,797
Mean PM2.5	Q3	-0.0365	(0.0088)	-1,832
	Q4	-0.0174	(0.0099)	-872
	Q5	0.0057	(0.0152)	284
Additional amenity controls		Yes		
Moving Costs, all		Yes		
Origin FEs		Yes		

Table 10: Preference heterogeneity by origin. Here, we highlight second stage parameter estimates from a single model run. Destination amenity levels are interacted with quintile dummy variables for the origin's amenity level. This therefore studies how amenity preferences may vary with a household's previous/"status quo" level of the amenity. First stage parameters - and resulting ASCs used for estimation - are those from Table 6. Parameters estimated by generalized least squares, with observations weighted by origin population divided by size of origin's choice set. No instrument for PM2.5 in this regression. All standard errors are het-robust. *Implied MWTP shown for a household at the US median net income.

Appendix 1: Additional results

	[1]		[2]		[3]	
Dependent variable: $\widehat{\epsilon_k}$	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
% population aged <18	0.0046	(0.0027)				
% population aged 18-24	0.0045	(0.0019)				
% population aged 25-34	0.0370	(0.0062)				
% population aged 35-44	-0.0308	(0.0105)				
% population aged 45-54	-0.0413	(0.0091)				
% population aged 55-64	0.0362	(0.0079)				
% population aged >65	-0.0136	(0.0025)				
% Owner occupied			-0.0054	(0.0017)		
% HHs with mortgage			0.0025	(0.0011)		
% HHs moved in pre-1990			0.0041	(0.0026)		
% HHs moved in 1990-1999			0.0035	(0.0071)		
% HHs moved in 2000-2009			-0.0114	(0.0041)		
% HHs moved in since 2010			0.0085	(0.0012)		
% HHs: <25k income					0.0036	(0.0026)
% HHs: 25-50k income					-0.0114	(0.0066)
% HHs: 50-75k income					0.0106	(0.0109)
% HHs: 75-100k income					-0.0036	(0.0135)
% HHs: 100-150k income					0.0051	(0.0129)
% HHs: 150-200k income					0.0005	(0.0203)
% HHs: >200k income					-0.0039	(0.0088)
Observations	344		344		344	
R^2	0.3933		0.315		0.0247	

Table A1.1: An exploratory analysis of how MSA's churn ($\widehat{\epsilon_k}$) varies with regional demographic and economic characteristics. See Equations (12)-(13) and Tables 4-5 for more explanation on this measure. Independent variables are measured at the MSA level and taken from 2012 (beginning of the study period). Column [1] examines whether life-cycle patterns make MSAs more or less "sticky". Column [2] considers the role of housing tenure, ownership, and household debt. Column [3] looks at whether income constraints are holding back free mobility. In all columns, estimation by OLS; standard errors are het-robust.

	Estimate	Implied MWTP ($)
Mean Net Income	4.26e-5	
	(2.38e-5)	
Annual Precipitation	-0.0003	-8.19
	(5.87e-05)	
Mean Winter Temperature	0.0195	457.75
	(0.005)	
Mean Summer Temperature	-0.0274	-643.19
	(0.008)	
Mean PM2.5	-0.0176	-413.15
	(0.008)	
Moving costs: distance	Yes	
Other amenity controls	Yes	
Origin FEs	Yes	
Observations	117648	

Table A1.2: Measures of MWTP for environmental amenities derived from an aggregate-data-only model. Estimation by Pseudo-Poisson Maximum Likelihood; standard errors are het-robust. Net income enters linearly in this stylized specification. MWTP is therefore a simple ratio of parameters.

ln(wage income)	Mean across MSAs	St. dev.
Age	0.0458	0.0068
Ageˆ2	-0.0004	0.0001
Nonwhite: Asian (base = Caucasian)	-0.0500	0.1068
Nonwhite: Black	-0.1074	0.0856
Nonwhite: Other	-0.0454	0.0609
Hispanic	-0.0607	0.0688
Graduate/Professional Degree (base = College degree)	0.2402	0.0496
HS diploma or some college	-0.2392	0.0438
Less than HS	-0.3743	0.0756
Unmarried	-0.1087	0.0288
Female	-0.1754	0.0340
No Children in HH	-0.0381	0.0220
Dahl probabilty	-0.1664	0.3309
Dahl probabilityˆ2	0.2372	0.4930
MSA Count:	344	
Year FEs	Yes	
Industry FEs	Yes	
Occupation FEs	Yes	
Weekly hours worked (binned)	Yes	
Observation level:	Household	

Table A1.3: Summary of counterfactual wage income regression parameters. Separate hedonic wage regressions were run in each of $N = 344$ MSAs. We summarize the resulting distribution of parameter estimates for each characteristic by the mean and standard deviation of values across MSAs. Counterfactual wage predictions then omit the control function parameters from Dahl (2003). All money measures were deflated by yearly CPI.

ln(Housing costs)	Mean across MSAs	St. dev.
Home is owned	-0.1022	0.1143
2 bedrooms (base = 1 bedroom)	0.0009	0.2758
3 bedrooms	0.1562	0.2855
4 bedrooms	0.2916	0.2863
5 bedroom	0.4232	0.29
6+ bedrooms	0.5164	0.2957
2 rooms (base = 1 room)	0.0034	0.3128
3 rooms	0.0197	0.2938
4 rooms	0.0261	0.2955
5 rooms	0.05	0.2987
6 rooms	0.1079	0.3044
7 rooms	0.1718	0.3092
8 rooms	0.23	0.3103
9 rooms	0.3198	0.3125
10 rooms	0.3391	0.3165
11+ rooms	0.4261	0.3306
Single family detached	0.0927	0.1041
Single family attached	0.0647	0.1065
2-family	-0.0167	0.1172
3-family	-0.0304	0.0899
5-9 family	-0.0342	0.0837
20-49 family	0.0225	0.1288
50+ family	0.0293	0.1541
Mobile home, boat, tent, other	-0.419	0.1498
Built: 1940-49	-0.0346	0.0961
Built: 1950-59	-0.0105	0.0959
Built: 1960-69	0.0088	0.1025
Built: 1970-79	0.0443	0.1052
Built: 1980-89	0.102	0.1048
Built: 1990-99	0.205	0.1197
Built: 2000-04	0.3138	0.1347
Built: 2005-09	0.3669	0.1425
Built: 2010-16	0.3581	0.1665
MSA Count:	344	
Year FEs:	Yes	
Observation level:	Household	

Table A1.4: Summary of counterfactual housing expenditure regression parameters. Separate hedonic price regressions were run in each of $N = 344$ MSAs. We summarize the resulting distribution of parameter estimates for each characteristic by the mean and standard deviation of values across MSAs. All money measures were deflated by yearly CPI.

	Outer bandwidth									
	800km		900km		1000km		1100km		1200km	
Inner bandwidth	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
80km	-0.0317	(0.0134)	-0.0305	(0.0134)	-0.0293	(0.0134)	-0.0291	(0.0134)	-0.0289	(0.0134)
90km	-0.0448	(0.0141)	-0.0433	(0.0141)	-0.0419	(0.0141)	-0.0416	(0.0141)	-0.0414	(0.0141)
100km	-0.0352	(0.0146)	-0.0335	(0.0146)	-0.032	(0.0146)	-0.0317	(0.0146)	-0.0314	(0.0147)
110km	-0.0543	(0.0144)	-0.0523	(0.0144)	-0.0504	(0.0144)	-0.05	(0.0144)	-0.05	(0.0144)
120km	-0.0714	(0.0148)	-0.069	(0.0149)	-0.0667	(0.0149)	-0.0662	(0.0149)	-0.0659	(0.0149)

Table A1.5: Robustness: instrument sensitivity to distance bandwidths. Inner bandwidth is the lower distance bound for a neighboring MSA's PM2.5 level to be included in the instrument's generation. Similarly, outer bandwidth is the upper distance bound. The displayed PM2.5 parameter estimates and standard errors are those resulting from the second stage model as in Table 6: the only things changing are the distance bandwidths used to generate the instrument.

	1 unit PM	1 degree C winter	1 degree C summer	Details
Sinha et al. (2018): full sample		\$1,787	-\$2,200	
Sinha et al. (2018): movers only sample		\$2,477	-\$2,795	Mean MWTP of households that moved in last 5 years
Hamilton and Phaneuf (2015): 2000 PM levels	-\$690			Median PM10 from paper: 33.87 μg/m^3
Hamilton and Phaneuf (2015): current PM levels*	-\$1,756			
Bayer et al (2009): 2000 PM levels	-\$448			Median PM10 from paper: 36.0 μg/m^3
Bayer et al (2009): current PM levels*	-\$1,212			
Theising and Phaneuf (2020)	-\$1,616	\$2,507	-\$1,287	Using Q3 (median) marginal utility of income

Table A1.6: Comparison with MWTP estimates from other equilibrium sorting models of a related nature. Notes: All mean MWTP values shown in table are in 2016 dollars. Sinha et al (SCC) appear to use 2000 prices, for which CPI to 2016 is 1.4. Hamilton and Phaneuf (HP) use 1990 prices, for which CPI to 2016 prices is 1.86. Bayer et al (BKT) use 1983 prices, for which CPI to 2016 prices is 2.42. We infer MWTP at the median sample incomes of \$25,683 (HP) and \$15,679 (BKP). *When converting between PM2.5 and PM10 for comparable "current PM levels", we use the conversion rate of PM2.5 = .65 PM10. So for the mean PM2.5 level in our sample of cities (8.649), this implies a mean PM10 level of 13.306. Thus, we calculate implied MWTP for a 1 unit PM10 reduction at the base level of 13.306.