Valuation of Local Public Goods: Migration as Revealed Preference for Place

Brett Close and Daniel J. Phaneuf*

University of Wisconsin - Madison

March 15, 2017

Abstract

We develop a residential sorting model based on a panel of county-to-county migration flows to estimate the marginal valuation of air pollution. Our approach exploits annual cross-sectional variation in migration flows to estimate mean location utilities at the county level, while flexibly controlling for moving costs. The mean utilities provide a time-varying, county level index of residential attractiveness. We then use panel variation in county characteristics to decompose mean utility into observable and unobservable components using county fixed effects, which allows us to estimate the marginal value of local amenities. In our application to air pollution, we use an instrumental variables approach to provide robust evidence that the concentration of fine particulate matter (PM 2.5) is a disamenity that negatively impacts location decisions. In our preferred specification, we find that the median household is willing to forgo 2.65 percent of annual income for a 1 $\mu g/m^3$ decrease in fine particulates.

*Contact Information: bclose@wisc.edu (Close), dphaneuf@wisc.edu (Phaneuf).
We would like to thank participants at seminars at the University of Wisconsin and attendees at the AERE Summer Conference for helpful feedback. All remaining errors are our own.

1 Introduction

Local public goods such as clean air, quality public schools, and safe neighborhoods can provide significant economic value. Though they are not purchased directly, households do in fact pay for local public goods indirectly through spatially differentiated housing rents, wages, and local taxes that vary with location characteristics. In choosing where to live, households implicitly select their bundle of local public goods, meaning they must strike a balance between location-specific amenities, and the private consumption goods they will be able to purchase with the income they can earn in the location, net of rents and local taxes.

Selecting a location involves more than trading off income for desirable local amenities, however. Accessing a different bundle of local public goods requires moving, which is highly disruptive. There are pecuniary and nonpecuniary costs; the latter includes, for example, leaving established social networks and familiar routines. Yet every year thousands of households do migrate. What role do local public goods play in this decision? And what can we learn about how much people value local amenities by their migration choices? In this paper, we address these questions by developing a residential sorting model using a panel dataset of migration flows between US counties. By exploiting a new dataset and methodology, we estimate mean utilities for counties across space and time, and then explain the variation in utility using time-varying county characteristics. In doing so, we contribute to the literature on residential sorting models, non-market valuation, estimating quality of life indices, and the broader literature on discrete-choice demand analysis.

Scholars have been interested in using location decisions to understand the value of local public goods at least since Tiebout (1956). Modern approaches rely on either the hedonic property value model, or residential sorting models. Hedonic approaches, based on Rosen (1974), estimate an equilibrium price function for housing, based on characteristics of properties, including neighborhood attributes such as local public goods. The gradient of the price function with respect to an amenity is then taken to be the marginal willingness to pay, as housing options varying along the price function should display this tradeoff.

Sorting models attempt to remedy a number of drawbacks of hedonic approaches, and have been described in-depth by Kuminoff et al. (2013). The primary drawback of the hedonic model is that it uses a costless mobility assumption to link the equilibrium price function to households' marginal willingness to pay. With costly migration, the price gradient is not necessarily equal to marginal willingness to pay. Horizontal sorting models address this problem by adopting a discrete-choice framework that directly estimates preference parameters. Agents select among a finite number of possible locations, trading off place-specific amenities against local wages and rents. Estimation relies on the econometric framework developed in Berry et al. (1995), which also accommodates household heterogeneity in preferences. The intuition from Roback (1982) implies that variation in local amenities will be reflected in wages and rents, as prices adjust to maintain equilibria in the housing and labor markets. As such, migration in response to variation in wages and amenities across space and time can be used to quantify household tradeoffs and marginal willingness to pay for local amenities.

The horizontal sorting framework was first developed in Bayer et al. (2004) and subsequently extended in a number of applications. Notable examples include Bayer et al. (2007) for public school quality, Klaiber and Phaneuf (2010) for local land use, and Depro et al. (2015) for environmental justice. Each of these papers model residential sorting within a metropolitan area. Sorting models have also been applied at the national scale to estimate the marginal cost of air pollution. The first contribution was from Bayer et al. (2009), who recognized the importance of including moving costs to reduce bias in their estimates of marginal willingness to pay. This is because the gains from moving to a higher-amenity location must compensate for the higher rents, lower wages, *and* moving costs. Hamilton and Phaneuf (2015) build on Bayer et al. (2009), marrying it to the micro-level sorting approaches by using a nested-logit formulation to represent two-stage budgeting of the choice of an optimal metropolitan area and optimal neighborhood within it. Other examples of sorting models include Tra (2010), Bayer and McMillan (2012), and Tra (2013). Properties of the general estimator have been derived in Bayer and Timmins (2005) and Bayer and Timmins (2007).

All of the models discussed above rely on observing where households' live at specific points in time, except Depro et al. (2015), who use population changes between census tracts in 2000 and 2010. Since the actual migration path is not known, they assume that the observed distribution of location choices represents a spatial equilibrium, which was generated by households' previously-made location choices. In contrast, our analysis relies on individual observation of migration flows, so that that our preference parameters are identified off current-period decisions to move or not move, and do not require a market-level spatial equilibrium to hold. We contribute methodologically to the sorting model literature by developing a generalization of the discrete choice approach that accommodates our flow data, exploits a multiple year panel, and flexibly accounts for costly moving.

Our application examines the marginal willingness to pay for air quality, which has been the focus of previous hedonic property value and sorting applications. Research in both the economics and health literatures shows that air pollution has significant impacts on human health and is an aesthetic disamenity in some locations. Concentrations of fine particulate matter can cause respiratory and cardiovascular problems, leading to elevated fatality risks in vulnerable populations, as well as increased symptoms for people with asthma and similar conditions. For the Obama Administration's Clean Power Plan, which is nominally a plan to reduce carbon dioxide emissions that contribute to climate change, the EPA estimates that as much as 60% of the benefits will actually come from reductions in health problems from co-pollutants, rather than reductions in green house gases (Environmental Protection Agency, 2016). Current estimates of the marginal willingness to pay for particulate matter reductions are based on the distribution of pollution and prices across space at a given time, or identified using changes in prices and pollution within locations across time. Common spatial units include counties or metropolitan statistical areas (MSAs) for national scale studies. For example, Chay and Greenstone (2005) use the Clean Air Act Amendments non-attainment status to instrument for total suspended particulates in a hedonic price model, using counties as the unit of observation. Bayer et al. (2009) and Hamilton and Phaneuf (2015) use changes in PM10 (particulates less than 10 μm in diameter) between 1990 and 2000 at the MSA level, along with 1990 and 2000 snapshots of residential location shares, to estimate the value of pollution reductions using sorting models. We contribute to this literature by providing a new estimate for the marginal value of fine particulate reductions, using a novel source of data and research design.

To estimate our sorting model, we rely on a relatively untapped data source within the economics literature: the Internal Revenue Service (IRS) data from tax filings that track the number of households filing in different counties, in successive years. The data provide counts of county-to-county migration flows for the years 2005-2006 through 2010-2011, which we combine with repeated cross-sectional variation in county level wages to estimate the role that wage differentials play in explaining annual migration flows. Our reliance on county-level data for the entire US allows us to evaluate migration-wage tradeoffs using a broader sample than is possible using metropolitan areas as the unit of analysis. In addition, because migration flows vary between pairs of counties as well as year over year, we are able to identify mean utilities that exhibit panel variation for each county. This allows us to estimate the determinants of mean utility using a second stage panel regression with county fixed effects. The IRS data also allows us to separate the extent to which counties draw migrants in, versus retaining current residents. With this we can estimate welfare impacts separately for migrants and non-migrants, and separately test whether changes in pollution levels impact households' move versus stay decisions, and conditional on moving, where a household goes.

We find strong evidence that fine particulate pollution has an impact on household migration decisions. Across a wide array of specifications, we find that PM 2.5 (particulates less than 2.5 μm in diameter) functions as a disamenity to households. Our preferred specification implies a non-migrant household earning \$50,000 per year would be willing to forgo \$1325 dollars to reduce their exposure to PM 2.5 by 1 $\mu g/m^3$ - about one-third of a standard deviation in concentrations. We also find some evidence that PM 2.5 concentrations have a negative impact on the attractiveness of a county to migrating households, although this effect is smaller than for non-migrant households.

By estimating a panel of mean utilities based on migration flows, our paper also contributes to the quality of life literature that focuses on general explanations for the relative attractiveness of different locations. This literature has used both sorting and hedonic approaches. In a sorting application, Kahn (1995) uses observations of households' wages and rents at their selected location to impute wages and rents at alternative locations. These are used as explanatory variables in a discrete choice location model that estimates location fixed effects in a cross section, which are then interpreted as an ordinal ranking of city quality. Work by Albouy and coauthors has examined quality of life in a hedonic framework, focusing on adjustments for taxation and non-rent prices (Albouy, 2012), and commuting costs (Albouy and Lue, 2015). Albouy et al. (2016) use the framework to estimate marginal valuation for climate amenities. Our mean utilities can be used to examine similar questions, without reliance on imputing location-specific variables or an assumption of free mobility.

The paper is organized as follows. We begin by laying out the behavioral model of household migration decisions and marginal valuations. We then present the data used in the estimation and explain sources of cross-sectional and panel variation, followed by a description of the application of the model to air pollution, including an instrumental variables strategy to address endogeneity of air pollution. This is followed by a description of the mechanics of estimating the model and the innovations we developed to the traditional horizontal sorting model. Finally, we provide results from the first-stage estimation of county-level mean utilities and the second-stage estimation of marginal willingness to pay to avoid air pollution, and offer concluding thoughts.

2 Behavioral Model

2.1 Choice Behavior

Consider household i residing in county k in year t. It faces a choice of which county to live in during the following year among the full set of counties $j = 0, \ldots, J$, including the choice to stay in k. We assume that the utility from residing in location j is,

$$\begin{aligned} V_{kj}^{it} &= \delta_j^t + \mathbb{1}\{k = j\}\alpha_j^t + (log(wage_j^t) - log(wage_k^t))\gamma_{wage}^t + Z_{kj}^{it}\gamma^t + \epsilon_j^{it} \\ &= \tilde{V}_{kj}^{it} + \epsilon_j^{it}, \end{aligned} \tag{1}$$

where δ_j^t is the mean utility (alternative-specific constant or ASC) that anyone selecting county j receives; $\mathbb{1}\{\cdot\}$ is the indicator function; α_j^t is the mean utility that agents in county k receive from staying in county k beyond what migrants receive; $wage_j^t$ is the wage in county j in year t; γ_{wage}^t is the marginal utility of income in year t; Z_{kj}^{it} is the set of factors whose value depends on characteristics of the agent and characteristics of county k relative to county j, and may include interactions; γ^t is a vector of marginal utilities of Z; and ϵ_j^{it} is an idiosyncratic term reflecting characteristics of county j for agent i. The household selects county j as its residence in the following year if

$$V_{kj}^{it} = \max_{l=0,\ldots,J}\{V_l^{it}\}. \tag{2}$$

This model has three direct output parameters, δ, γ, and α. We can interpret them as follows: δ_k represents the mean representative utility of the decision to move into county k in the following year, γ_g represents the slope of representative utility with respect to element g of Z, and α_k represents the mean representative utility among residents of county k to staying in county k, apart from the benefits to newcomers.

Of these parameters, γ is a typical coefficient in a discrete-choice model. The primary alternative-specific constant δ is similarly familiar from discrete choice models and a larger value of δ makes a county unambiguously more desirable. As a choice-context-specific and alternative-specific constant, α is less familiar and somewhat more ambiguous. A larger value of α could be either positive, denoting a high quality of life due to rich local social networks or location amenities that require local knowledge, or negative, due to chronic poverty that makes leaving difficult. As such, it can be interpreted as a location-specific fixed (utility) cost of moving away, beyond the cost embedded in γ. This contrasts with the traditional Roback-style approach which assumes free mobility, and complements the approach introduced in Bayer et al. (2009) that includes a fixed cost to moving. Generally, we can think of δ as a measure of attractiveness of the county, α as a measure of retentiveness, and $-\gamma$ as variable moving cost. The sum $\alpha + \delta$ is the mean utility of residents of continuing to reside in the county. The main parameters of interest in the model are $\alpha + \delta$ and δ, as these correspond to the mean utility of non-migrants and migrants, respectively.

2.2 Marginal Valuation

The county-level mean utility parameters can be decomposed as:

$$\delta_j^t = X_j^t \beta_1 + \phi_{1j} + \psi_{1j}^t + \xi_{1j}^t, \tag{3}$$

$$\alpha_j^t = X_j^t \beta_3 + \phi_{3j} + \psi_{3j}^t + \xi_{3j}^t, \tag{4}$$

and

$$\alpha_j^t + \delta_j^t = X_j^t \beta_2 + \phi_{2j} + \psi_{2j}^t + \xi_{2j}^t, \tag{5}$$

where X is a set of characteristics of counties that affect people's valuation of locations, each β is a parameter vector indicating the partial derivative of the value component with respect to the elements of each X, and $\phi_{\cdot j}$ and $\psi_{\cdot j}^t$ are county and (possibly regionally varying) time fixed effects.

Given these decompositions, we can can rewrite the utility function for a migrant as,

$$V_{kj}^t = X_j^t \beta_1 + \phi_{1j} + \psi_{1j}^t + \xi_{1j}^t + (log(wage_j^t) - log(wage_k^t))\gamma_{wage}^t + Z_{kj}^t \gamma^t + \epsilon_j^t, \tag{6}$$

and,

$$V_{kj}^t = X_j^t \beta_2 + \phi_{2j} + \psi_{2j}^t + \xi_{2j}^t + Z_{kj}^t \gamma^t + \epsilon_j^t, \tag{7}$$

for a non-migrant. Taking a total differential for a migrant gives,

$$dV_j^t = dX_j^t\beta_1 + d(log(wage_j^t))\gamma_{wage}^t + dZ_{kj}^t\gamma^t. \tag{8}$$

Setting all differentials to zero except for elements g of X and the wage element,

$$0 = dX_{gj}^t\beta_{1g} + d(log(wage_j^t))\gamma_{wage}^t \implies \left(\frac{d(log(wage_j^t))}{dX_{1gj}^t}\right)\Bigg]_{dV=0} = -\frac{\beta_{g1}}{\gamma_{wage}^t}, \tag{9}$$

which is the marginal willingness to pay for changes in X_{gj}.

Equation 3 is the structural equation for the mean utility of migrants, whereas Equation 5 is the structural equation for the mean utility of non-migrants. Because the vast majority of households do not migrate in any given year, Equation 5 is more relevant for welfare analysis. Equation 3 still has welfare implications for households that do migrate, and is relevant for a test of the Tiebout hypothesis that amenities drive location decisions: $\beta_{1g} > 0$ implies people "vote with their feet" for amenity g. Equation 5 provides the converse hypothesis:$\beta_{2g} > 0$ implies people are more likely not to leave a place with a high value for X_g.

2.2.1 Marginal Utility of Income and Willingness to Pay

We estimate the marginal utility of income in the first stage using the difference in the log of average wages. This allows us to recover estimates based on variation across counties within a year, reflecting the tradeoffs people are making. We then use these estimates to calculate the marginal willingness to pay.

The marginal utility of income is thus estimated separately for each year in the panel. In order to compare across years, we directly estimate marginal willingness to pay for each element of X by running our second stage model on $\tilde{\delta}_j^t = \frac{\delta_j^t}{\gamma_{wage}^t}$, or the equivalent expression for other parameters. Then the coefficient on X becomes a direct estimate of $\frac{\beta_1}{\gamma_{wage}}$.

In addition to improving the estimate of marginal willingness to pay, it also provides scale normalization to the model, ensuring that the error distribution is the same across years and all values are comparable. Because the overall scale of utility is irrelevant in any discrete choice model, parameters are only identified up to the scale of the unobserved utility term; in a logit model, this scaling is done automatically by the distributional assumption. But we still cannot distinguish between a model with parameter θ from a model with parameter estimate $\frac{\theta}{\sigma}$ for any positive σ (which would simply reflect a different scale for the error term). Therefore, parameters estimated from different data cannot be directly compared without accounting for scale, but ratios of parameters can be directly compared because the scale parameter will cancel out. Thus, while we cannot directly compare the magnitude of δ_j^t and δ_j^s, we can compare the magnitudes of $\frac{\delta_j^t}{\gamma_{wage}^t}$ and $\frac{\delta_j^s}{\gamma_{wage}^s}$ (which, due to location normalization reflect comparisons to the mean utility of the reference location in years t and s). We assume that the true value of γ_{wage} is fixed (other elements may change due to changes in technology or travel costs over the period), and each estimate of γ^t_{wage} simply reflects differences in unobserved factors in different years.

2.3 Identifying Variation

By relying on migration flows and different migration shares for different origin counties, the model uses rich variation for identification. In the first stage of the model, all parameters are identified off of variation across counties in a year, specifically in the variation in differences between origins and destinations across all the origin counties. γ is identified off of the differences between counties with respect to each other. If Z contains distance elements, the identification comes from the fact that the set of distances to other counties is unique. For example, Los Angeles county in California will be a more desirable destination for people in Washington state than people in Maine because of its closer proximity. In terms of wages, a destination with moderate wages will be more desirable to residents of low-wage counties than residents of high-wage counties.

The parameter δ is identified off of migrants moving from county to county. Larger migration flows for destinations with equivalent values of $Z\gamma$ indicate larger values of δ, and this variation is what identifies the parameter. $\alpha + \delta$ is identified off of the portion of the population that chooses to remain in a county, relative to the desirability of all the other counties available; α is then based off of the values of the other three parameters, as discussed below.

In the second stage, β_1, β_2 and β_3 are identified off of panel variation across years using county and time fixed effects models. That is, we look at how δ_j and α_j vary across years, taking into account differences across counties within a year.

3 Data

3.1 Migration Flow Data

The primary data source for this analysis is a data set of migration flow estimates produced by the Internal Revenue Service (IRS) for the years 2005-2006 through 2010-2011. For each year, the IRS produces tabulations for each county of the number of tax returns, the number of exemptions and the total adjusted gross income associated with tax payer IDs that were filed in each other county in the previous year. That is, the agency produces a dataset that quantifies the number of households (proxied by returns) and individuals (proxied by exemptions) who moved from one county to another and the number who stayed, and the income (proxied by the adjusted gross income) of those households. In principle, each county pair would then be reflected twice in the dataset: migrants from A to B, and from B to A.

The dataset does not report the direct county-to-county numbers for combinations of counties between which fewer than 10 returns moved. Thus, the reported values are censored. The censored returns are reported in aggregate for each county to others in the same state, and to each of four regions: Northeast, Midwest, South, and West. If any of these regional categories had fewer than 10 returns, they are aggregated into an "all other regions" category. It is not possible to distinguish a county pair between which no one moved and one between which fewer than 10 households moved. See Section 5.2.3 for details on how the censoring is addressed.

The IRS migration data are relatively untapped in the economics literature. While these data have been used within the geography literature, the only study within economics that we are aware of uses them to describe overall migration trends (Molloy et al., 2011).

The IRS migration data provide rich cross-sectional and temporal variation that our model is able to exploit. Figure 1 shows rates of in-migration, out-migration, and net migration for counties between 2005 and 2006, binned into nine quantiles with light yellow corresponding to low values and dark blue corresponding to high values. There is clear cross-sectional variation, both between regions and between counties near each other. For example, both in-migration and out-migration rates are high in the inter-mountain West, but there is sufficient variation between these rates that net migration was much more positive (more in-migration than out-migration) in Nevada counties than Utah counties.

In addition to cross-sectional variation, there is also significant panel variation. Figure 2 shows net migration for 2005-06 and for 2009-10, again with yellow indicating more out-migration and blue indicating more in-migration. For 2005-06 rates are generally negative in the Plains states, but positive in the Southwest. However, the rates are relatively lower in the Southwest in 2009-10, while many of the Plains counties see relatively high rates of in-migration, although with significant variation county-to-county. This panel variation can be seen, for example, in California, as shown in Figure 3. Coastal counties had much higher rates of net in-migration in 2009-10 compared with 2005-06.

Table 1 provides summary statistics for each year of the panel, including the average, and 5% and 95% quantiles of out-migration and in-migration rates; the correlation between in-migration and out-migration rates at the county level; the average number and standard deviation of observed destination counties for each source county; and the average and standard deviation distance households moved. We can see that approximately 8% of households migrate in a year, meaning that approximately 92% remain in their current county. The correlation between in- and out-migration is high for all years except 2005, meaning that the overall pattern is of high turnover in some locations, as opposed to movement primarily away from certain counties into others. The anomaly in 2005 is likely due to disruption from Hurricane Katrina. Migrants from each county move to an average of about 27 counties each year (among counties where at least 10 households are observed to migrate), at an average distance of about 100 miles.

These figures and tables highlight the cross sectional and panel variation in aggregate migration at the county level, but one of the major strengths of our model is the ability to deal with county-to-county migration, rather than just aggregate migration rates. That is, it allows for variation in the in-migration rates for each county, from each other county. Figure 4 shows 2005-06 migration rates from Los Angeles and Santa Clara Counties in California, and New York County in New York (the Borough of Manhattan in New York City). Counties in white have fewer than 10 returns moving between them in those years. All three share some similar patterns with the majority of migration going to the West Coast, Florida, and the Northeast corridor from Washington, D.C. to Boston. But there is signficant variation between them as well. Los Angeles County, which had the highest number of migrants of any county, has a more broad-based out-migration than the others. Santa Clara County migrants moved primarily along the West Coast, while New York County migrants moved primarily along the Northeast Corridor and to Florida.

We can see the differences more clearly by focusing on migration to California counties, as shown in Figure 5. Unsurprisingly, New York County migrants moved to a much smaller set of counties than Los Angeles County or Santa Clara County migrants, focusing along the coasts. Los Angeles County migrants focused more on Southern California, while Santa Clara County migrants moved more around the Bay Area, and to Northern California and the Central Valley.

In addition to migration flows, we draw population estimates from the IRS dataset as the category "All U.S. Migrants and Non-migrants." Thus, it does not correspond to the traditional definition of the population of the county, but reflects the total sample observed. In a small number of cases, the number of migrants to counties and regions was less than the population due to aggregation between counties. In these cases, the shortfall in observed migration was added to the "all other regions" category.

Certain returns are not included in the dataset, including those that are filed extremely late, which, according to IRS documentation, tend to be complicated returns associated with extremely wealthy households. In addition, households that do not file federal income taxes will not be reflected.

3.2 Control Data

We include controls in both the first and second stage estimations. First-stage controls, which must vary at the county-to-county level, include distances; indicators for whether the counties are in the same Core-based Statistical Area (CBSA, roughly equivalent to an MSA), state, and region; relative log wages; and the squared difference in relative rurality. Second-stage controls, which vary at the county level year over year, include industry composition controls, air pollution, housing market characteristics, and unemployment rate. Distance data come from NBER county-to-county distance files. County-to-region and county-to-state distances were calculated as the mean distance between a county to the set of counties in a region, or within the same state, which were not directly reflected in the IRS database. Thus, while the distance between any two counties is fixed throughout time, the distance from a county to a region will vary between years, based on which counties in that region had observed migration and which did not.

Over the study period, certain county definitions changed due to renaming, or the creation or dissolution of existing counties. In the case of renaming, the new county inherited distance from the old county. In the case of creation, the new county's distances to any county is calculated as the average of the distances of the predecessor counties. Any counties that have been dissolved prior to a year have no observed migration and are dropped from the analysis.

Industrial composition and wage data come from the Bureau of Labor Statistics Quarterly Census of Employment and Wages. Figure 6 shows average wage quantiles at the county level for 2005, with high wages shown in dark blue and low wages shown in light yellow. Again, there is significant variation in the cross-section. This includes regional variation, but also variation between near-by counties. While there is a clear rural-urban difference, with high wages in many of the urban areas, there is also variation between rural areas. For example, wages in rural eastern Oregon are much higher than in rural northern Nebraska.

Rurality index data come from Waldorf and Kim (2015). Their Index of Relative Rurality is a composite index reflecting county population, population density, distance to a major population center, and percent of land built up. Values are between zero and one and reflect the average placement of the county between the maximum and minimum value for each of the criteria. We rely on the 2010 estimates and use the absolute difference between counties' indexes.

Industry composition controls include the number of establishments, aggregate employment levels, and annual wages overall and for the set of industries for which data are available at the county level and that we assume to be most polluting, based on NAICS classification: 23 (Construction), and 31-33 (Manufacturing). Unemployment data are the county level annual average from the Bureau of Labor Statistics.

Air pollution data are from the EPA's AirData project.[1] The variable of interest is the average concentration of particulate matter (PM 2.5). These value are not monitored in all counties, so these analyses restrict the data set. Figure 7 shows quantiles of average PM 2.5 concentrations in 2009. Higher concentrations are dark blue and lower concentrations are light yellow, while counties in white do not have any EPA monitoring of PM 2.5. The figure demonstrates both the geographic extent of the monitoring and the amount of variation within regions. Monitoring is not limited to urban counties, but includes both rural and urban counties from across the country.

There is variation in particulate concentration across time as well as across space. This can be seen in concentrations for California counties in 2005 and 2009, as shown in Figure 8. Concentrations in northern California have decreased, relative to concentrations in the Central Valley and southern California.

Housing data come from the American Community Survey conducted by the US Census bureau. Data include median rents, median property values, average number of rooms in a dwelling, and the average year built.

Emissions data used in the calculation of the instrument, as discussed in Section 4.1, come

[1] http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html

from EPA's national emissions inventory (NEI). Because county-level data are not available for all years, annual data at the state level are used for each emissions source tier for interpolations.

Table 2 provides summary statistics for the 2005 cross section, while Table 3 gives measures of variation in the panel, including the overall standard deviation, the "within" standard deviation (the standard deviation after subtracting the county mean), and the mean range within a county, reflecting another measure of total variation at the county level. PM 2.5 concentrations and unemployment rate have significant variation within the panel, as well as overall in the cross section. Wages, on the other hand, have significantly more variation in the cross-section than in the panel.

4 Application: Particulate Pollution

We use our model framework to investigate marginal valuation of fine particulate pollution for the years 2005 through 2011. Fine particulate matter, particles with diameter less than $2.5\mu m$ known as PM 2.5, has been shown to have serious implications for health. It can cause irritation of the eyes, nose, throat, and lungs. When inhaled these very small particles can travel deep into the lungs and enter the bloodstream. Exposure has been linked to increased health problems such as asthma and heart disease as well as increased hospital admissions. Fine particulate matter is produced by fuel combustion as well as created by chemical reactions in the atmosphere, especially due to the presence of sulfur dioxide (SO2).

We look at concentrations of fine particulate matter to estimate their impact on the structural parameters $\tilde{\alpha} + \tilde{\delta}$ and $\tilde{\delta}$ and to estimate average marginal willingness to pay to avoid PM 2.5 concentrations. Our first-stage structural equation, representing Equation 1, includes: the log of the distance between the two counties; indicators for whether the counties are in the same core-based statistical area (CBSA), and census region; and the difference in log of average wages. Our second-stage structural equations of interest are,

$$\begin{aligned} \tilde{\delta}_j^t &= \beta_{1,pm} PM_j^t + \beta_1' X_j^t + \phi_{1j} + \psi_{1j}^t + \xi_{1j}^t, \\ \tilde{\alpha}_j^t + \tilde{\delta}_j^t &= \beta_{2,pm} PM_j^t + \beta_2' X_j^t + \phi_{2j} + \psi_{2j}^t + \xi_{2j}^t, \end{aligned} \tag{10}$$

where $\beta_{1,pm}$ and $\beta_{2,pm}$ are the parameters of interest, representing the proportion of income that households are willing to trade off to receive a unit decrease in pollution concentrations; PM_j^t is the average PM 2.5 concentration in county j in year t; X_j^t is a vector of controls for job market and housing conditions, and industry composition within the county; β_1, and β_2 are coefficients on the controls; and ξ_{1j}^t and ξ_{2j}^t are disturbances.

The controls included in X_j^t include industry controls, housing controls, and the unemployment rate. Industry controls include the number of private employers, average employment, and average wages for all private businesses and for establishments classified as construction or manufacturing. Housing controls include the median rent, the median housing value, the median number of rooms per home, and the median year built.

Because much of the pollution in a location is produced by economic activity, such as manufacturing and construction, which is correlated with job market outcomes, pollution and job market outcomes are likely to be endogenous and direct estimation of Equation 10 by linear panel methods is likely to lead to biased estimates of the parameters of interest. That is, having more factories producing at higher levels means a county has more jobs that pay well, which is desirable, but also more pollution, which is not. Direct estimation of the impact of pollution on desirability will likely yeild inconsistent estimates because PM_j^t is likely correlated with ξ_{1j}^t and ξ_{2j}^t.

4.1 Instrumental Variables Approach

To address this challenge, we develop an instrument for PM 2.5 concentrations, adopting a practice used elsewhere, including Hamilton and Phaneuf (2015). The logic of the instrument is that local emissions are likely to be correlated with positive job market outcomes, but some of the pollution in a location is due to emissions elsewhere that flow in via weather patterns. This pollution is unlikely to be correlated with local job market outcomes or other components of Equation 10, conditional on included fixed effects.

This instrument is calculated by aggregating emissions at the county level for large point source polluters with smoke stacks greater than 500 feet high, primarily power plants, which are monitored by the EPA. Pollutant inflow is then calculated based on a matrix of transfer coefficients describing the fraction of emissions that will flow from the point source to each county in the country. Because a major source of PM 2.5 pollution is particles that are created from SO2, which is emitted primarily from coal-fired power plants, we use SO2 emissions from all counties outside of a band around the county of 100 miles, with dead-bands of 30, 50, and 75 miles as robustness checks.

Because the NEI does not produce estimates each year at the source level, we interpolate values for 2006, 2007, 2009, and 2010 based on NEI data for 2005, 2008, and 2011. Because the 2008 and 2011 databases do not include the stack height at the facility level, we estimate high-stack emissions based on the distribution of stack heights in each county from 2005. We use a second-order Taylor polynomial with the second derivative estimated based on NEI state-level emissions for each source tier (type of emissions source). So for year t, state s and emissions tier r, we estimate the proportional second derivative in the emissions rate as,

$$\Delta_{srt}^2 = \frac{E_{sr,t+1} + E_{sr,t-1} - 2E_{srt}}{E_{srt}}, \tag{11}$$

where E_{srt} is the emissions from state s, tier r and year t. The first derivative of emissions for county k is estimated based on the county-level emissions in the next year for which data are available, $\bar{t}$, and the most recent year for which data are available, $\underline{t}$ (for example, the first derivative for 2006 is based on 2005 and 2008), as,

$$\Delta_{krt} = \frac{E_{kr\bar{t}} - E_{kr\underline{t}}}{3} \tag{12}$$

Then the emissions for county k are estimated for the year for which data are not available as,

$$\hat{E}_{krt} = E_{kr\underline{t}} + \Delta_{krt}(t - \underline{t}) + \frac{\left(.5(t-\underline{t})^2\Delta^2_{s(k)r\underline{t}}E_{kr\underline{t}} + .5(\bar{t}-t)^2\Delta^2_{s(k)r\bar{t}}E_{kr\bar{t}}\right)}{2}, \tag{13}$$

where $s(k)$ is the state of county k. These county-level emissions are then multiplied by a transmission factor calculated by the EPA that estimates the quantity of emissions in a county that flow to each other county, based on prevailing weather patterns. For each receptor county j, these individual contributions are summed up across source counties outside of a deadband distance. That is, inflow SO2 is calculated as,

$$SO2^t_{-j} = \sum_{k=0}^{J} \mathbb{1}\{dist_{kj} > D\} b_{kj}\hat{E}_{krt}, \tag{14}$$

where $dist_{kj}$ is the distance between county k and county j, D is the deadband distance, and b_{kj} is the transfer coefficient between county k and county j.[2]

Figure 9 shows the quantiles of the calculated values of the instrument for New York County, NY and Allegheny County, PA, home to Pittsburg, in 2005 from surrounding states. For New York County, much of the inflow comes from western Pennsylvania. But for Allegheny county, the sources in Western Pennsylvania are within the deadband and excluded in the calculation of the instrument. Although these emissions sources are far distant, the instrument is quite strong, as shown in Section 6.3.

The structural equation for PM 2.5 pollution is then,

$$PM^t_j = \Gamma_{1,SO2}SO2^t_{-j} + \Gamma'_1 X^t_j + \mu^t_{1j}, \tag{15}$$

where $SO2^t_{-j}$ is SO2 inflow from far-away counties, $\Gamma_{1,SO2}$, and Γ_1 are the projection coefficients for SO2 inflow and the vector of controls discussed above, respectively. This gives the reduced-form estimating equations,

$$\begin{aligned} \tilde{\delta}^t_j &= \beta_{1,pm}\widehat{PM}^t_j + \beta'_1 X^t_j + \phi_{1j} + \psi^t_{1j} + \nu^t_{1j}, \\ \tilde{\alpha}^t_j + \tilde{\delta}^t_j &= \beta_{2,pm}\widehat{PM}^t_j + \beta'_2 X^t_j + \phi_{2j} + \psi^t_{2j} + \nu^t_{2j}, \end{aligned} \tag{16}$$

[2] An early study using this source receptor matrix is Shadbegian et al. (2007). These authors worked with EPA staff and analysts at Abt Associates to document appropriate use of the matrix, which was originally described in Latimer (1996), with additional technical information in Abt Associates (2000). Bayer et al. (2009) and Hamilton and Phaneuf (2015) received necessary files and documentation from Wayne Gray, which we have used for this paper. Documentation, including email correspondence between developer Douglas Latimer and Shadbegian et al., are available upon request.

where $\widehat{PM}_j^t$ is the predicted PM 2.5 concentration from Equation 15, which is uncorrelated with stochastic disturbance terms ν_{1j}^t and ν_{2j}^t.

5 Estimation

5.1 Model Likelihood

Based on the behavioral model laid out in 2.1, assuming that ϵ_j^{it} is distributed type 1 extreme value, the probability that agent i decides to reside county j in the following year is,

$$Pr_j^{it} = \frac{exp(\tilde{V}_j^{it})}{\sum_{l=0}^J exp(\tilde{V}_l^{it})}. \tag{17}$$

Now, consider not only agent i in county k but N_k agents living in county k, who are identical up to the idiosyncratic terms and act according to the behavioral model laid out above. Because agents are identical we can rewrite $Z_{kj}^{it} = Z_{kj}^t$ and $Pr_j^{it} = Pr_{kj}^t$. The likelihood of observing a migration flow for county k from year t to year $t+1$ is,

$$L_k = \prod_{i=1}^{N_k} Pr_{j(i)}^{it} = \prod_{i=1}^{N_k} Pr_{k,j(i)}^t = \prod_{j=0}^{J} (Pr_{kj}^t)^{M_{kj}}, \tag{18}$$

where $j(i)$ is the county selected by agent i and M_{kj} is the number of agents observed moving from county k to county j. The second equality is due to agents being identical within a county, and the third equality from the probabilities for a destination being the same within the same origin county. Taking the log gives the log-likelihood,

$$\mathcal{L}_k = \sum_{j=0}^{J} M_{kj} * ln(Pr_{kj}^{it}). \tag{19}$$

The log-likelihood of the full set of national migration is then the sum of the log-likelihood contributions from each source county,

$$\mathcal{L} = \sum_{k=0}^{J} \mathcal{L}_k = \sum_{k=0}^{J}\sum_{j=0}^{J} M_{kj} * ln(Pr_{kj}^t). \tag{20}$$

Rewriting $M_{kj} = N_k * m_{jk}$ in terms of population N_k and migration share m_{kj}, gives the log-likelihood function for the model:

$$\begin{aligned}\mathcal{L}(\delta,\alpha,\gamma|M) = & \sum_{k=0}^{J} N_k * \left(\sum_{j=0}^{J} m_{kj} * ln(Pr_{kj}^t)\right) \\ = & \sum_{k=0}^{J} N_k * \left(\sum_{j=0}^{J} m_{kj} * ln\left(\frac{exp(\tilde{V}_j^{it})}{\sum_{l=0}^{J} exp(\tilde{V}_l^{it})}\right)\right) \\ = & \sum_{k=0}^{J} N_k * \left(\sum_{j=0}^{J} m_{kj} * \left(\delta_j^t + \mathbb{1}\{k=j\}\alpha_j^t + \gamma^t Z_{kj}^t - \right.\right. \\ & \left.\left.\ln\left(\sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma^t Z_{kl}^t)\right)\right)\right)\end{aligned} \tag{21}$$

where δ is full set of δ_j for $j = 0, \ldots, J$, α is the full set of α_j for $j = 0, \ldots, J$, and M is the full set of M_{kj} for $k = 0, \ldots, J$ and for $j = 0, \ldots, J$. To normalize the model, fix $\delta_0 = 0$, so the dimension of the estimable parameter δ is J.

5.2 Contraction Mapping

With more than 3000 counties in the United States, direct estimation of the maximum likelihood estimator of $\{\delta, \alpha, \gamma\}$ is infeasible. The contraction mapping algorithm due to Berry et al. (1995), hereafter BLP, provides a tractable alternative in the case of a vector of market shares, but this model includes $J+1$ vectors of market shares, each of length $J+1$, making direct adoption of this approach impossible. Fortunately, the logic of the approach can be adapted to provide a solution.

The general logic of the BLP approach is that the first-order conditions for the maximum likelihood estimator in a logit-type model imply that the predicted market shares equal the observed market shares. The approach splits estimation into two parts: first, holding constant the coefficients on exogenous variables that vary at the individual level, estimate the values of the alternative specific constants via a simple algorithm that is guaranteed to converge to a fixed point that maximizes the log likelihood. Second, fix the value of the alternative-specific constants and estimate the coefficients on the individual variables by gradient or other traditional methods. Our model requires two generalizations: two sets of alternative-specific constants, and a matrix of market shares rather than a vector of market shares.

5.2.1 New Alternative-Specific Constant The first generalization is straight forward. The parameter δ functions as the traditional alternative-specific constant in the BLP formulation, while α is new. The value of the full vector δ is relevant to choice probabilities across all choice occasions (i.e. in every county), whereas α_k is only relevant for agents currently in county k. Fortunately, there is a closed-form solution for the maximum-likelihood estimator of each α_k

from the first-order conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \alpha_k^t} &= N_k * \left(\sum_{j=0}^{J} m_{kj}^t * \left(\mathbb{1}\{k=j\} - \frac{exp(\delta_k^t + \mathbb{1}\{k=l\}\alpha_k^t + \gamma Z_{kk})}{\sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma Z_{lk})} \right) \right) \\
&= N_k \left(m_{kk}^t * \left(1 - \frac{exp(\delta_k^t + \mathbb{1}\{k=l\}\alpha_k^t + \gamma Z_{kk})}{\sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma Z_{lk})} \right) \right. \\
&\quad \left. -(1 - m_{kk}^t) * \left(\frac{exp(\delta_k^t + \mathbb{1}\{k=l\}\alpha_k^t + \gamma Z_{kk})}{\sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma Z_{lk})} \right) \right) \\
&= m_{kk}^t - \frac{exp(\delta_k^t + \mathbb{1}\{k=l\}\alpha_k^t + \gamma Z_{kk})}{\sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma Z_{lk})} \\
&= m_{kk}^t - Pr_{kk}^t = 0.
\end{aligned} \tag{22}$$

That is, the maximum likelihood estimator for α_k is the value that predicts non-migration perfectly. The second-to last line of Equation 22 implies that

$$\begin{aligned}
m_{kk}^t = &\quad \frac{exp(\delta_k^t + \alpha_k^t + \gamma Z_{kk})}{\sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma Z_{lk})} && \Longrightarrow \\
m_{kk}^t \sum_{l=0}^{J} exp(\delta_l^t + \mathbb{1}\{k=l\}\alpha_l^t + \gamma Z_{lk}) = &\quad exp(\delta_k^t + \alpha_k^t + \gamma Z_{kk}) && \Longrightarrow \\
m_{kk}^t \sum_{l \neq k}^{J} exp(\delta_l^t + \gamma Z_{lk}) = &\quad (1 - m_{kk}^t) exp(\delta_k^t + \alpha_k^t + \gamma Z_{kk}) && \Longrightarrow \\
\frac{m_{kk}^t}{1 - m_{kk}^t} \sum_{l \neq k}^{J} exp(\delta_l^t + \gamma Z_{lk}) = &\quad exp(\delta_k^t + \alpha_k^t + \gamma Z_{kk}),
\end{aligned} \tag{23}$$

which can be solved to find the maximum-likelihood estimator,

$$\hat{\alpha_k^t} = ln \left(\frac{m_{kk}^t}{1 - m_{kk}^t} \sum_{l \neq k}^{J} exp(\delta_l^t + \gamma Z_{lk}) \right) - \delta_k^t - \gamma Z_{kk}. \tag{24}$$

5.2.2 Generalized Contraction The BLP methodology is attractive because it is guaranteed to find the maximum likelihood estimate of the parameters (given suitable starting values), and has low computational requirements. The first is because the algorithm is a contraction and will lead to a fixed point, which is equal to the maximum likelihood estimate. The second is because the algorithm relies on very simple calculations. In this section, we describe the traditional approach, lay out a generalization for the current model, and show that the generalization still constitutes a contraction.

The traditional application of the BLP estimator involves only the parameters δ and γ, but not α, and proceeds in two parts. The maximum-likelihood estimate in a logit model has the property that the predicted probabilities (in the form of predicted market shares) equal the observed probabilities (in the form of observed market shares). This provides an intuitive logic to the estimator. Given a value of γ and a starting estimate δ^0, an updated estimate of δ is calculated as

$$\delta^{r+1} = \delta^r - \left(ln(\sigma) - ln(\hat{\sigma}^r) \right), \tag{25}$$

where $\hat{\sigma}^r$ is the vector of the model's predicted market shares (based on δ^r and γ), and σ is the vector of observed market shares. This algorithm is repeated until a fixed point is found in δ, which occurs when the predicted market shares are equal to the observed market shares. This is guaranteed to converge to a fixed point because Equation 25 constitutes a contraction. Then δ is held fixed, and the estimate of γ is found via gradient or other methods. These two steps, the contraction on δ and the gradient search on γ, are repeated until a fixed point is found in the two parameter sets, which is the maximum likelihood estimate.

In the national migration model, both $\hat{\sigma}$ and σ take on matrix values (i.e. one vector for each source county). To accommodate this, the contraction can be updated to the form,

$$\begin{aligned} \delta_j^{r+1} &= \delta_j^r + \left(ln\left(\frac{\sum_{k\neq j} N_k * m_{kj}}{\sum_{k\neq j} N_k}\right) - ln\left(\frac{\sum_{k\neq j} N_k * Pr_{kj}(\delta^r|\gamma)}{\sum_{k\neq j} N_k}\right)\right) \\ &= \delta_j^r + \left(ln\left(\sum_{k\neq j} N_k * m_{kj}\right) - ln\left(\sum_{k\neq j} N_k * Pr_{kj}(\delta^r|\gamma)\right)\right) \\ &= \delta_j^r + S_j(\delta^r), \end{aligned} \tag{26}$$

where the time dependence has been dropped for clarity, and $S(\delta^r)$ is the step for iteration r. The value of δ is updated based on the total number of people moving into a county versus the number predicted to move into it, the average national migration share, rather than an origin-county-specific migration share for the destination county.

Berry et al. (1995, Appendix I) showed conditions under which a function constitutes a contraction and showed that their algorithm satisfies those conditions. The theorem is as follows: consider metric space $(\mathbb{R}^J, d)$ with $d(x,y) = \|x-y\|$ and let $f\colon \mathbb{R}^J \to \mathbb{R}^J$ have the following properties:

1. $\forall x \in \mathbb{R}^J$, $f(x)$ is continuously differentiable, with $\forall j$ and l, $\frac{\partial f_j(x)}{\partial x_l} \geq 0$ and $\sum_{l=1}^J \frac{\partial f_j(x)}{\partial x_l} < 1$.
2. $\min_j \inf_x f(x) \equiv x > -\infty$.
3. There is a value, $\bar{x}$, with the property that if for any j, $x_j \geq \bar{x}$, the for some l (not necessarily equal to j), $f_l(x) < x_l$.

Then there is a unique fixed point, x_0, to f in $\mathbb{R}^J$. Further, let the set $X = [\underline{x}, \bar{x}]^J$, and define the truncated function, $\hat{f}\colon X \to X$, as $\hat{f}_j(x) = min\{f_j(x), \bar{x}\}$. Then $\hat{f}(x)$ is a contraction of modulus less than one on X.

To show that $f(\cdot) = S(\cdot)$ satisfies this definition, we show that property 1 holds. The other two properties follow exactly the proof in Berry et al. (1995). For property 1:

$$\begin{aligned}\frac{\partial S_j(\delta_j)}{\partial \delta_j} &= 1 - \frac{\sum_{k\neq j} N_k * \frac{\partial Pr_{kj}(\delta^r|\gamma)}{\partial \delta_j}}{\sum_{k\neq j} N_k * Pr_{kj}(\delta^r|\gamma)} \\ &= 1 - \frac{\sum_{k\neq j} N_k * \left(Pr_{kj} - \left(Pr_{kj}\right)^2\right)}{\sum_{k\neq j} N_k * Pr_{kj}} \\ &= \frac{\sum_{k\neq j} N_k \left(Pr_{kj}\right)^2}{\sum_{k\neq j} N_k * Pr_{kj}} > 0,\end{aligned} \tag{27}$$

and for $l \neq j$

$$\begin{aligned}\frac{\partial S_j(\delta_l)}{\partial \delta_l} &= -\frac{\sum_{k\neq l} N_k * \frac{\partial Pr_{kj}(\delta^r|\gamma)}{\partial \delta_l}}{\sum_{k\neq l} N_k * Pr_{kj}(\delta^r|\gamma)} \\ &= \frac{\sum_{k\neq l} N_k * Pr_{kl} * Pr_{kj}}{\sum_{k\neq l} N_k * Pr_{kl}} > 0.\end{aligned} \tag{28}$$

Each of these partial derivatives is the weighted sum of probabilities. And because one of the counties is held out as an outside option to provide normalization of the model, for any k, $\sum_{j=1}^{J} Pr_{kj} < 1$. Therefore, the full sum is a weighted sum of probabilities that sum to strictly less than 1:

$$\sum_{l=1}^{J} \frac{\partial S_j(\delta_j)}{\partial \delta_l} = \sum_{l=1}^{J} \frac{\sum_{k\neq l}(N_k * Pr_{kl}) * Pr_{kj}}{\sum_{k\neq l}(N_k * Pr_{kl})} < 1, \tag{29}$$

and the second part of the first property holds.

5.2.3 Censoring and Zeros The elements laid out above in Subsection 2.3 are in principal all that are needed for estimating the model. But in the present empirical application, there is censoring in the data that must be dealt with. This leaves the reported migration between many pairs of counties within a year (and the true migration for many, no doubt) as zero. This is not a problem for the algorithm as long as each county has some in-migration, but there are cases where counties do not receive measurable in-migration from any county in a year, but do include observations of out-migration. Rather than drop these counties from the analysis and lose their observations of out-migration which are used in identifying parameters for other counties, we utilize the "numerical patch" technique from Timmins and Murdock (2007) for estimation of the model. This amounts to adding a very small increment to each migration value and each predicted probability and equivalently adjusting the denominator of Equation 26. For 2005-06, roughly 99% of the approximately 9.8 million cells in the migration matrix were zeros, but roughly 98.5% of locations decisions were observed at the county-level, rather than an aggregate level, and 162 counties had no observed in-migration. This adjustment only affects the parameter values of counties with no observed in-migration, which are not used in the second-stage analysis.

5.2.4 Numerical Refinements While Equation 26 is a contraction that is guaranteed to lead to a fixed point, it may converge slowly. This makes intuitive sense as the model is trying to fit approximately J^2 migration shares with approximately $2 * J$ parameters, compared to a traditional BLP model where J shares are fit with approximately J parameters. To aid in the estimation, we adopt a hybrid approach: changes in δ are always in the same direction as in Equation 26, but the step size is shortened in some cases. The algorithm becomes,

$$\delta_j^{r+1} = \delta^r + \hat{l}^r \left(ln \left(\sum_{k \neq j} N_k * m_{kj} \right) - ln \left(\sum_{k \neq j} N_k * Pr_{kj}^r \right) \right) = \delta^r + \hat{l}^r S^r, \tag{30}$$

where $\hat{l}^r \in [0, 1]$ is a step length modifier.

Rather than continue the contraction in Equation 26 until a fixed point is found, at each step of the algorithm the log-likelihood is calculated. After the step S^r is calculated, the log-likelihood is then recalculated for $\delta^r + S^r$. If the log-likelihood has improved, the new value of δ is accepted with $\hat{l}^r = 1$ and the algorithm proceeds to the next iteration. If the log-likelihood has not improved, then the optimal step length modifier is calculated. Because this step length modifier is constrained between zero and one, the algorithm is still a contraction, but avoids taking steps that are not improving the likelihood.

6 Results

This section presents the findings of the migration model applied to PM 2.5 pollution concentrations. General patterns from the first-stage estimation are presented in Section 6.1. They are consistent with stylized facts about what locations attract people, as well as the findings in the Quality of Life literature. Direct estimates of a Quality of Life index are presented in Section 6.2. Results from the second-stage estimation of marginal willingness to pay to avoid PM 2.5 pollution are presented in Section 6.3. We find a significant willingness to pay to avoid PM 2.5 pollution that is robust to changes in specification.

Because estimation of the model proceeds in two stages, with the second stage being a two-stage instrumental variables model itself, it is easy to create confusion in discussing results. To avoid this confusion, we refer to the overall first stage involving the estimation of δ, α, and γ as the first stage; the estimation of the structural equation for the endogenous regressor (commonly called the “first-stage") as the endogenous projection; and the estimation of the reduced-form relationship to estimate the parameters of interest as the reduced-form. The last two together are the second-stage.

6.1 First-Stage Results

Figure 10 shows a map of counties split into nine categories based on the quantile in $\tilde{\delta}$ for 2005 and 2010. Darker blue corresponds to a higher value and yellow to a lower value. The map indicates attractive counties along the West Coast, in the Southwest, South Florida, the Northeast Corridor, and around urban centers. A swath of the Plains state from the Dakotas to Texas has low attractiveness. The overall pattern is very similar in 2005 and 2010.

Table 4 shows the 25 counties with the highest $\tilde{\delta}$ values averaged from 2005-2010. Many of the counties on the list are in the Southwest, California, Texas, and Florida, and most are home to major cities, including Honolulu, HI, Phoenix, AZ, Las Vegas, NV, El Paso, TX, and Los Angeles, CA in the top five spots. Table 5 displays the 25 counties with the lowest identified $\tilde{\delta}$ values. Fourteen of the top 25 are in a swath of land in the center of the country from the Dakotas to Texas.

There are many similarities between these results and the ranking reported in Albouy (2012), including Honolulu as the number one location, and the general favorability of areas on the West Coast and Southwest. Our estimates show less favorability for the many of the other up-market communities that dominate the top of Albouy's list, and we find Phoenix, AZ and Houston,TX to be much more attractive. But it should be noted that quality of life, as measured by wage-rent differentials, and attractiveness are distinct concepts. Some of the places that are highly attractive in our model are attractive because of high wages and low rents, which may compensate for lower amenities. And very expensive communities are not attractive due to costs, despite the quality of amenities there.

Figure 11 shows the values for $\tilde{\alpha} + \tilde{\delta}$ with the same color scheme as in Figure 10. Retentive counties are clustered in the Rust Belt and Upper South. The West has generally low retentiveness. Again the overall pattern is very similar between 2005 and 2010.

Fourteen of the top 25 counties are in Tennessee, Kentucky, or Missouri. Counties in which there was no observed in-migration do not have $\tilde{\delta}$ and $\tilde{\alpha}$ identified separately, do have $\tilde{\alpha} + \tilde{\delta}$ well-identified.

While α is highly negatively correlated with δ, and is a function of δ, $\alpha + \delta$ and δ have very small correlation at about -0.046 for all estimated parameters from 2005 through 2010.

Parameter values also appear to be highly correlated over time. The correlation between $\tilde{\delta}$ for 2005 and for 2010 is 0.80 and between $\tilde{\alpha} + \tilde{\delta}$ for 2005 and for 2010 is 0.94. Figure 12 shows the quantiles of the differences in $\tilde{\delta}$ and $\tilde{\alpha} + \tilde{\delta}$ between 2005 and 2010. While there is no strong pattern among the change in $\tilde{\alpha} + \tilde{\delta}$, the areas that saw the highest increase in attractiveness appear to be in the same region with the least attractive counties, indicating the largest improvements were at the bottom of the distribution.

Values of γ^t, the gradient of the utility function equal to the marginal utility of income, are also very stable across years. Results are presented in Table 6. As expected, distance has a negative marginal utility, while being in the same CBSA or the same state, and wages have positive marginal utility. Being in the same census region has a small negative impact, likely due to also conditioning on distance and the other indicators. The coefficient on wages is higher in 2007 than for other years.

6.2 Quality of Life

In addition to estimation of marginal valuations for local public goods, our model provides a framework for estimating aggregate local amenities in the spirit of the literature on quality of life. The basic idea of that literature is that people value consumption and quality of life. Then, under an assumption of free mobility, people do not need to be compensated to live where they do beyond what they earn in income. This leads to an implication that the quality of life is equal to the overall expenditure level in a location minus after-tax wages. Albouy (2012) suggests constants to adjust housing costs to estimate the overall price level and after-tax income, as well as a number of additional refinements.

We adopt the adjustments for housing costs and wages, but are not forced to rely on the free mobility assumption and simplify the model to rely on only a single type. From our behavioral model and assuming the marginal utility of income is the marginal utility of consumption, the utility to move to county j if one is compensated for moving costs is,

$$V_j^t = \gamma_{wage} Q_j + V_j^{ct} + \mu_j^t = \gamma_{wage} Q_j + \gamma_{wage}^t * (.51 * log(wage_j^t) - .33 * log(housing_j^t)) + \mu_j^t, \quad (31)$$

where V^{ct} is the value of consumption and Q is the quality of life, normalized for convenience by γ_{wage}, and *housing* is median housing costs. Setting this equal to our estimated value $V_j^t = \delta_j^t + \gamma_{wage} * log(wage_j^t) + \epsilon_j^t$ from a model without county rurality, we can rearrange to calculate the quality of life as,

$$Q_j = \frac{\delta_j^t}{\gamma_{wage}^t} + .49 * log(wage_j^t) + .33 * log(housing_j^t) + \epsilon_j^t - \mu_j^t. \quad (32)$$

Then the quality of life can be estimated as the mean values of $\frac{\delta_j^t}{\gamma_{wage}^t} + .49 * log(wage_j^t) + .33 * log(housing_j^t)$ in the county over the study period (equivalent to the fixed effect in a fixed-effect-only model). Note that because δ is calculated already accounting for the full value of mean wages, we add that value back in and subtract off the net income in calculating the quality of life index.

Results are shown in Table 7. It should be noted that quality of life in this context does not mean the quality of the life people lead there, but the quality of the local amenities after adjusting for prices. The list is very similar to the list of most attractive counties with the top four spots being identical. But the ordering is somewhat different, with some of the Texas counties dropping in the rankings, and urban coastal counties like King, WA (Seattle); San Diego, CA; and Miami-

Dade, FL rising.

6.3 Second-Stage Results

The parameter values discussed in Section 6.1 provide inputs for a second-stage regression to explain $\tilde{\alpha} + \tilde{\delta}$ and $\tilde{\delta}$ in terms of the characteristics of counties. This second stage is based on the subset of counties for which control data are available, whereas the first stage is based on migration flows for nearly all counties in the US. Concerns about sample selection issues are discussed in Appendix A. We do not find evidence of selection effects that would cause inconsistency of estimate parameters.

A naive implementation of the model would estimate a second stage using an OLS fixed effects model. Results for such an implementation are given in Table 8. Coefficients are quite small in magnitude across models with little statistical significance, although with a fairly constant magnitude across models.

As discussed in Section 4.1, PM 2.5 is likely to be endogenous in the structural equation due to correlation with unobserved job market characteristics. Because of this, a better approach is a linear panel instrumental variables model that instruments for particulate concentrations using pollution inflow from distant counties. We find very consistent coefficient values for PM 2.5 across a variety of specifications.

The endogenous projection (i.e. panel first stage) is quite strong, as shown in Table 9. The table presents results for three models: each has the same included regressors (discussed above), and county fixed effects. The first includes year fixed effects, the second census region-by-year fixed effects, and the third EPA region-by-year fixed effects. The census categorizes four regions: Northeast, Midwest, South, and West. The EPA categorizes 10 regions. In each case, the F-statistic is well above the rule of thumb threshold, indicating that inflow SO2 is not a weak instrument.

Table 10 shows coefficients on PM 2.5 for a progression of model specifications, starting with only census region-by-year fixed effects, then sequentially adding the unemployment rate, average wage, median rent, housing controls and industry controls. The parameter values are very similar across specifications, giving a strong indication that the reduced-form coefficient is well-identified. These results imply that PM 2.5 concentrations influence households' decisions of whether to migrate away from their current county of residence.

Reduced-form results are given in Table 11 for the same models as Table 9 and the final model of Table 10. The coefficients reflect marginal valuations of non-migrant households and, due to the normalization based on the first-stage coefficient on log wages, reflect semielasticities with respect to income, giving the proportional change in income households will accept for a unit change in air pollution. The first two models show significant coefficients with the expected sign for the parameter of interest on average dissolved PM 2.5. The third does not have the expected sign and is not significant at conventional levels. With only 461 clusters, there may not be sufficient variation in pollution after accounting for 60 time fixed effects. Our preferred results are the model with census region-by-year fixed effects which provides greater robustness to differential time trends, without soaking up all the useful variation into fixed effects.

The coefficient on wage is negative and significant across the models. This is likely due to the inclusion of cross-sectional wages in the first stage so that there is no structural interpretation to the coefficient, but it acts as a proxy for other unobserved job market characteristics. The coefficient on median rent is negative, but not significant for all three models. While rents are an equilibrium outcome that is endogenous to the structural equation, we do not believe that rents are meaningfully partially correlated with PM 2.5, which is required for the endogeneity to cause inconsistency of the parameter of interest.

Equivalent results for a progression of models of $\tilde{\delta}$ are shown in Table 8. Again, the coefficient on PM 2.5 is largely stable across the models, though of varying significance. Results for models with a full set of controls and varying time fixed effects are given in Table 13. These coefficients reflect the valuations of households that decide to migrate between counties. A negative coefficient indicates support for the Tiebout hypothesis that location decisions reflect the value of PM 2.5 pollution and people "vote with their feet" to select lower PM 2.5 concentrations, all else equal. These results are consistent with the Tiebout hypothesis, but not strong evidence for it. Again, wages have negative coefficients, with an ambiguous structural interpretation. Both housing costs and unemployment have negative coefficients, as expected, though the housing cost coefficients are not significant for the first two models.

Table 14 shows F-statistics for the endogenous projection at deadbands of 30, 50, 75, 150, and 200 miles, in addition to the base specification of 100. That is, it provides evidence that the instrument has remained strong even as we exclude counties from a larger and larger radius in estimating the amount of SO2 that is flowing in. Table 15 shows the reduced-form results for these varied deadband instruments, including the base specification. The coefficient on PM 2.5 concentrations increases with larger deadbands, possibly reflecting the fact that emissions from closer counties may be more correlated with local job market conditions, weakening the exclusion restriction on the instrument. We retain the 100 mile deadband as we believe the exclusion restriction remains valid and the instrument remains strong.

6.3.1 Marginal Willingness to Pay Table 16 provides estimates of marginal willingness to pay to avoid PM 2.5 concentrations in 2005 dollars. Because the coefficients are semielasticities, the value depends on an income level. The table reflects value for a household earning $50,000 per year, close to the median income during the middle of our study period. Results are presented for non-migrants and for migrants based on the census region-by-year fixed effects models for a one standard deviation change, a one unit change, and a one percent change. A one standard deviation change in PM 2.5 concentrations is roughly equivalent to moving from New York City to St. Paul, MN. The results reflect annual values based on the tradeoff on wage levels.

These results are not directly comparable to findings from Bayer et al. (2009) or Hamilton and

Phaneuf (2015) due to their use of a different class of particulate matter, PM 10. But based on the average ratio of PM 10 to PM 2.5 in our sample, and converting to 1982-84 dollars, our preferred estimate is $360 for a one unit change, compared a range of $149 to $185 in Bayer et al. (2009) and $114 to $413 in Hamilton and Phaneuf (2015).

As a point of comparison, we have also estimated marginal willingness to pay using the hedonic framework. That is, we regress various measures of housing prices on instrumented PM 2.5 pollution and a similar set of county-level controls. This approach does not take into account sorting due to preference heterogeneity or moving costs. None of these models had statistically significant coefficients when including census region-by-year fixed effects, so the results in Table 17 reflect models with year fixed effects. The table presents estimates of marginal willingness to pay based on the migration sorting model as well as hedonic models using median rent, median housing costs (based on both renters and owners), median owner costs, and median property values. The first four are each annual measures of willingness to pay, while the final represents a long-run estimate. Each of the annual hedonic estimates is significantly smaller than the sorting estimate, as is the property value estimate for reasonable long-run time scales. This result highlights the importance of accounting for heterogeneous preference sorting and moving costs.

7 Conclusions

This study develops an innovative model to address the important and relevant question of the value of air quality and local public goods. By tapping a rich but unused data source, it uses county-level variation in migration shares to estimate mean utility parameters, and cross-sectional variation in wages to estimate a marginal utility of income, and then uses panel variation of county characteristics to estimate marginal valuations for clean air. By relying on migration flow data, including households who remain in the same county, we are able to separately estimate impacts on households that do not move and households that move with changes in local amenities, and allow for flexible heterogeneous average moving costs. By relying on IRS migration estimates, we are able to broaden our analysis beyond large cities to include a larger cross-section of the country.

We find that the most attractive counties on average across the sample are located primarily in urban areas along the coasts and in the Southwest and Texas, with Honolulu, HI claiming the most attractive spot. The least attractive counties lie mostly in a band in the Planes States from the Dakotas south to Texas. But the largest increases in attractiveness occurred in this same region over our study period.

Our preferred specification gives a marginal willingness to pay to avoid $1\mu g/m^3$ of $1325 for non-migrant households, and $484 for migrant households. This estimate of marginal willingness to pay is not directly comparable to Chay and Greenstone (2005), Bayer et al. (2009), or Hamilton and Phaneuf (2015) because they rely on broader classes of particulate matter, but they find smaller values, though of the same order of magnitude. Our result has the advantage of being based on a panel of mean utilities developed using the decisions households make each year due to the pollution concentrations they face and the costs of benefits of moving to other locations.

The model we develop could be used to address a wide array of questions regarding valuation of local amenities beyond the current focus on air quality. An ideal extension of the model would be to marry the aggregate migration data to micro-level data to deal directly with greater household heterogeneity. By generalizing the estimation framework of Berry et al. (1995), the model provides a method for estimating demand for goods and services that are mutually exclusive during fixed periods of time and where there could be loyalty effects,including insurance policies and phone, Internet, and cable providers.

8 Tables and Figures

Table 1: Migration Summary Statistics

Year	2005	2006	2007	2008	2009	2010
Mean Out-migration Rate	0.084	0.079	0.078	0.078	0.074	0.074
5% Out-migration Percentile	0.050	0.048	0.048	0.047	0.044	0.045
95% Out-migration Percentile	0.124	0.119	0.120	0.115	0.110	0.112
Mean In-migration Rate	0.082	0.078	0.077	0.076	0.073	0.071
5% In-migration Percentile	0.044	0.044	0.044	0.043	0.041	0.041
95% In-migration Percentile	0.137	0.129	0.125	0.118	0.113	0.113
Correlation, In- and Out-migration	0.590	0.935	0.946	0.985	0.980	0.975
Mean Migration Distance	97.2	94.6	95.1	93.8	91.3	96.5
SD Migration Distance	146.2	141.8	144.4	143.6	143.6	147.2
Mean Observed Destination Counties	28.3	27.8	28.4	27.6	26.1	27
SD Observed Destination Counties	58.1	57.6	58.1	56.6	53.9	55

Table 2: Summary Statistics for Controls, 2005 Cross-section

Statistic	Mean	St. Dev.	Min	Max
PM 2.5($\mu g/m^3$)	12.7	3.1	3.3	21.6
Total Establishments (1,000s)	12.6	22.9	1.0	367.4
Average Employment (1,000s)	209.6	332.5	13.0	4,082.5
Average Wage (1,000s)	36.9	7.9	24.8	84.2
Construction Establishments	1,161.7	1,372.0	102	13,047
Construction Employment (1,000s)	11.3	17.6	0.0	161.0
Construction Wage (1,000s)	39.7	8.0	22.4	74.2
Manufacturing Establishments	542.7	1,019.0	27	15,889
Manufacturing Employment (1,000s)	20.3	33.5	0.0	467.3
Manufacturing Wage (1,000s)	48.0	12.1	26.5	123.1
Median Monthly Rent	700.6	170.1	412	1,287
Median Property Value (100,000s)	189.1	125.7	53.1	763.1
Median Rooms	5.5	0.4	3.2	7.0
Median Year Built	1972	10.6	1940	1993
Unemployment Rate	5.2	1.5	2.3	16.1

Note: Wages, rent, and property values are in 2005 dollars

Table 3: Panel Variation for Key Variables

	Mean	Standard Deviation	Within Standard Deviation	Mean Range
PM 2.5 Concentration ($\mu g/m^3$)	11.1	2.9	1.5	3.7
Mean Wage	39,938	8,918	1,961	5,015
Median Monthly Rent	769	196	49.8	125.4
Unemployment Rate	6.5	2.9	2.2	4.9

Note: Wages and rent are in 2005 dollars

Table 4: Most Attractive Counties on Average, 2005-2010

	County	State
1	Honolulu County	HI
2	Maricopa County	AZ
3	Clark County	NV
4	Los Angeles County	CA
5	Harris County	TX
6	El Paso County	TX
7	Bexar County	TX
8	San Diego County	CA
9	Anchorage Borough	AK
10	Cook County	IL
11	Riverside County	CA
12	Salt Lake County	UT
13	Bernalillo County	NM
14	King County	WA
15	Pima County	AZ
16	San Bernardino Count	CA
17	Broward County	FL
18	Tarrant County	TX
19	Spokane County	WA
20	El Paso County	CO
21	Miami Dade County	FL
22	Dallas County	TX
23	Hidalgo County	TX
24	Franklin County	OH
25	Orange County	FL

Table 5: Least Attractive Counties on Average, 2005-2010

	County	State
1	Hartley County	TX
2	Wheeler County	TX
3	Real County	TX
4	Haskell County	KS
5	Boone County	NE
6	Hamilton County	NY
7	Ontonagon County	MI
8	Wayne County	UT
9	Kimble County	TX
10	Hettinger County	ND
11	Pendleton County	WV
12	Sierra County	CA
13	Sharkey County	MS
14	Randolph County	GA
15	Hancock County	TN
16	Nelson County	ND
17	Grant County	OR
18	Harper County	OK
19	Pocahontas County	WV
20	Edwards County	KS
21	Sanborn County	SD
22	Custer County	ID
23	Oneida County	ID
24	Corson County	SD
25	De Baca County	NM

Table 6: First Stage Gradient Estimates

	2005	2006	2007	2008	2009	2010
Log Distance	-1.279	-1.266	-1.281	-1.286	-1.295	-1.245
Same CBSA	0.749	0.727	0.725	0.714	0.758	0.727
Same State	1.379	1.385	1.392	1.421	1.444	1.446
Same Census Region	-0.064	-0.071	-0.095	-0.099	-0.115	-0.101
Log Wage	1.947	1.888	2.542	1.816	1.703	1.801
Difference in Rurality	-2.346	-2.353	-2.102	-2.292	-2.348	-2.412

Table 7: Highest Quality of Life

	County	State
1	Honolulu County	HI
2	Maricopa County	AZ
3	Clark County	NV
4	Los Angeles County	CA
5	San Diego County	CA
6	Anchorage Borough	AK
7	Harris County	TX
8	King County	WA
9	Riverside County	CA
10	Broward County	FL
11	Bexar County	TX
12	Cook County	IL
13	San Bernardino Count	CA
14	Miami Dade County	FL
15	Salt Lake County	UT
16	El Paso County	TX
17	El Paso County	CO
18	Tarrant County	TX
19	Bernalillo County	NM
20	Pima County	AZ
21	Palm Beach County	FL
22	Orange County	CA
23	Washoe County	NV
24	Orange County	FL
25	Spokane County	WA

Table 8: Progression of OLS Alpha+Delta Models Adding More Controls

	Dependent variable:					
	Normalized Alpha+Delta					
	(1)	(2)	(3)	(4)	(5)	(6)
Average Dissolved PM 2.5	-0.00339	-0.00326*	-0.00260	-0.00252	-0.00184	-0.00212
	(0.00184)	(0.00145)	(0.00147)	(0.00146)	(0.00135)	(0.00130)
County Fixed-Effects	X	X	X	X	X	X
Census Region-year Fixed-Effects	X	X	X	X	X	X
Unemployment Rate		X	X	X	X	X
Wage			X	X	X	X
Median Rent				X	X	X
Industry Controls					X	X
Housing Controls						X
Observations	2648	2642	2642	2642	2642	2602
R^2	0.864	0.901	0.911	0.911	0.915	0.927

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Standard errors clustered at the county level

Table 9: Endogenous Projection:PM2.5

	Instrumental Variable:		
	Inflow SO2 Beyond 100 Miles		
	(1)	(2)	(3)
Cragg-Donald Wald F statistic	139.1	81.8	41.0
Stock-Yogo 10% Critical Value		16.38	
County Fixed-Effects	X	X	X
Time Fixed-Effects	Yr	Census Reg-Yr	EPA Reg-Yr
Observations	2597	2597	2597
Clusters	461	461	461

Table 10: Progression of IV Alpha+Delta Models Adding More Controls

	Dependent variable:					
	Normalized Alpha+Delta					
	(1)	(2)	(3)	(4)	(5)	(6)
Average Dissolved PM 2.5	-0.0229*	-0.0280***	-0.0259**	-0.0261**	-0.0239**	-0.0265***
	(0.00910)	(0.00838)	(0.00801)	(0.00807)	(0.00809)	(0.00667)
County Fixed-Effects	X	X	X	X	X	X
Census Region-year Fixed-Effects	X	X	X	X	X	X
Unemployment Rate		X	X	X	X	X
Wage			X	X	X	X
Median Rent				X	X	X
Industry Controls					X	X
Housing Controls						X
Observations	2648	2640	2640	2640	2597	2597
R^2	0.848	0.873	0.887	0.886	0.893	0.900

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors clustered at the county level

Table 11: IV Alpha+Delta Models

	Dependent variable:		
	Normalized Alpha+Delta		
	(1)	(2)	(3)
Average Dissolved PM2.5	-0.0345***	-0.0265***	0.00128
	(0.00498)	(0.00667)	(0.00803)
County Fixed-Effects	X	X	X
Time Fixed-Effects	Yr	Census Reg-Yr	EPA Reg-Yr
Observations	2597	2597	2597
Clusters	461	461	461
R^2	0.860	0.900	0.937

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors clustered at the county level.
Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

TABLE 12: PROGRESSION OF IV DELTA MODELS ADDING MORE CONTROLS

	Dependent variable:					
	Normalized Delta					
	(1)	(2)	(3)	(4)	(5)	(6)
Average Dissolved PM 2.5	-0.00324	-0.0122	-0.00960	-0.0104	-0.00644	-0.00968
	(0.00982)	(0.00965)	(0.0103)	(0.0103)	(0.00861)	(0.00797)
County Fixed-Effects	X	X	X	X	X	X
Census Region-year Fixed-Effects	X	X	X	X	X	X
Unemployment Rate		X	X	X	X	X
Wage			X	X	X	X
Median Rent				X	X	X
Industry Controls					X	X
Housing Controls						X
Observations	2648	2640	2640	2640	2640	2597
R^2	0.554	0.581	0.605	0.604	0.674	0.687

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Standard errors clustered at the county level

Table 13: IV Delta Models

	Dependent variable:		
	Normalized Delta		
	(1)	(2)	(3)
Average Dissolved PM2.5	-0.0276***	-0.00968	0.0000971
	(0.00664)	(0.00797)	(0.0109)
County Fixed-Effects	X	X	X
Time Fixed-Effects	Yr	Census Reg-Yr	EPA Reg-Yr
Observations	2597	2597	2597
Clusters	461	461	461
R^2	0.595	0.687	0.720

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors clustered at the county level
Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

Table 14: Endogenous Projection At Various Deadbands

	Cragg-Donald Wald F statistic					
	30 Mi	50 Mi	75 Mi	100 Mi	150 Mi	200 Mi
Year FE	137.1	137.7	138.2	139.1	127.3	113.0
Census Reg-Year FE	76.5	77.6	79.7	81.8	69.8	55.9
Stock-Yogo 10% Critical Value	16.38					
County Fixed-Effects	X	X	X	X	X	X
Observations	2597	2597	2597	2597	2597	2597
Clusters	461	461	461	461	461	461

Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

Table 15: Comparison of Various Pollution Flow Deadbands

	Dependent variable:					
	Normalized Alpha + Delta					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Dissolved PM 2.5	-0.0193** (0.00613)	-0.0210*** (0.00625)	-0.0248*** (0.00644)	-0.0265*** (0.00667)	-0.0309*** (0.00789)	-0.0401*** (0.0102)
Deadband	30	50	75	100	150	200
County FE	X	X	X	X	X	X
Time FE	Cen. Reg-Yr	Cen. Reg-Yr	Cen. Reg-Yr	Cen. Reg-Yr	Cen. Reg-Yr	Cen. Reg-Yr
Observations	2597	2597	2597	2597	2597	2597
R^2	0.914	0.911	0.904	0.900	0.890	0.863

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors clustered at the county level.
Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

Table 16: Estimated Marginal Willingness to Pay

	Group:	
Decrease in Average PM2.5	Non-migrants (Alpha+Delta)	Migrants (Delta)
1 s.d. of overall mean	$3800	$1388
1 μg per m^3	$1325	$484
1% of overall mean	$146	$53
County FE	X	X
Time FE	Census Reg-Yr	Census Reg-Yr

Note: Estimates reflect marginal willingness to pay in 2005 dollars avoid the indicated average concentration of PM 2.5 for a family with $50,000 per year income.

Table 17: Comparison of Sorting and Hedonic Estimates

	Model:				
Decrease in Average PM2.5	Sorting	Hedonic: Rent	Hedonic: Housing Cost	Hedonic: Owner Cost	Hedonic: Property Value
1 μg per m^3	$1725	$81	$153	$161	$4645
Time Scale	Year	Year	Year	Year	Long-run
County FE	X	X	X	X	X
Time FE	Year	Year	Year	Year	Year

Note: Estimates reflect marginal willingness to pay in 2005 dollars avoid the indicated average concentration of PM 2.5 for a family with $50,000 per year income, or at the average of the indicated housing cost measure.

Figure 1: Migration Cross-sections for 2005-06

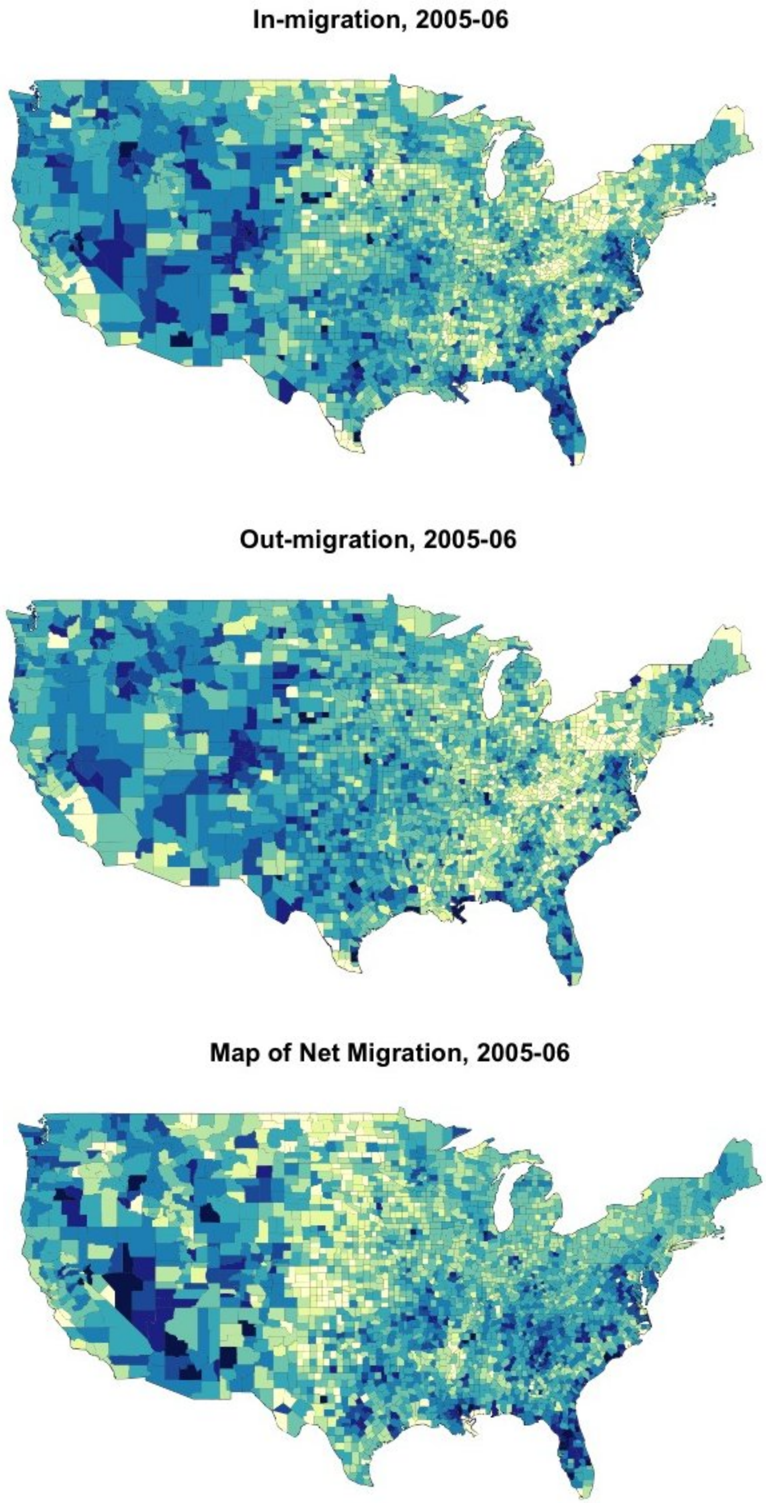


Figure 2: Net Migration for 2005-06 and 2009-10

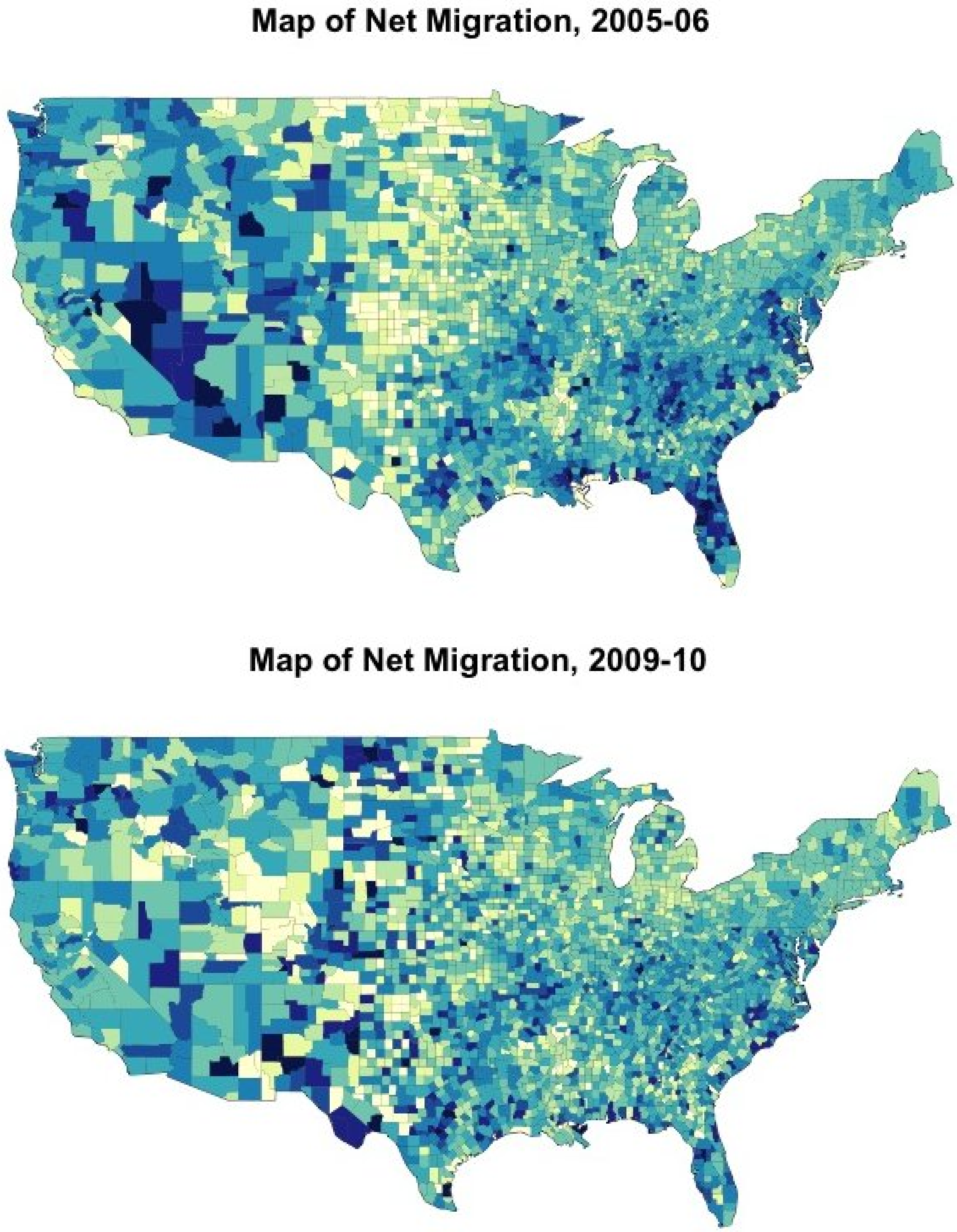


Figure 3: California Net Migration for 2005-06 and 2009-10

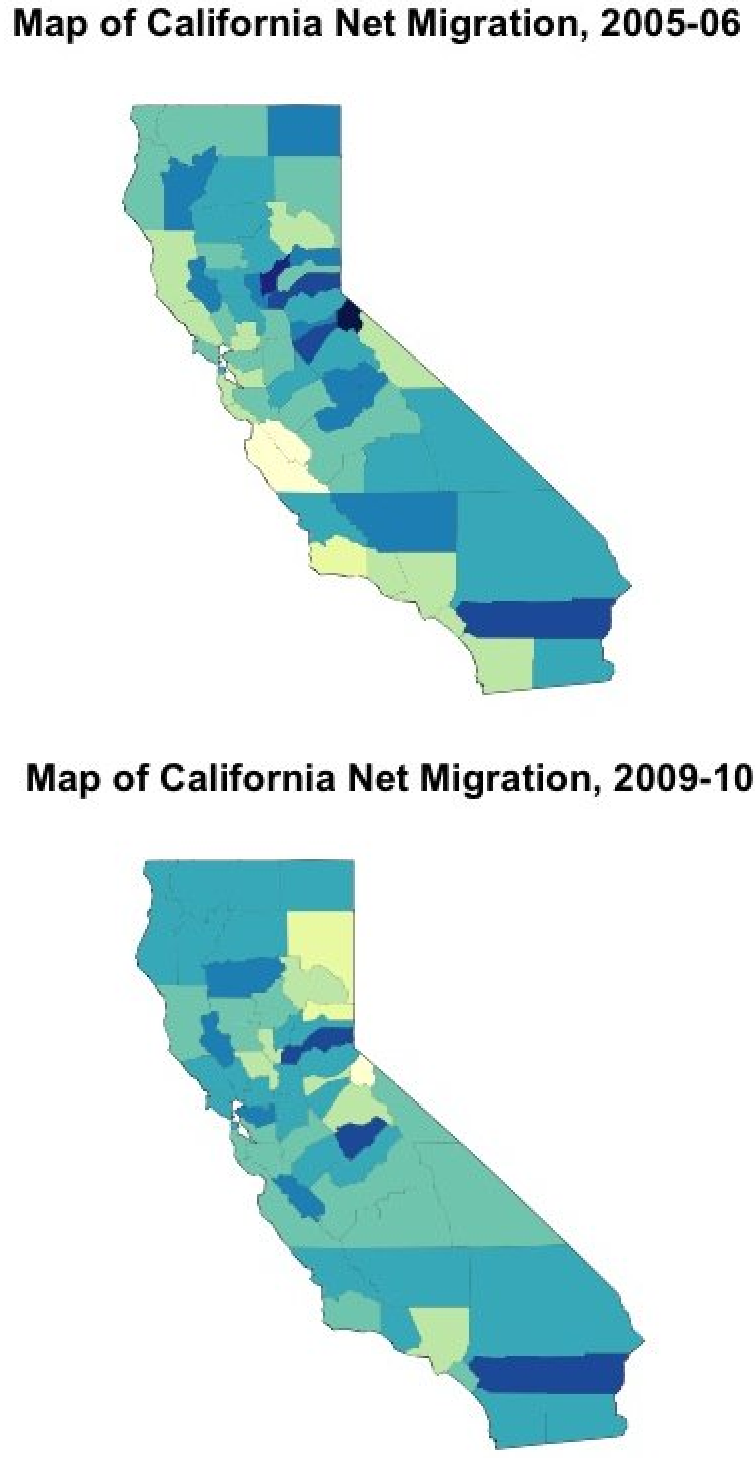
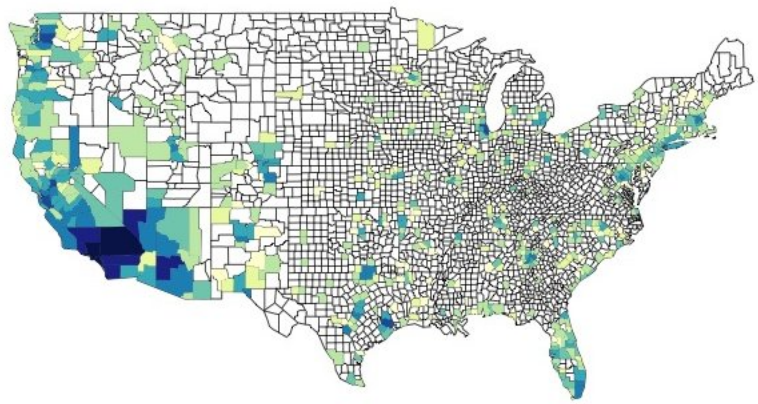
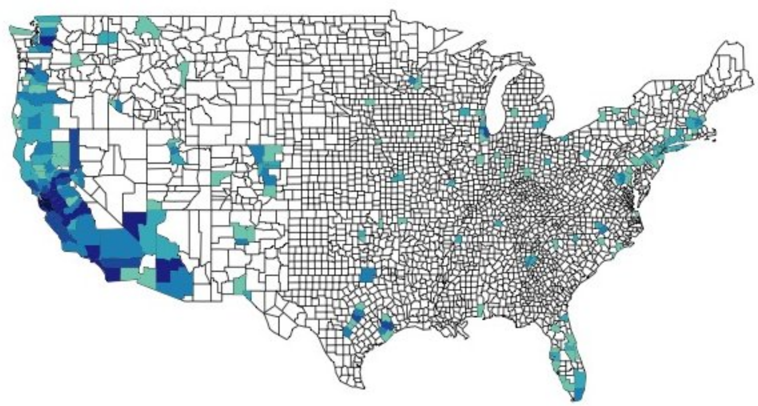


Figure 4: County-to-county Migration Rates for 2005-06 from Los Angeles, Santa Clara, and New York counties

Los Angeles County Migration, 2005-06



Santa Clara County Migration, 2005-06



New York County Migration, 2005-06

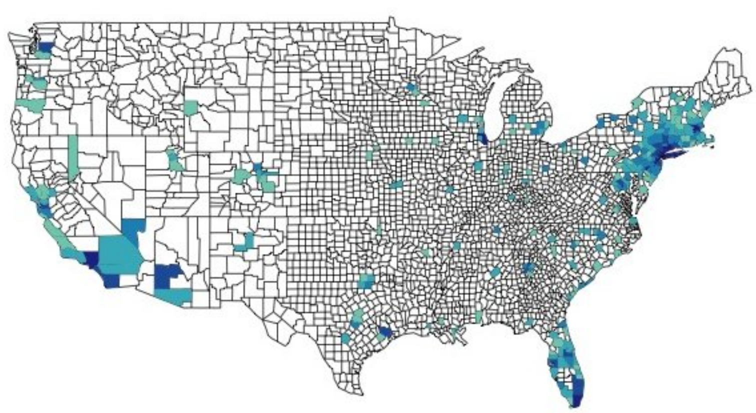


Figure 5: County-to-county Migration Rates to California in 2005-06 from Los Angeles, Santa Clara, and New York counties

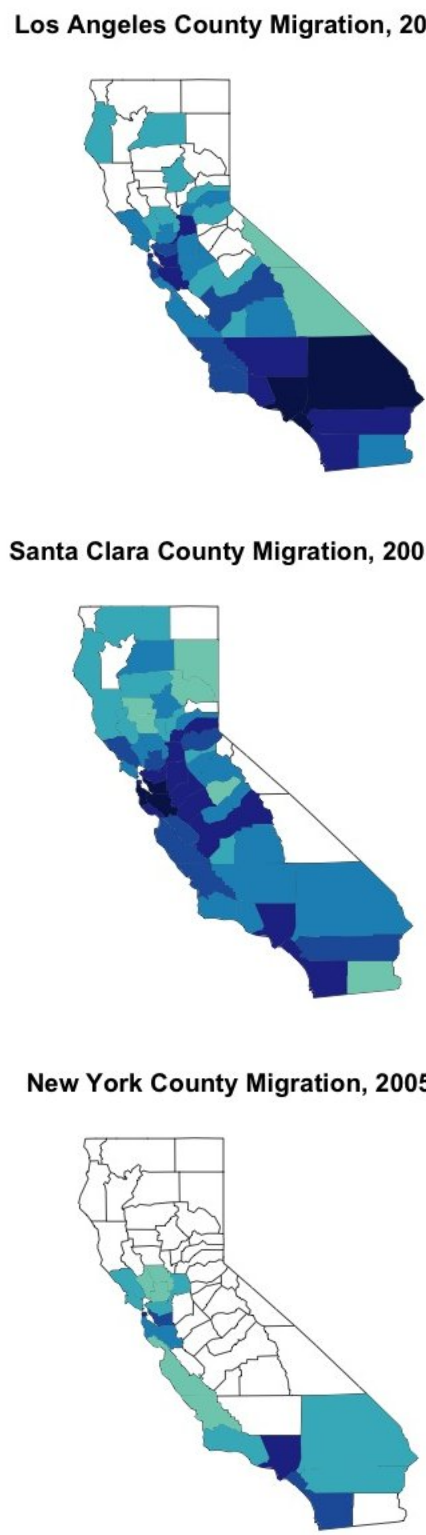


Figure 6: County-level Average Wages, 2005

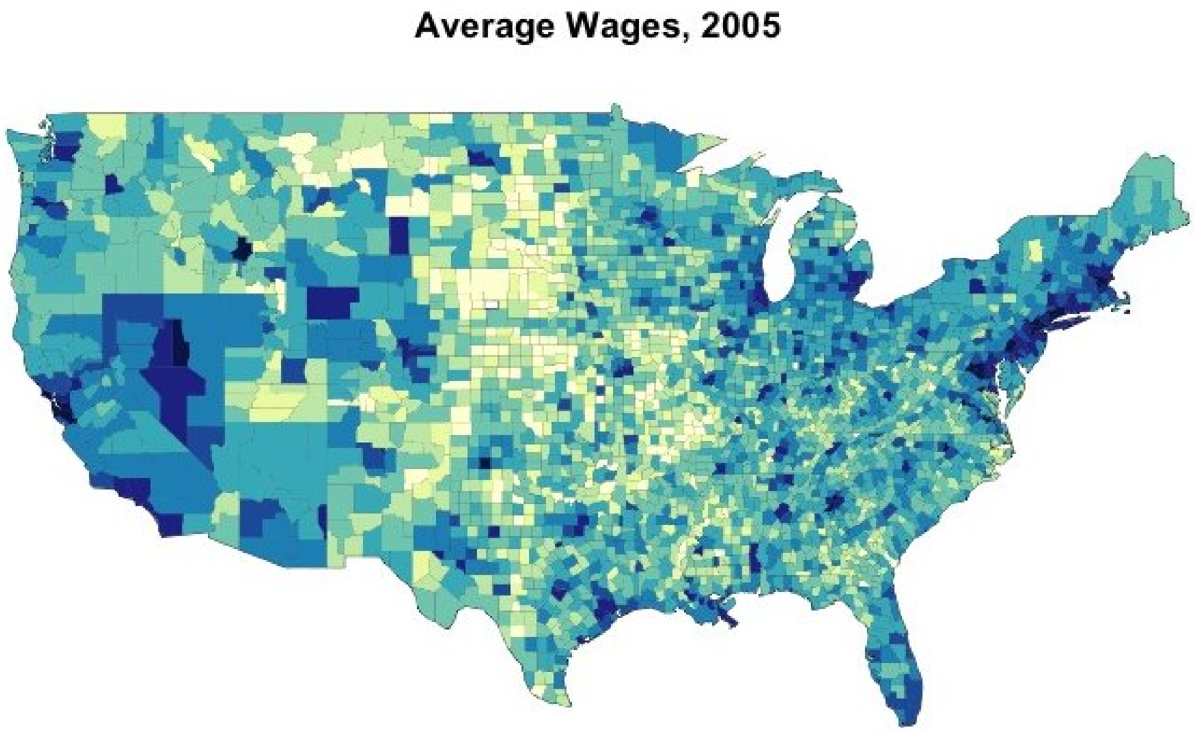


Figure 7: Average PM 2.5 Concentrations, 2009
Counties in white do not have EPA monitoring

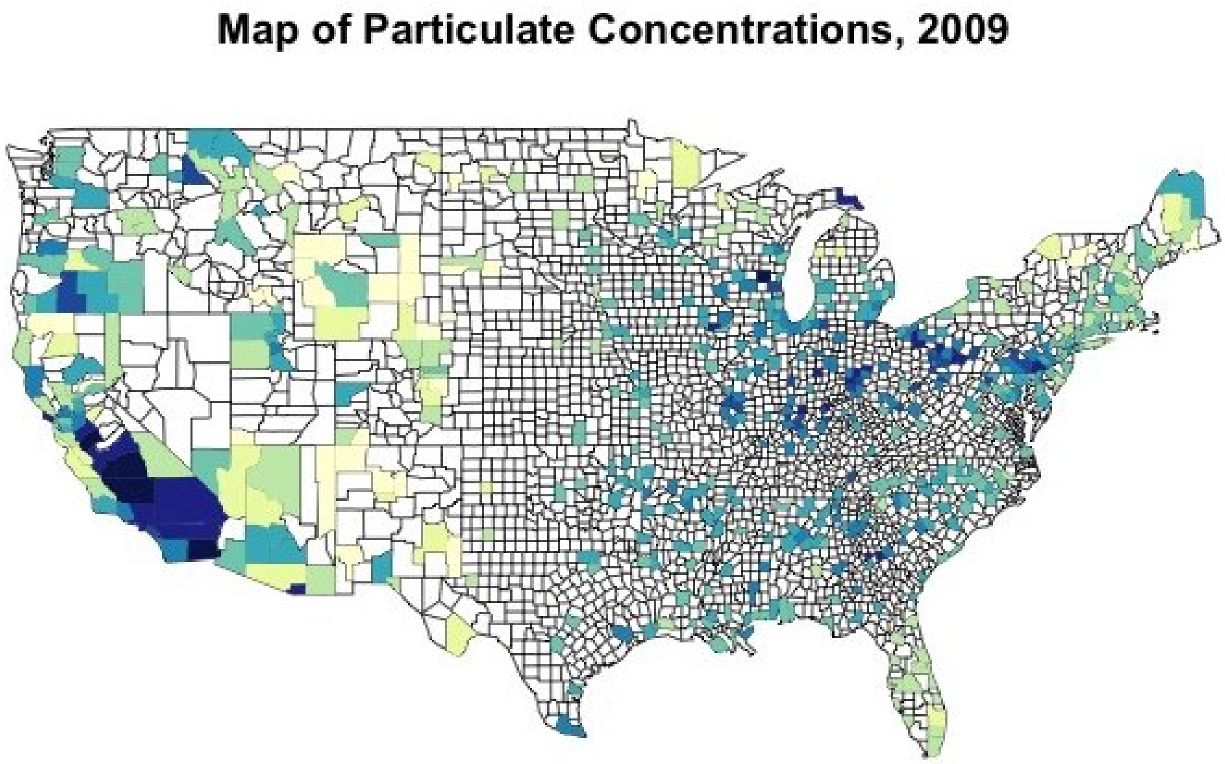


Figure 8: Average California PM 2.5 Concentrations, 2005 and 2009
Counties in white do not have EPA monitoring

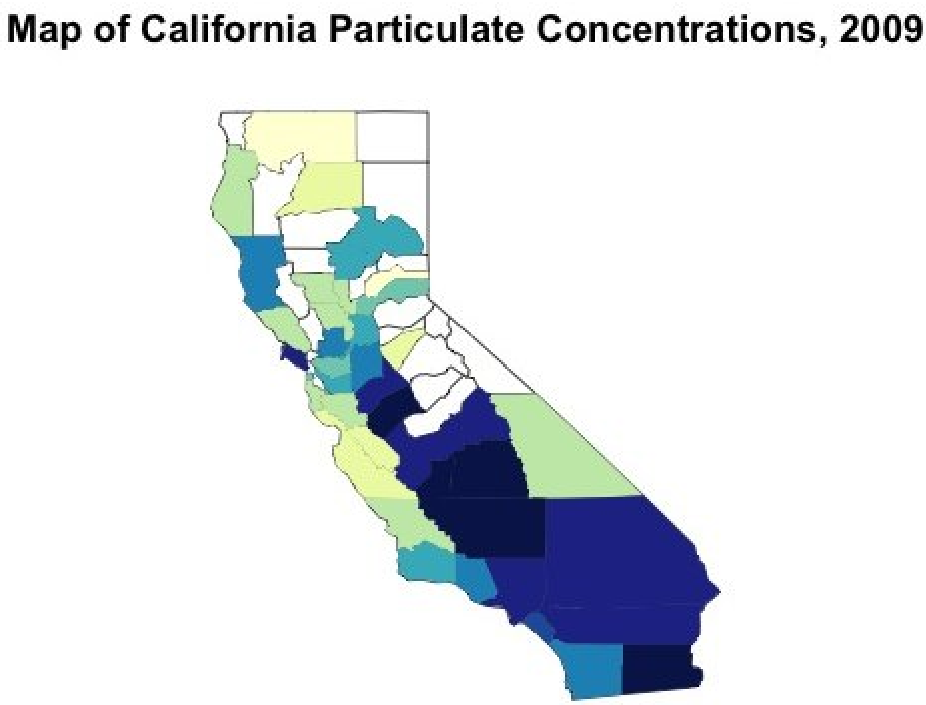


Figure 9: SO2 Point Source Emissions Flowing to New York County, NY and Allegheny County, PA, 2005
Counties in white do not have point source emitters

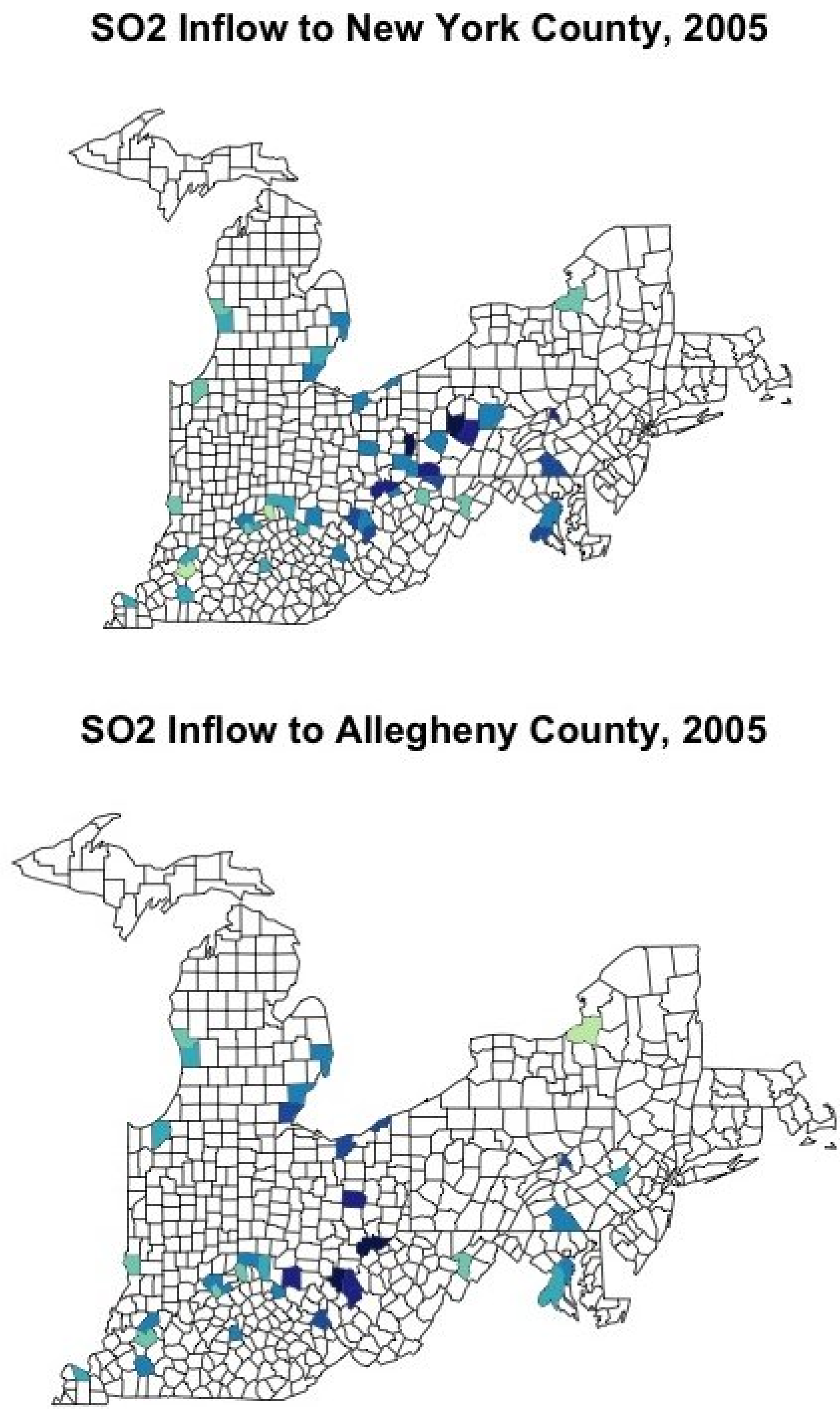
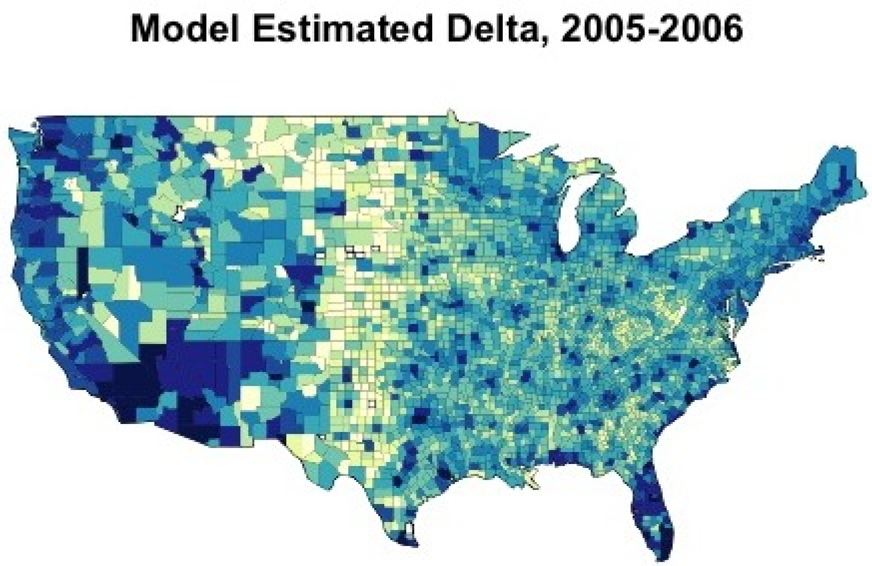


Figure 10: County Attactiveness, 2005 and 2010



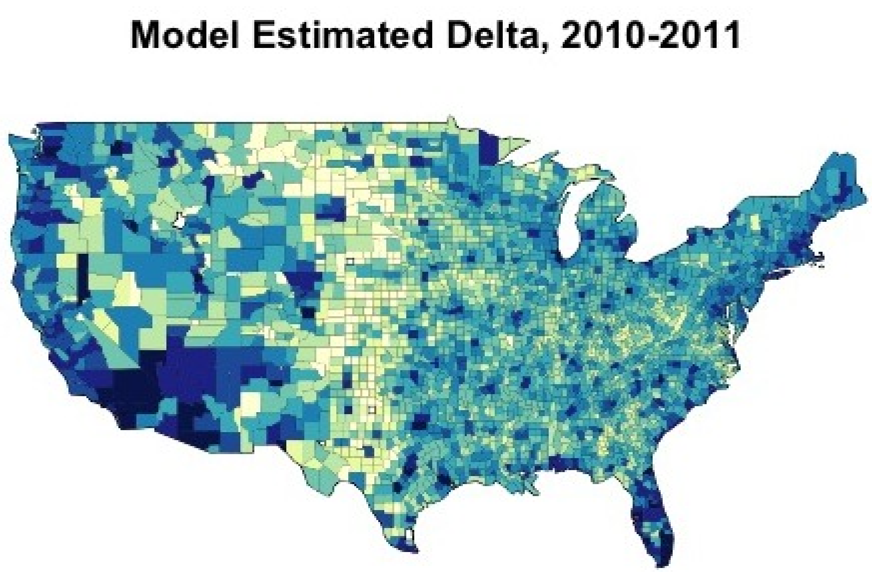


Figure 11: County Retentiveness, 2005 and 2010

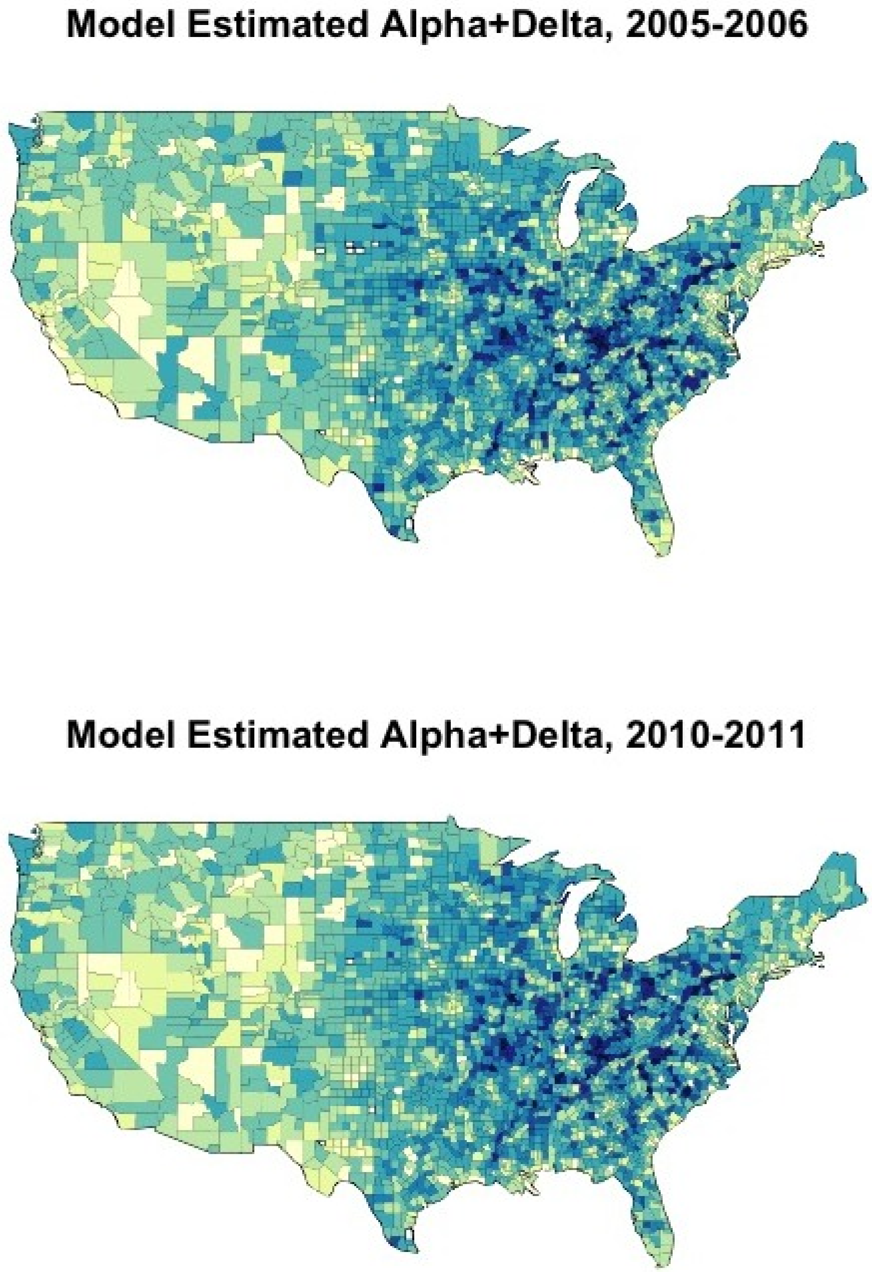
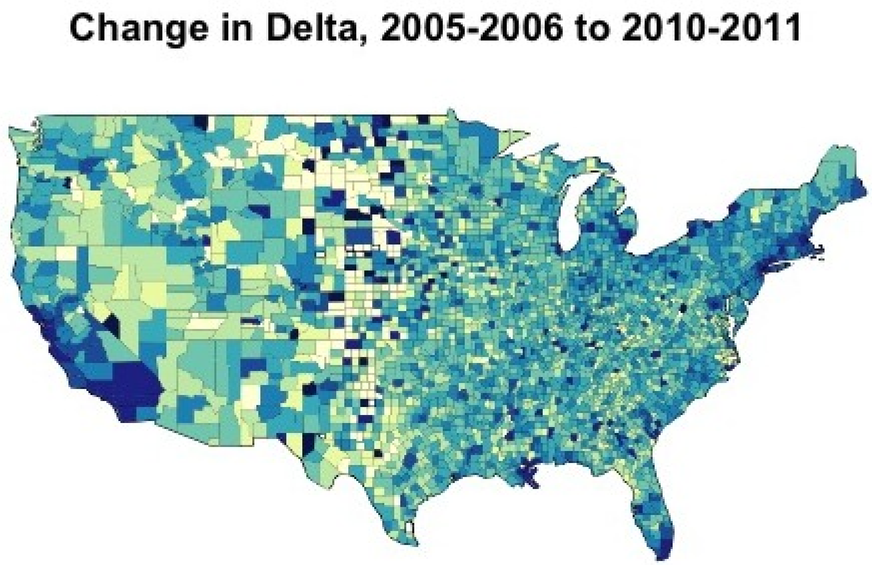
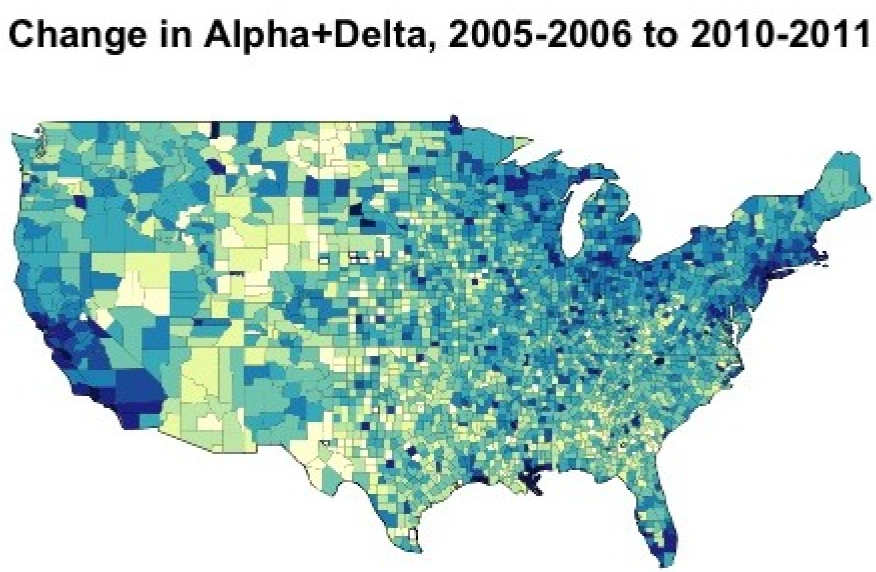


Figure 12: Changes in Parameter Values, 2005-2010





A Selection

While the first stage model is evaluated on nearly all counties, the endogenous projection and reduced form are evaluated only on the counties for which the control variables are available. This raises two potential concerns: an overall concern about incidental truncation that there may be unobserved factors causing inconsistency in estimated parameter, and that the panel is unbalanced and counties that enter the sample may do so due in a manner correlated with the characteristic of interest. We assess both of these issues.

A.1 Incidental Truncation

Of the roughly 3100 counties and county-equivalents in the US, only 461 enter our main analysis in estimating marginal willingness to pay. If selection into the panel with the full data, which is limited primarily by American Community Survey housing data and EPA particulate pollution data, is related to the unobserved idiosyncratic error in the model, then the parameter estimates will not be consistent for the full population's parameters. As this sample represents the vast majority of the population, the parameters would still be of interest, but not universally applicable. Because inclusion in the panel is restricted by multiple data sources, we cannot fully test for incidental truncation, but we can conduct a partial test comparing counties that have all the relevant data to counties that have all relevant data except the pollution data.

Testing for an impact of incidental truncation involves estimating a probit selection equation and including the Inverse Mills Ratio (IMR) for the included observations in the full second stage regression. Then a test of the significance of coefficient for the IMR is a test for effects of incidental truncation. There is no need to make additional adjustments for the asymptotic distribution due to the IMR being a calculated regressor, and the selection equation doesn't even need to be correct. It does need to reflect the control values from each year and not just the contemporaneous values, but this can be accomplished by using the mean value for each county for each variable. (Remedying problems of incidental truncation require more care.)

Results are shown in Table 18. The coefficient on the IMR is not significantly different from zero at any conventional level for any of the models. Thus there is no evidence from this test that incidental truncation is causing inconsistency of our estimates for the population parameters.

A.2 Changing Panel Composition

The results presented in Section 6.3 are for the sample of observations where full data are available, which is an unbalanced panel due primarily to the changing set of EPA monitoring data over the study period. A second concern is that counties are entering the panel in an a way that correlated with the error term. This can be tested by including a lagged indicator of whether the county is in the sample in the full second stage. Results are shown in Table 19. The lagged sample inclusion indicator is not significant any of the models, so there is not an indication of problems with the unbalanced panel. Additionally, Table 20 shows results that rely only on a balanced panel, excluding counties that are not included in all years. The parameter estimates are very similar between the models.

A.3 Selection Tables

Table 18: Incidental Truncation

	Dependent variable:		
	Normalized Alpha+Delta		
	(1)	(2)	(3)
Inverse Mills Ratio	-0.0148	-0.0594	-0.0322
	(0.0520)	(0.0492)	(0.0449)
County Fixed-Effects	X	X	X
Time Fixed-Effects	Yr	Census Reg-Yr	EPA Reg-Yr
Observations	2574	2574	2574
Clusters	445	445	445
R^2	0.850	0.890	0.936

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Standard errors clustered at the county level
Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

Table 19: Test of Effect of Unbalanced Sample

	Dependent variable:		
	Normalized Alpha+Delta		
	(1)	(2)	(3)
Lagged PM 2.5 Sample Indicator	-0.00283	0.0101	0.00258
	(0.0217)	(0.0126)	(0.0122)
County Fixed-Effects	X	X	X
Time Fixed-Effects	Yr	Census Reg-Yr	EPA Reg-Yr
Observations	2145	2145	2145
Clusters	444	444	444
R^2	0.780	0.909	0.935

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Standard errors clustered at the county level
Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

Table 20: Comparison of Balanced and Unbalanced Panels

	Dependent variable:			
	Normalized Alpha+Delta			
	(1)	(2)	(3)	(3)
Average Dissolved PM2.5	-0.0345***	-0.0358***	-0.0265***	-0.0266***
	(0.00498)	(0.00512)	(0.00667)	(0.00684)
Balanced Panel		X		X
County Fixed-Effects	X	X	X	X
Time Fixed-Effects	Yr	Yr	Census Reg-Yr	Census Reg-Yr
Observations	2597	2404	2597	2404
Clusters	461	403	461	403
R^2	0.860	0.858	0.900	0.902

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors clustered at the county level
Controls include unemployment rate, mean wages, median rents, housing controls, and industry controls.

References

Albouy, David (2012), "Are big cities bad places to live? estimating quality of life across metropolitan areas." Technical Report 14472, National Bureau of Economic Research.

Albouy, David, Walter Graf, Ryan Kellogg, and Hendrik Wolff (2016), "Climate amenities, climate change, and american quality of life." *Journal of the Association of Environmental and Resource Economists*, 3, 205–246.

Albouy, David and Bert Lue (2015), "Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life." *Journal of Urban Economics*, 89, 74–92.

Bayer, Partrick, Nathaniel Keohane, and Christopher Timmins (2009), "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management*, 58, 1–14.

Bayer, Partrick and Christopher Timmins (2005), "On the equilibrium properties of locational sorting models." *Journal of Urban Economics*, 57, 462–477.

Bayer, Partrick and Christopher Timmins (2007), "Estimating equilibrium models of sorting across locations." *The Economic Journal*, 117, 353–374.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007), "A unified framework for measuring preferences for schools and neighborhoods." *Journal of Political Economy*, 18, 588–638.

Bayer, Patrick and Robert McMillan (2012), "Tiebout sorting and neighborhood stratification." *Journal of Public Economics*, 96, 1129–1143.

Bayer, Patrick, Robert McMillan, and Kim Rueben (2004), "An equilibrium model of sorting in an urban housing market." Working Paper 10865, NATIONAL BUREAU OF ECONOMIC RESEARCH.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 63, 841–890.

Chay, Kenneth Y and Michael Greenstone (2005), "Does air quality matter? evidence from the housing market." *Journal of Political Economy*, 113, 376–424.

Depro, Brooks, Christopher Timmins, and Maggie 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, 439–468.

Environmental Protection Agency (2016), "Fact sheet: Clean power plan benefits of a cleaner, more efficient power sector." Technical report, US EPA, URL https://www.epa.gov/cleanpowerplan/fact-sheet-clean-power-plan-benefits-cleaner-more-efficient-power-sector.

Hamilton, Timothy L and Daniel J 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.

Kahn, Matthew E (1995), "A revealed preference approach to ranking city quality of life." *Journal of Urban Economics*, 38, 221–235.

Klaiber, Alan and Daniel Phaneuf (2010), "Valuing open space in a residential sorting model of the twin cities." *Journal of Environmental Economics and Management*, 60, 57–77.

Kuminoff, Nicolai, V. Kerry Smith, and Christopher Timmins (2013), "The new economics of equilibrium sorting and policy evaluation using housing markets." *Journal of Economic Literature*, 51, 1007–1062.

Latimer, Douglas A. (1996), "Particulate matter source-receptor relationships between all point and area sources in the united states and psd class i area receptors." Technical report, US EPA, OAQPS.

Molloy, Raven, Christopher L Smith, and Abigail K Wozniak (2011), "Internal migration in the united states." Technical report, National Bureau of Economic Research.

Roback, Jennifer (1982), "Wages, rents, and the quality of life." *The Journal of Political Economy*, 1257–1278.

Rosen, Sherwin (1974), "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy*, 82, 34–55.

Shadbegian, R. J., W. Gray, and C. Morgan (2007), "Benefits and costs from sulfur dioxide trading: A distributional analysis." In *Acid in the Environment: Lessons Learned and Future Prospects* (G. Visiglio and D. Whitelaw, eds.), 249–259, Springer US, Boston, MA.

Tiebout, Charles (1956), "A pure theory of local expenditures." *Journal of Political Economy*, 64, 416–424.

Timmins, Christopher and Jennifer Murdock (2007), "A revealed preference approach to the measurement of congestion in travel cost models." *Journal of Environmental Economics and Management*, 53, 230–249.

Tra, Constant I (2010), "A discrete choice equilibrium approach to valuing large environmental changes." *Journal of Public Economics*, 94, 183–196.

Tra, Constant I (2013), "Measuring the general equilibrium benefits of air quality regulation in small urban areas." *Land Economics*, 89, 291–307.

Waldorf, Brigitte and Ayoung Kim (2015), "Defining and measuring rurality in the us: From typologies to continuous indices." In *Workshop on Rationalizing Rural Classifications*, Washington, D.C., URL http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_168031.pdf.