

# Tiebout Sorting and Toxic Releases\*

Dakshina G. De Silva<sup>†</sup> Anita R. Schiller<sup>‡</sup> Aurelie Slechten<sup>§</sup> Leonard Wolk<sup>¶</sup>

## Abstract

Combining detailed county-to-county migration data with Toxic Release Inventory data, and fine-scale  $PM_{2.5}$  concentration levels, we investigate the relationship between internal migration, income of migrant and non-migrant households and county-level differences in environmental quality. We show that households moving to “cleaner” counties are relatively “richer”— a result consistent with a sorting by income in the spirit of Tiebout (1956). An implication of this finding is that internal migration could contribute to the persistence of disparities in pollution exposure at the county-level.

Keywords: Migration, Toxic Release Inventory, Air Quality, Income Sorting

JEL Classification : Q50, Q53, R23

---

\*We thank George Deltas, seminar audiences from Lancaster University and KU Leuven, and participants at the 2022 Annual European Association of Environmental and Resource Economists Conference, the 48th annual Regional Science Association International-British and Irish Section conference, and the 2023 Regional Studies Association annual conference for comments and suggestions. We also thank the editor and two anonymous referees for helpful comments that significantly improved the paper. The authors declare that they have no relevant material or financial interests that relate to the research described in this paper.

<sup>†</sup>Department of Economics, Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, UK (e-mail: d.desilva@lancaster.ac.uk).

<sup>‡</sup>Department of Economics, Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, UK (e-mail: anita.schiller@lancaster.ac.uk).

<sup>§</sup>Department of Economics, Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, UK (e-mail: a.slechten@lancaster.ac.uk). *Corresponding author.*

<sup>¶</sup>Department of Finance, School of Business and Economics, Vrije Universiteit Amsterdam, the Netherlands (e-mail: l.wolk@vu.nl).

# 1 Introduction

Inequalities in exposure to pollution among income groups are a source of major public concern as they seem to persist over time (Colmer et al., 2020; Jbaily et al., 2022). From the perspective of environmental justice, the question is then whether households self-select by income across areas with different levels of environmental quality. An ancillary question is “Are these correlations between income and pollution the result of firms’ strategic decisions based on local demographic characteristics?” This paper explores the first question using internal migration data from the United States (US).<sup>1</sup> We analyze whether out-moving households’ destination choices are consistent with a sorting by income across levels of pollution.

The idea that internal migration could, at least partially, create the income-pollution correlations reported in the literature can be theoretically motivated by Tiebout (1956)’s canonical sorting model. Households “vote with their feet” and sort across areas that provide their optimal bundle of private goods, housing, and neighborhood characteristics. If environmental quality is an amenity that is valued by households, but these households differ in terms of their income, a location choice model in the spirit of Tiebout (1956) predicts a stratification of local areas by income, where poorer households end up in more polluted areas.

To explore this mechanism empirically, we use detailed county-to-county migration data from 2010 to 2014 provided by the US Internal Revenue Service (IRS). We combine this dataset with county-level pollution and demographic data for the home and destination counties. An interesting feature of the IRS dataset is that we observe the income of moving and non-moving households. To the best of our knowledge, this is the first study that directly identifies a sorting by income across areas with different levels of environmental quality using county-to-county migration data. The IRS county-to-county migration data series provides a comprehensive view of internal

---

<sup>1</sup>De Silva et al. (2016) and De Silva et al. (2021) examine the second question and find evidence that firms choose both their location and waste management expenditures based on local demographic characteristics.

migration patterns within the US. It captures information from 95% to 98% of all tax filers, making it the most extensive data source available for tracking population movements between counties.

Our outcome of interest is the relative income of households moving from a home county  $h$  to a destination county  $d$ . It is defined for each home and destination county pair  $(h, d)$  as the ratio of the average income of households moving from  $h$  to  $d$  to the average income of households staying in  $h$ . Using a linear specification with county-pair fixed effects, we estimate the relationship between the relative income of out-migrant households and environmental quality at their chosen destination. If households self-select across locations as predicted by Tiebout's sorting model, we expect out-moving households from county  $h$  to sort by income with lower-income households from  $h$  moving to the more polluted destinations among the set of destinations of county  $h$ , while higher-income households choose less polluted areas.

To measure environmental quality in a county, we consider two indicators of local pollution:  $PM_{2.5}$  concentrations (Meng et al., 2019) and the number of facilities reporting to the Environmental Protection Agency (EPA)'s Toxic Release Inventory (TRI).  $PM_{2.5}$  is known to have adverse health effect and is considered as a good general measure of air pollution. The chemicals covered by the TRI Program are typically local and were chosen because they pose a threat to the environment and human health. The data about toxic waste management is easily accessible on the EPA website, where households can obtain information about polluting facilities in their home and potential destination counties.<sup>2</sup>

In all our specifications, we use a wide range of econometric controls that could affect households' decision to move to a particular county (e.g. employment opportunities, amenities, and other demographic characteristics). Destination choices are based on the comparison between the home area and the possible destinations, and the attributes of a destination county might be viewed differently by individuals living

---

<sup>2</sup>In addition, local newspapers regularly report on the toxic waste releases of local polluting firms (Campa, 2018).

in different home counties. For this reason, all the county-level attributes are expressed as the difference between the values for the home and destination counties. To address concerns regarding the endogeneity of local air pollution levels, we adopt an instrumental variable (IV) approach, following prior research (Bento et al., 2015; Lang, 2015). This approach exploits changes in  $PM_{2.5}$  concentrations driven by a county's designation of non-attainment or re-attainment status by the EPA under the Clean Air Act Amendments (CAAA).

Consistent with the existence of a sorting by income, we find that on average, households that move to a county with a lower  $PM_{2.5}$  concentration, or less TRI reporting facilities, are relatively richer. We check for the robustness of our results by looking at various sub-samples of our data. We restrict our analysis to within-state or out-of-state migration, we exclude large and sparsely populated counties, etc. Overall, our findings suggest that household self-selection across destinations could play a role in the persistence of inequalities in exposure to pollution at the county-level.

With our specification, we observe the destinations households chose, but we do not observe what they could have selected from (their choice set). In the Appendix, we provide additional insights to address this concern by examining a destination model with home and destination counties characteristics. This allows us to study internal migration patterns and the trade-offs between moving costs, job opportunities, and other amenities faced by US households.

We contribute to the literature in several ways. We first add to the growing literature studying destination choices of internal migrants. Using the IRS data, Curtis et al. (2015) and DeWaard et al. (2016) investigate migrant destinations after Hurricane Katrina, while Frey (2009) and Molloy et al. (2011) explore more generally the possible explanations to the US migration slowdown. Davies et al. (2001) use repeated cross-sections of the IRS data to study the relationship between interstate migration, relative economic opportunities, and cost of moving.<sup>3</sup> The novelty of our paper lies

---

<sup>3</sup>A few studies use direct measures of internal migration in other countries than the US. Hatton and Tani (2005) derive a measure of internal migration from the records of the British National Health Services, while Beine and Coulombe (2018) use internal migration data derived by Statistics Canada

in the combination of migration patterns at the county level with measures of local environmental quality. This allows us to study how households sort across destination counties based on their income.

The heterogeneity in the responses to changes in location characteristics among different types of migrants has received limited attention in the internal migration literature. Recent exceptions are Baum-Snow and Hartley (2020), Aydemir and Duman (2021) and Chen et al. (2022).<sup>4</sup> Chen et al. (2022) is of particular interest for our paper, as they find different effects of air pollution on net internal migration flows in China across gender, education and occupation. Besides the fact that we focus on an alternative household attribute (income), our study differs in the identification strategy. Instead of estimating our destination choice model for different income groups, we explore the existence of a sorting by income by regressing a direct measure of relative income of out-moving households on differences in local pollution.

Research focusing on residential mobility and local environmental quality commonly uses changes in aggregate demographic characteristics and difference-in-difference (or similar) approaches to investigate whether households locally or regionally sort by income (Been and Gupta 1997; Kahn 2000; Cameron and McConnaha 2006). However, studies have shown that these approaches might be problematic. Building on a standard sorting model, Banzhaf and Walsh (2008) illustrate that the predictions of difference-in-difference models regarding the effect of changes in local environmental quality on average income are ambiguous, except for “large” changes.<sup>5</sup> More fundamentally, Depro et al. (2015) note that these approaches do not allow to identify the impact of local pollution on residential mobility because this impact depends on the

---

from income tax reports. Both studies focus on the impact of immigration on internal migration.

<sup>4</sup>Aydemir and Duman (2021) examine how the role of migrant networks at destination differs across migrant types (in terms of skills, age at migration, and reason of migration), while Baum-Snow and Hartley (2020) document that the trade-offs between local amenities and economic opportunities vary substantially across households with different socioeconomic characteristics.

<sup>5</sup>Using a two-community example, they demonstrate that improvements in the environmental quality of the most polluted area give rise to in-migration from the poorest households of the other community, leading to an increase in average income in both communities. Only large improvements (i.e improvements that reverse the ordering of communities in terms of local pollution) produce non-ambiguous predictions.

characteristics of the home and destination areas. Without information about the destination county, we don't know whether households who changed residence, moved to a more (or less) polluted area compared to the home county.

In our paper, we overcome this issue by relying on the IRS migration data, which is very detailed and refers directly to the yearly flows of in- and out-migrants in a given county. An alternative approach based on the structural estimation of households' willingness to pay is proposed in Depro et al. (2015) or Freeman et al. (2019). While their analyses concentrate on specific communities within Los Angeles County or selected cities in China, benefiting from detailed data, our study encompasses all counties in the contiguous US. One exception is Close and Phaneuf (2017) who also use the IRS county-to-county migration data to estimate residents' marginal willingness to pay to avoid air pollution. However, our analysis diverges as we specifically examine the interplay between migrants' income and local environmental quality.

## 2 Conceptual Framework

The motivation for our analysis is based on Tiebout's (1956) canonical model of residential sorting, which has inspired more recent general equilibrium models of location choice (e.g. Banzhaf and Walsh 2008). In these models, there is a set of possible locations that differ in terms of their level of public goods or amenities and households have to choose where to live subject to a budget constraint. Because households prefer neighborhoods with "better amenities", the housing demand in those areas is higher, leading to higher housing prices. Households are therefore facing a trade-off between consumption and local amenities. In particular, a low-income household might not be willing to pay as much as a high-income household to live in a high-amenity neighborhood, as they have to prioritize necessary goods, such as everyday clothing or food. At equilibrium, if households differ only in terms of their income, they will sort by income across locations with different levels of amenities. If the local amenity of interest is environmental quality, Tiebout's result implies that poorer households end up in more

polluted areas, while richer households can afford areas with better environmental quality.

To examine whether the observed correlations between income and pollution could be, at least partially, attributed to households self-sorting across locations by income, we rely on internal migration data and explore the relationship between destination choices and income of households changing residence in a given year. Consider the set of households who decided to move from their home area  $h$  in year  $t$  and the set of associated destinations  $d \in \{1, \dots, N\}$  that differ in their level of environmental quality. If these households choose where to move subject to a budget constraint, a residential sorting model in the spirit of Tiebout (1956) predicts that out-migrant households sort by income level from the most polluted to the least polluted destinations in  $\{1, \dots, N\}$ . Out-migrant households with an average income higher than the average income of households staying in area  $h$  move to locations that are less polluted than their home area and we expect the least polluted destinations to attract the households with the highest income (relative to the income of households in  $h$ ). Similarly, out-migrant households with an average income lower than the average income of households staying in  $h$  move to more polluted destinations.

This motivates the following empirical specification:

$$\text{Relative Income}_{d,h,t} = \gamma_1 \Delta E_{d-h,t-1} + \gamma_2 \Delta X_{d-h,t-1} + \alpha_{dh} + \tau_t + \varepsilon_{d,h,t} \quad (1)$$

where our dependent variable is the relative income of out-migrant households and is defined for each pair of home and destination locations  $(h, d)$  as the ratio of the average income of households moving from  $h$  (where they were living in year  $t - 1$ ) to  $d$  to the average income of households staying in county  $h$  in year  $t$ :

$$\text{Relative Income}_{d,h,t} = \left( \frac{\text{average household income of out-migrants}}{\text{average household income of non-migrants}} \right)_{d,h,t}$$

The vector  $\Delta E_{d-h,t} = E_{d,t} - E_{h,t}$  contains our pollution measures and indicates how polluted a destination  $d$  ( $E_{d,t}$ ) is, relative to the level of pollution in the home area

( $E_{h,t}$ ).  $\Delta X_{d-h,t} = X_{d,t} - X_{h,t}$  includes a set of control variables (differences between destination and home characteristics). This specification captures the idea that destination choices are based on the comparison between the home area and the possible destinations. The location-pair fixed effects,  $\alpha_{dh}$ , account for the unobservable time-invariant heterogeneity among pairs of locations.  $\tau_t$  represents the year fixed effects and  $\varepsilon_{d,h,t}$  is the error term. In equation (1),  $\Delta E_{d-h,t}$  and  $\Delta X_{d-h,t}$  are adjusted for a one-year lag to capture the environmental and socio-economic conditions prevalent in the home and destination areas during the year preceding households move to their new residence in area  $d$ .

The coefficient of interest is  $\gamma_1$ . Given the characteristics of a home area  $h$ , it captures how the relative income of out-moving households is associated with differences in pollution across destinations. A sorting by income across destinations is consistent with  $\gamma_1 < 0$ : when the level of pollution at destination decreases relative to the home area, the income of households who choose that destination increases (relative to the income of households staying back).

### 3 Data and preliminary evidence

We compile a comprehensive county-level data file on local pollution and migration for the lower 48 states in North America. We supplement this data file with county-level demographic data. We provide descriptive statistics in Table 1. We have data for 3,109 counties over a five-year sample period, which gives us 15,545 county-year observations. 37% of counties are categorized as metropolitan counties and 42.5% are defined as urban counties.<sup>6</sup> Finally, about 7% of the counties are coastal counties (excluding the Great Lakes).

---

<sup>6</sup>To define metropolitan (metro), urban and rural counties, we use the 2013 rural-urban continuum codes provided by the US Department of Agriculture’s Economic Research Service. Counties are divided as either metro or non-metro counties (2013 Office of Management and Budget). Metro counties are then distinguished by the population size of their metro area, and non-metropolitan (non-metro) counties by their degree of urbanization and adjacency to a metro area.



### 3.1 Migration data

We gather county-to-county migration data from the IRS data files for the 3,109 counties located in the contiguous US between 2010 and 2014. The IRS uses individual federal tax returns between two tax years and identifies migrants and non-migrants in a county.<sup>7</sup> A snapshot of this data, where the home county is Autauga in Alabama (State code = 1 and County code = 1), is provided in Table A1 (Appendix A). For example, we observe that 10 households move to Shelby County in Tennessee from Autauga in Alabama and, on average, they have the highest level of income—they earn about 2.17 times more than non-migrants in Autauga County. By contrast, the average income of households moving to Clayton County in Georgia is 86% lower than the average income of non-migrants in Autauga.

As shown in Table 1, around 1,800 households move out of a county every year and household net migration is slightly positive.<sup>8</sup> The average relative income is 0.717 indicating that the income of households moving from a home county  $h$  to a different county was on average 28% lower than the income of households staying in county  $h$ . In Figure 1, we present the average net migration for all US counties between 2010 and 2014. Counties with positive net migration are blue and those with negative net migration are red. We observe that many households move to Florida, the suburbs of Dallas, Houston, and San Antonio/Austin areas, while many seem to move out of Northeast counties.

In terms of destination choices, 2,968 counties sent households to at least one other county in the US in 2010, with an average of 25 different destinations per county. Finally, every year, the proportion of out-moving households relocating to a different state is about 30%.

The IRS dataset is considered as an attractive data source to conduct migration

---

<sup>7</sup>Note that the income reported in this dataset reflects the income earned during the year preceding the filing year. Depending on the exact date at which a household moved from one county to the other, this income may include income earned both in the home and destination county.

<sup>8</sup>Net migration in a county in a given year is defined as the difference between the total number of households that move in that county from all the other counties (in-migration) and the total number of households that move out (out-migration).

research in the US (Frey, 2009; Molloy et al., 2011; Curtis et al., 2015). Because this dataset is based on administrative records, it is available annually and relatively comprehensive (it covers 95-98% of the US tax filers and their dependents). There is one main limitation of this dataset. Even though the overwhelming majority of householders file tax returns, some categories of the population are most likely to be underrepresented in the data, namely the undocumented populations, the elderly, and college students (Gross, 2003; DeWaard et al., 2016). Hauer and Byars (2019) provide a comparison of the three main sources of migration data in the US (i.e. the Decennial Census long form, the American Community Survey, and the IRS data). They note that, despite the limitation in terms population coverage, the IRS dataset is the largest migration data source for count flows between counties (e.g. the ACS data contains about 2% of the observations in the IRS data) and is more appropriate for researchers interested in annual comparisons of migration patterns.<sup>9</sup>

### 3.2 Pollution measures

For each county, we construct two measures that capture different aspects of local environmental quality.

#### ***PM<sub>2.5</sub> data***

$PM_{2.5}$  is an ambient fine particulate matter with a diameter that is generally 2.5 micrometers and smaller.  $PM_{2.5}$  can remain airborne for long periods and travel hundreds of miles. Concentration in a given location consists of both locally emitted  $PM_{2.5}$  (due to industrial activity or traffic congestion), but also pollution released elsewhere that is transported by the wind. Particulate Matters are a widely used measure of local environmental quality (Deryugina et al. 2019; Chen et al. 2022; Greenstone and Hanna 2014). It is well established that ambient  $PM_{2.5}$  has adverse effects on the human respiratory system, especially for children, and increases the mortality risk.

---

<sup>9</sup>Hauer and Byars (2019) explain that this data exists in multiple files making its analysis rather cumbersome. This has probably hindered the widespread adoption of this valuable resource for US migration scholarship.

This paper takes advantage of the availability of recent and fine-scale *annual*  $PM_{2.5}$  *concentration* estimates compiled by Meng et al. (2019).<sup>10</sup> The annual ambient  $PM_{2.5}$  concentration data is available at a 0.01 degree by 0.01 degree resolution. We therefore map the concentration estimates to US counties boundaries.<sup>11</sup> The average  $PM_{2.5}$  concentration at the county level is  $7.63\mu g/m^3$  over our sample period.

### ***TRI data***

The TRI is a US database established by law that requires private and government facilities to report annually their waste management and pollution prevention activities. The reporting requirements are detailed in Section 313 of the EPCRA (Emergency Planning and Community Right to Know Act). A plant has to report to the TRI if that plant belongs to a North American Industry Classification System (NAICS) code, identified by the EPA’s TRI Program, and manufactures, processes or uses designated hazardous or toxic chemicals above a reporting threshold set by the EPA. In addition, these plants must have at least 10 full-time employees. The facilities subject to mandatory reporting are denoted as TRI reporting facilities or TRI reporters.

The TRI Program covers hundreds of chemicals that are known to pose a threat to human health and the environment, including lower birth weight or higher infant mortality rates (Currie and Schmieder 2009; Agarwal et al. 2010; Currie et al. 2015). The EPA website provides plant-level waste management information on quantities of toxic waste recycled, combusted for energy recovery, treated, released (to the air, water, and land) or otherwise disposed of, both on- and off-site for each chemical. The EPA has also developed a toxicity-weighted index, which gives the total plant-level environmental releases (on-site and off-site) across all media and all chemicals (in pounds).

---

<sup>10</sup>These estimates were obtained from chemical transport modeling, satellite remote sensing, and ground-based measurements.

<sup>11</sup>Note that these estimates of  $PM_{2.5}$  concentration provide a better measure of the county-level concentrations than air pollution data from the EPA’s Air Quality System database. This database provides hourly data at the pollution-monitor level for pollutants that are regulated by the Clean Air Act. A county’s pollution-monitor readings may not adequately measure the average pollution exposure for county residents due to the sparse placement of monitors within counties (Deryugina et al., 2019).

We construct our TRI-related measure in two different ways: (1) *the number of TRI reporting facilities in a county*, and (2) *county-level total on-site toxic releases*.<sup>12</sup> Note that we exclude off-site releases as they correspond to toxic chemicals transferred to a receiving facility (for disposal), which may not necessarily be located in the same county. Given the self-reported nature of TRI releases, the number of TRI reporting facilities in a county is our preferred measure of pollution.

Even though preliminary data on plant-level waste management practices is available with a one-year lag on the EPA website, the final TRI National Analysis is released by the EPA with a lag of two years. For example, the TRI data available to potential migrant households in 2010 provide information about toxic chemicals released in 2008. For this reason, we use the number of TRI reporting facilities and total toxic releases on-site lagged by two years in all our estimations. On average, there are about 7 TRI reporting facilities in a county and about 960,000 pounds of toxic releases (on-site and off-site) per county. 21.3% of counties, did not have any TRI reporting facility over the sample period (about half of them were metro or urban counties). These numbers are relatively stable overtime.

Compared to  $PM_{2.5}$ , which can travel thousands of miles, the chemicals covered by the TRI Program are very local pollutants and a county might be too large to represent the population closely exposed to polluting activities.<sup>13</sup> However, our focus is not on the actual health impact of toxic chemicals but rather on household reactions' to an increase in the number of polluting facilities in their area. The publication of TRI data might play a role in their migration decisions even if they are not directly affected by the presence of a TRI facility.<sup>14</sup>

---

<sup>12</sup>Not all TRI reporting plants necessarily report chemical releases, as they might be managing their toxic waste in a different way. We provide an example of TRI reporters and total releases in Table A2. The facilities with TRI plant IDs *36067NNCMP100JE* and *3606WSHRMN74CUN* reported releases for all the years shown in the sample while the plant with TRI plant ID *36066TWBCL17JES* did not have any releases in 2010 and 2011, but still reported to the TRI Program. In our dataset, they all appear as TRI reporting facilities for the entire sample period.

<sup>13</sup>Using individual level data, Currie et al. (2015) show that the openings or closings of toxic plants (i.e. plants reporting a release to the TRI Program) have an impact on birth outcomes within a 1-mile radius of the plant location. Other studies (Currie and Schmieder, 2009; Agarwal et al., 2010) identify health effects of TRI chemicals at the county level

<sup>14</sup>Some studies using hedonic methods (Currie et al., 2015; Mastromonaco, 2015) have found that

Since the first public announcement in June 1989, the TRI data has received continuous media attention. Hamilton (2005) shows that public concerns about toxic pollution can spillover to adjacent areas through media reports. Saha and Mohr (2013) document that many of the newspaper articles released shortly after the publication of EPA data reported toxic releases from the largest polluters in relatively large areas (States, Counties, or Metropolitan areas).

### 3.3 Other factors affecting migration decisions

Based on the determinants identified in the migration literature, we consider 3 categories of control variables ( $\Delta X_{d-h,t}$ ) that affect county-to-county migration patterns and might also lead to a sorting by income: (1) factors related to income and employment opportunities, (2) other local amenities, and (3) social and demographic characteristics.

The significant expenses associated with changing residence, including moving costs, have been recognized as key factors influencing migration patterns within the migration literature. While the distance between counties serves as a standard measure of such costs, in our model, we capture this factor through the incorporation of county-pair fixed effects.

#### 3.3.1 Income and employment opportunities

Previous research (Davies et al., 2001; Hatton and Tani, 2005; Beine and Coulombe, 2018) find that differences in earnings, local economic or labor market conditions are correlated with households' decision to change residence. The *Unemployment rate* at the county level is compiled from the US Department of Commerce, Bureau of Economic Analysis, while the *annual household median income* comes from the Small

---

the information provided by the TRI affects the housing market within one or two miles of the plant location only, but Bayer et al. (2009) show that costs associated with re-location might cause hedonic methods (which typically assume that individuals are free to move) to underestimate households willingness-to-pay for air quality.

Area Income and Poverty Estimates Program (US Census Bureau).<sup>15</sup> Over our sample period, the median household income is about \$45,000, while the unemployment rate is about 8%.

We also collect data from the Quarterly Census of Employment and Wages (QCEW) provided by the Bureau of Labor Statistics (BLS). The county business patterns report the number of establishments and employment by industry. To account for the size of the economy in the county or the employment opportunities (without capturing activities related to TRI chemicals), we consider the *number of non-TRI reporting establishments*, which are all the establishments that are not in a NAICS code identified by the TRI Program. To address the concern that our measures of local pollution capture the industrial composition of a county, we control for the *Number of manufacturing establishments*, defined as the number of establishments in the NAICS codes 31-33.

### 3.3.2 Other amenities and county demographic characteristics

Counties with desirable recreational amenities might be more attractive to moving households. Using the county business patterns from the BLS, we compute the county-level *number of amenity establishments* as the number of establishments in NAICS 71 (Arts, Entertainment, and Recreation), and NAICS 72 (Accommodation and Food Services). We account for Metro, urban, and rural moving patterns within a county pair using a dummy variable to identify movements from a Metro or urban county to a rural county.

As suggested by some studies (McCormick, 1997; Hatton and Tani, 2005), house prices could be an important driver of destination choices. House prices could also capitalize the value of local amenities. In our empirical analysis, we use the *house price index* from the Federal Housing Finance Agency, which captures the evolution of house prices within an area.<sup>16</sup> We expect areas with a higher index (i.e. that

---

<sup>15</sup>The SAIPE are available for all counties, starting in 2011. For 2009 and 2010, we linearly extrapolate the county median income based on the data from 2011-2014.

<sup>16</sup>We use 2008 as a base year.

have experienced a larger increase in house prices relative to the base year) to attract relatively richer households.

Internal migration and destination choices are also affected by demographic characteristics, e.g. age and level of education (Baum-Snow and Hartley, 2020), presence of minorities (Boustan, 2010), or population density (Davies et al., 2001).<sup>17</sup> In our empirical specification, we include the following county-level demographic characteristics: college ratio (American Community Survey), estimated county population, median age and Black and Hispanic population ratios (US Census Bureau, Population Division).<sup>18</sup> Over our sample period, the average county has around 100,000 inhabitants, and the average Black and Hispanic ratios are both around 9%.

### 3.4 Preliminary evidence of disparities in destination choices

Figures 2-4 illustrate the large inter-county variations in pollution over our sample period. Darker red indicates a higher level of pollution ( $PM_{2.5}$  concentration, number of TRI reporters normalized by each county’s total number of establishments or toxic releases), while non-polluting counties are in white. TRI reporters (and toxic releases) are present in counties all across the US (other than in the middle of the country). This is very different from Figure 2, in which the highest  $PM_{2.5}$  concentrations are all located in the East part of the country. This suggests that the TRI and  $PM_{2.5}$  pollution measures capture different aspects of pollution. Over our sample period, the correlation between  $PM_{2.5}$  and the number of TRI reporting facilities is 0.120.

A prerequisite for identifying a sorting by income using migration data is that there is enough variation in out-migrant income and county-level pollution across destina-

---

<sup>17</sup>Davies et al. (2001) document that in the context of state-to-state migration, individuals are more likely to move to more populated areas and provide two interpretations for this result: (1) these areas are perceived as offering better economic opportunities, and (2) search costs are lower as information about available opportunities in those areas is usually more easily accessible.

<sup>18</sup>The college ratio is defined as the proportion of the population (older than 25 years old) with a college degree. The Black (respectively Hispanic) population ratio is defined as the Black (respectively Hispanic) population divided by the total county population. The Black (respectively Hispanic) population are the individuals who indicated “Black or African American alone” (respectively “Hispanic alone”) as the response to the question “What is this person’s race?” in the US Census.

tions chosen by households from a given home county. Table 2 provides descriptive statistics by year for home counties with at least two destinations. In 2010, the average home county has a relative income range of 0.99. In other words, given our definition of relative income (see section 2), the difference in income between the richest and the poorest out-migrant households is on average equivalent to the home county median income. For the average home county in 2010, the range of  $PM_{2.5}$  concentrations at destination is  $3.53\mu g/m^3$ . This corresponds to approximately 1/3 of the range of  $PM_{2.5}$  concentrations across all counties in 2010.

## 4 Instrumental Variables for $PM_{2.5}$

The evolution of air pollution levels over time is potentially endogenous due to the presence of omitted variables that could be correlated with both local air pollution and households' decisions to relocate. To address this concern, an important body of literature (Chay and Greenstone, 2005; Bento et al., 2015; Lang, 2015; Isen et al., 2017) uses EPA non-attainment designations after the enactment of the CAAA as an instrument to measure changes in air pollution.

### 4.1 The Clean Air Act Amendments (CAAA)

In response to the adverse health impacts of consistently elevated concentrations of major air pollutants, the US Congress enacted the Clean Air Act of 1970. Major amendments were subsequently added in 1977 and 1990. A pivotal component of the 1970 Clean Air Act is the establishment of federal air quality standards, known as the National Ambient Air Quality Standards (NAAQS), for key air pollutants. This legislation mandates the EPA to designate each county as either in attainment or non-attainment status for each pollutant, contingent upon whether the relevant standard is exceeded.

A county's designation as out of attainment has important implications due to stringent regulations imposed by the CAAA on polluting entities within non-attainment



areas. Specifically, states and counties are obligated to devise a State Implementation Plan (SIP) for EPA approval, outlining strategies to mitigate pollution levels in non-attainment counties and establish plant-specific regulations for each significant pollution source.

In 1987, the EPA began regulating particulates less than 10 micrometers in diameter ( $PM_{10}$ ), for which the negative health effects were deemed particularly severe.<sup>19</sup> Designation of non-attainment areas for  $PM_{10}$  occurred in 1990. In July 1997, EPA promulgated new standards for  $PM_{2.5}$  specifically and federal designation of attainment status began in 2005 (based on 2001 through 2003 air quality monitoring data), with a subsequent revision in 2006 and re-designation starting in 2009 (based on concentrations from 2006 to 2008). After consistently meeting the EPA’s standard for a pollutant for three consecutive years, a county can apply for designation as a “re-attainment county”. However, if the county fails to maintain this standard at any point, it is re-designated as a non-attainment county.

## 4.2 Attainment status as an instrument

We focus on changes in county’s attainment designation for particulate matter emissions to identify exogenous changes in air pollution. We extract county-level attainment status data for the six key pollutants covered by the 1990 CAAA from the “Green Book Non-attainment Areas” available on the EPA’s website.<sup>20</sup> In 2009 (beginning of the re-designation following the revision of  $PM_{2.5}$  standards), 265 counties in our sample were in non-attainment status for  $PM_{2.5}$ . 89 counties changed status over our sample period.

The model outlined in section 2, i.e. equation (1), is based on disparities in characteristics among county pairs. Our primary instrument for the difference between desti-

---

<sup>19</sup>The 1970 Clean Air Act authorized the EPA to enforce a NAAQS for total suspended particles (particles less than 100 micrometers in diameter).

<sup>20</sup>The legal framework permits counties to be categorized as “partial attainment”, wherein only specific geographic regions within the county are designated as non-attainment. For analytical purposes, we treat these partially attaining counties as non-attainment counties, as they remain subject to regulatory measures.

nation and home county  $PM_{2.5}$  concentrations is therefore the divergence in attainment status for particulate matters ( $\Delta PM NonAttainment_{d-h,t} = PM NonAttainment_{d,t} - PM NonAttainment_{h,t}$ ). Rather than employing a simple binary indicator capturing whether a county is in attainment or not, we adopt a more nuanced approach reflecting the persistence of non-attainment status and its potentially heterogeneous impact on air quality improvements. For each county  $k$  in our sample ( $k = h, d$ ), the variable  $PM NonAttainment_{k,t}$ , spans values from 0 to 3. It is equal to zero if county  $k$  is designated in attainment in year  $t$ . Values from 1 to 3 indicate the frequency of non-attainment designation up to year  $t$  (based on the 1987, 1997 and 2006 standards).

Moreover, we consider an over-identified model by adding a second instrument,  $\Delta Attainment County_{d-h,t} = Attainment County_{d,t} - Attainment County_{h,t}$ , where the dummy variable  $Attainment County_{k,t}$  indicates whether a county  $k$  ( $k = h, d$ ) is designated as a non-attainment county for any pollutant covered by the CAAA in year  $t$ . Formally, the first-stage regression in this two-stage least squares estimator is:

$$\begin{aligned} \Delta \text{Log}(PM_{2.5})_{d-h,t} = & \beta_1 \Delta PM NonAttainment_{d-h,t} + \beta_2 \Delta Attainment County_{d-h,t} \\ & + \beta_3 \Delta X_{d-h,t} + \alpha_{dh} + \tau_t + \nu_{d,h,t} \end{aligned} \quad (2)$$

The other controls ( $\Delta X$ ) are the same as in our baseline model (1). This formulation is consistent with Auffhammer et al. (2009) who examine how a county's designation in year  $t$  affects  $PM_{10}$  concentrations in year  $t$  at different monitors located in this county. It reflects the fact that non-designations in year  $t$  are based on monitoring data from the past three years (see also Lang 2015).<sup>21</sup> In the second stage, we use the predicted differences in  $PM_{2.5}$  concentrations from equation (2) in place of the actual values in equation (1).

Differences in attainment status of destination and home counties serve as a valid

---

<sup>21</sup>As a robustness check, we also run our models including  $\Delta PM NonAttainment_{d-h,t-1}$  instead of  $\Delta PM NonAttainment_{d-h,t}$  in equation (2). The coefficients estimated in the second stage regression are virtually identical to those shown in Table 3.

instrument. Prior research confirms that county-level non-attainment status is correlated with subsequent air pollution levels within a county (Auffhammer et al., 2009; Lang, 2015). Moreover, conditional on other observable home and destination counties characteristics, differences in non-attainment status affect migration decisions only through their impact on local air pollution. One issue would be if efforts aimed at reducing pollution to maintain attainment (e.g., limiting manufacturing activity) independently affect location choices. However, we believe that with our extensive set of control variables (including, for example, the number of manufacturing establishments, or the unemployment rate), we effectively control for such possible channels.

## 5 Results

Based on the conceptual framework outlined in section 2, we explore the existence of a sorting by income across US counties by estimating how the relative income of households moving from county  $h$  to county  $d$  covaries with differences in environmental quality between home and destination counties (coefficient  $\gamma_1$  in equation (1)). For a given year, we can only compute the relative income of a county pair  $(h, d)$  if some households move from  $h$  to  $d$ . Therefore, we don't have a balanced panel.

With our specification, we observe the destinations households selected, but we don't observe their original choice set. To gain insights into the factors influencing households' destination choices, we analyze, in Appendix B, the correlation between household migration flows and the characteristics of both the origin and destination counties. We find that disparities in local air pollution levels ( $PM_{2.5}$ ) are strongly associated with the patterns of migration between counties. In other words, alongside employment opportunities (proxied by the difference in unemployment rates between origin and destination) and moving costs (proxied by the distance between two counties), environmental quality seems to play a role in households' selection of destination. However, this analysis does not allow us to identify sorting by income, as households moving to different locations might have heterogeneous attributes.

In Table A3, we provide summary statistics for the regression variables in equation (1). The average relative income for a county pair  $(h, d)$  is lower than 1, suggesting that households moving to a different county have a lower income than households staying back. At the same time, the difference in the number of TRI reporting plants between the destination county and home county is slightly positive. On average, the destination counties have more TRI reporting facilities than the home counties.

The estimation results of equation (1), with county-pair fixed effects and time effects, for all US counties are presented in Table 3.<sup>22</sup> In all specifications, we cluster standard errors at county-pair level. For completeness, Columns 1 and 2 report the results from OLS specifications. Columns 3 and 4 show the results when  $\Delta \text{Log of } PM_{2.5}$  is instrumented using differences in counties attainment status ( $\Delta PM \text{ NonAttainment}$  and  $\Delta \text{Attainment County}$ ). For these specifications, first-stage results, the Hansen  $J$ -statistic and relevant  $F$ -statistics are reported. In Columns 1 and 3, the TRI-related measure is the number of TRI reporting facilities, while Columns 2 and 4 show the results when the TRI-related measure is the county-level total toxic releases on site.

For the variables  $\Delta \text{Log number of TRI reporters}_{d-h,t}$  and  $\Delta \text{Log total releases on site}_{d-h,t}$ , a zero may have two different significations. First, it is possible that the destination and home counties' TRI reporting plants or total releases are the same. Second, it is possible that both counties have no TRI reporting facilities or no toxic releases. To identify this condition, we include a dummy when both counties do not have any TRI reporters and a dummy when both counties do not have any toxic releases.

In all specifications, the coefficients of our pollution measures have the expected sign. When the county-level  $PM_{2.5}$  concentration (or the number of TRI reporting facilities) increases in the destination county compared to the home county, the relative income of households moving to that county decreases. In the IV specifications (Columns 3 and 4), both the TRI-related measures and  $PM_{2.5}$  levels are statistically

---

<sup>22</sup>We estimate equation (1) for various subsets of our control variables and results (available upon request) remain qualitatively unchanged.

significant at a level of 1%. Our results are therefore consistent with a sorting by income across destinations with different levels of environmental quality: households that move to “cleaner” counties are relatively “richer”. In Columns 3 and 4, first-stage results confirm that our instruments are highly correlated with  $PM_{2.5}$  concentration levels. Moreover, the  $F$ -Statistic suggests that weak instruments are not an issue, while the p-value associated with the Hansen  $J$ -statistic implies that we cannot reject that the choice of instruments are valid. It is also interesting to note that the IV estimates of the  $PM_{2.5}$  coefficients are much larger (in absolute terms) than OLS estimates.

Beyond these variables of particular interest to us, estimates in Table 3 reveal that households also sort across locations with different economic opportunities. Destinations where the median income and the level of economic activity (measured by the number of non-TRI establishments) are higher or the unemployment rate is lower (relative to the home county) attract on average relatively richer households.

Wealthier households also tend to prefer less populated areas (everything else being equal). As expected, there is a positive (but only statistically significant at a level of 10%) association between the relative household income of out-migrants and the difference in house price index, i.e. households moving to counties where property values have increased more (relative to the base year) are “richer”. Finally, differences in Hispanic population ratio between the home and destination counties are strongly correlated with the relative household income of out-migrants. Destination counties with a larger proportion of Hispanic residents (compared to the home county) attract relatively poorer households.

## 5.1 Robustness checks

In this section, we explore the robustness of the relationship between differences in pollution measures and relative income of out-migrants. All the results are reported in Tables A4 and A5 and are obtained by estimating equation (1) using the IV estimator

with county-pair fixed effects and time effects.

### 5.1.1 Within-state and out-of-state migration

The 2019 Current Population survey documents that the primary motivation for moving varies by type of movers. In particular, long-distance moves are primarily motivated by employment opportunities, while shorter-distance moves are mostly associated with housing-related reasons.

Table A4 (panel A) reports the estimates of the main variables of interest for the within-state sample in column 2 (home and destination counties belong to the same state) and out-of-state samples in column 3 (column 1 shows the results for all contiguous counties and is the same as column 3 in Table 3). The results for the other explanatory variables are available upon request. The  $PM_{2.5}$  concentrations and TRI-related measures have the expected sign. However, the coefficient associated with the number of TRI reporters is statistically insignificant for interstate migration.

### 5.1.2 Personal Income

Next, we use relative individual income (instead of household income) as our dependent variable in equation (1). The personal income of out-migrants (respectively non-migrants) is obtained by dividing the aggregate income of out-migrants (respectively non-migrants) by the number of individuals moving out (respectively staying back). The sign and magnitude of the coefficients associated with differences in  $PM_{2.5}$  concentrations (Table A4, panel B) are similar to those obtained using relative household income in Table 3 and panel A of Table A4.

### 5.1.3 Sub-samples of counties

To address the concern that counties are relatively large areas in which pollution or income might vary significantly, we estimate equation (1) for different sub-samples of our data. First, we restrict our sample to (1) county-to-county migration within metro counties (which are typically smaller), (2) county-to-county migration within urban

and metro counties, or (3) county-to-county migration from urban/metro counties to any type of county. Results are reported in Table A5 (columns 1-3). Second, we exclude from our sample the 10% largest counties in terms of land area (column 4). Finally, in columns 5 and 6, we split our sample of counties into two groups based on their population density.<sup>23</sup> In sparsely populated counties, residential options are likely limited to a few small towns. As a result, individuals may have fewer choices about where to live to avoid pollution. In contrast, densely populated counties offer households greater flexibility to relocate within the county. Our results for  $PM_{2.5}$  remain qualitatively the same for all sub-samples of counties.

#### 5.1.4 Crime rate

The crime rate is another type of amenity known to affect migration decisions, we do not include this variable in our main analysis as the FBI crime rate in per capita terms is missing for around one third of our observations, mostly rural counties. Note that our results (available upon request) remain qualitatively unchanged when we estimate our model with crime rate as an additional control.

## 6 Conclusion

This study introduces a novel method for examining the presence of income-based sorting across locations with diverse environmental qualities. Our empirical approach integrates detailed county-to-county IRS migration data, fine-scale  $PM_{2.5}$  concentrations data, and TRI data to explore the relationship between pollution levels and the migration patterns of US households from 2010 to 2014. Our findings suggest that destination counties with lower pollution levels than the migrants' home counties attract wealthier households. From the perspective of Environmental Justice, this outcome is consistent with households self-selecting based on income across areas with different

---

<sup>23</sup>We rank all the counties according to their population density in 2010 (first year of our sample) and counties in the top decile are categorized as high density counties (column 6 of Table A5). The results for the other lower-density counties are shown in column 5.

levels of environmental quality. Furthermore, our research contributes to the existing literature on internal migration by emphasizing how socioeconomic characteristics shape households' responses to differences in destination attributes.

Given the growing political attention to inequalities in pollution exposure, our findings hold relevance for informing environmental policies. In particular, our results illustrate that addressing environmental justice issues requires not only refining air-quality standards and TRI reporting but also addressing income disparities, as households "vote with their feet".

However, this study has its limitations. Firstly, beyond income, a fundamental issue underlying the Environmental Justice movement is the correlation between race and pollution exposure. It would be interesting to apply the methodology proposed in this paper to investigate how differences in local environmental quality influence the racial/ethnic composition (as defined and collected by the US Census) of migration flows between different counties. Unfortunately, the lack of such information in the IRS data does not allow us to perform this analysis.

A second potential limitation of our study is the necessity to aggregate pollution data at the county level due to the availability of IRS data only at that level. As highlighted by Banzhaf et al. (2019), this aggregation may hide within-county variations in pollution exposure and give rise to the "ecological fallacy". However, these authors note that the ecological fallacy generally tends to mask environmental injustices in coarser data, suggesting that our results could be interpreted as conservative lower bounds. An ideal future extension would involve combining county-level migration data with detailed individual data to better capture household-specific and within-county pollution variations.

Finally, in our sample, we only observe household income when households relocate from one location to another. The original choice set remains unobserved. While the destination choice model in Appendix B partially addresses this concern, fully characterizing and modeling the complete choice set would extend beyond the scope of this paper. Despite these limitations, our work underscores the importance of



considering socioeconomic factors in addressing environmental inequalities.

## References

- Agarwal, N., C. Banerghansa, and L. T. Bui (2010). Toxic exposure in America: Estimating fetal and infant health outcomes from 14 years of TRI reporting. *Journal of Health Economics* 29(4), 557–574.
- Auffhammer, M., A. M. Bento, and S. E. Lowe (2009). Measuring the effects of the clean air act amendments on ambient pm10 concentrations: The critical importance of a spatially disaggregated analysis. *Journal of Environmental Economics and Management* 58(1), 15–26.
- Aydemir, A. and E. Duman (2021). Migrant Networks and Destination Choice: Evidence from Moves across Turkish Provinces. *IZA Discussion Paper*.
- Banzhaf, H. S. and R. P. Walsh (2008). Do people vote with their feet? An empirical test of Tiebout. *American Economic Review* 98(3), 843–63.
- Banzhaf, S., L. Ma, and C. Timmins (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives* 33(1), 185–208.
- Baum-Snow, N. and D. Hartley (2020). Accounting for central neighborhood change, 1980–2010. *Journal of Urban Economics* 117, 103228.
- Bayer, P., N. Keohane, and C. Timmins (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management* 58(1), 1–14.
- Been, V. and F. Gupta (1997). Coming to the nuisance or going to the barrios—a longitudinal analysis of environmental justice claims. *Ecology Law Quarterly* 24, 1.
- Beine, M. and S. Coulombe (2018). Immigration and internal mobility in Canada. *Journal of Population Economics* 31(1), 69–106.

- Bento, A., M. Freedman, and C. Lang (2015). Who benefits from environmental regulation? Evidence from the Clean Air Act Amendments. *Review of Economics and Statistics* 97(3), 610–622.
- Boustan, L. P. (2010). Was postwar suburbanization “white flight”? Evidence from the Black migration. *The Quarterly Journal of Economics* 125(1), 417–443.
- Cameron, T. A. and I. T. McConnaha (2006). Evidence of environmental migration. *Land Economics* 82(2), 273–290.
- Campa, P. (2018). Press and leaks: Do newspapers reduce toxic emissions? *Journal of Environmental Economics and Management* 91, 184–202.
- Chay, K. Y. and M. Greenstone (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy* 113(2), 376–424.
- Chen, S., P. Oliva, and P. Zhang (2022). The effect of air pollution on migration: evidence from China. *Journal of Development Economics* 156, 102833.
- Close, B. T. and D. J. Phaneuf (2017). Valuation of local public goods: Migration as revealed preference for place.
- Colmer, J., I. Hardman, J. Shimshack, and J. Voorheis (2020). Disparities in PM2.5 air pollution in the United States. *Science* 369(6503), 575–578.
- Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review* 105(2), 678–709.
- Currie, J. and J. F. Schmieder (2009). Fetal exposures to toxic releases and infant health. *American Economic Review* 99(2), 177–183.
- Curtis, K. J., E. Fussell, and J. DeWaard (2015). Recovery migration after Hurricanes Katrina and Rita: Spatial concentration and intensification in the migration system. *Demography* 52(4), 1269–1293.

- Davies, P. S., M. J. Greenwood, and H. Li (2001). A conditional logit approach to US state-to-state migration. *Journal of Regional Science* 41(2), 337–360.
- De Silva, D. G., T. P. Hubbard, and A. R. Schiller (2016). Entry and exit patterns of “toxic” firms. *American Journal of Agricultural Economics* 98(3), 881–909.
- De Silva, D. G., R. P. McComb, A. R. Schiller, and A. Slechten (2021). Firm behavior and pollution in small geographies. *European Economic Review*, 103742.
- Depro, B., C. Timmins, and M. O’Neil (2015). White flight and coming to the nuisance: can residential mobility explain environmental injustice? *Journal of the Association of Environmental and Resource Economists* 2(3), 439–468.
- Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review* 109(12), 4178–4219.
- DeWaard, J., K. J. Curtis, and E. Fussell (2016). Population recovery in New Orleans after Hurricane Katrina: Exploring the potential role of stage migration in migration systems. *Population and Environment* 37(4), 449–463.
- Freeman, R., W. Liang, R. Song, and C. Timmins (2019). Willingness to pay for clean air in China. *Journal of Environmental Economics and Management* 94, 188–216.
- Frey, W. (2009). The great American migration slowdown. *Brookings Institution, Washington, DC*.
- Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, 681–700.
- Greenstone, M. and R. Hanna (2014). Environmental regulations, air and water pollution, and infant mortality in India. *American Economic Review* 104(10), 3038–72.

- Gross, E. (2003). US population migration data: Strengths and limitations. *Internal Revenue Service Statistics of Income Division, Washington, DC*. [http://www.irs.gov/pub/irs-soi/99gross\\_update.doc](http://www.irs.gov/pub/irs-soi/99gross_update.doc).
- Hamilton, J. (2005). *Regulation through revelation: the origin, politics, and impacts of the Toxics Release Inventory Program*. Cambridge University Press.
- Hatton, T. J. and M. Tani (2005). Immigration and inter-regional mobility in the UK. *The Economic Journal* 115(507), 342–358.
- Hauer, M. and J. Byars (2019). IRS county-to-county migration data, 1990–2010. *Demographic Research* 40, 1153–1166.
- Isen, A., M. Rossin-Slater, and W. R. Walker (2017). Every breath you take—every dollar you’ll make: The long-term consequences of the Clean Air Act of 1970. *Journal of Political Economy* 125(3), 848–902.
- Jbaily, A., X. Zhou, J. Liu, T.-H. Lee, L. Kamareddine, S. Verguet, and F. Dominici (2022). Air pollution exposure disparities across US population and income groups. *Nature* 601(7892), 228–233.
- Kahn, M. E. (2000). Smog reduction’s impact on California county growth. *Journal of Regional Science* 40(3), 565–582.
- Lang, C. (2015). The dynamics of house price responsiveness and locational sorting: Evidence from air quality changes. *Regional Science and Urban Economics* 52, 71–82.
- Mastromonaco, R. (2015). Do environmental right-to-know laws affect markets? Capitalization of information in the Toxic Release Inventory. *Journal of Environmental Economics and Management* 71, 54–70.
- McCormick, B. (1997). Regional unemployment and labour mobility in the UK. *European Economic Review* 41(3-5), 581–589.

- McFadden, D. (1973). Conditional logit analysis of quantitative choice behavior. *Frontiers in Econometrics*.
- Meng, J., C. Li, R. V. Martin, A. van Donkelaar, P. Hystad, and M. Brauer (2019). Estimated long-term (1981–2016) concentrations of ambient fine particulate matter across North America from chemical transport modeling, satellite remote sensing, and ground-based measurements. *Environmental Science & Technology* 53(9), 5071–5079.
- Molloy, R., C. L. Smith, and A. Wozniak (2011). Internal migration in the United States. *Journal of Economic Perspectives* 25(3), 173–96.
- Rosenthal, S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2119–2171. Elsevier.
- Saha, S. and R. D. Mohr (2013). Media attention and the toxics release inventory. *Ecological Economics* 93, 284–291.
- Silva, J. S. and S. Tenreyro (2011). Poisson: Some convergence issues. *The Stata Journal* 11(2), 207–212.

# Figures

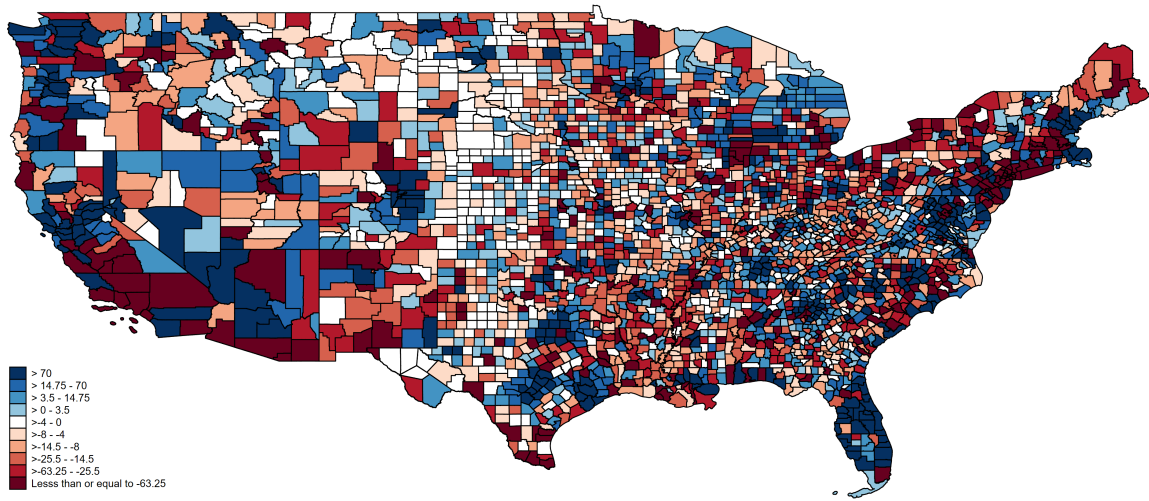


Figure 1: Net migration patterns

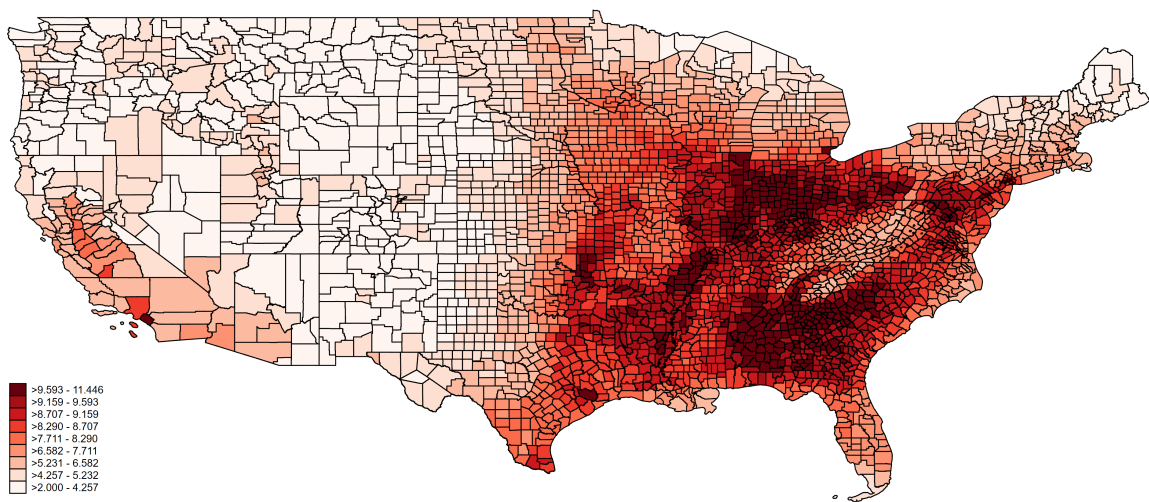


Figure 2: PM<sub>2.5</sub> patterns

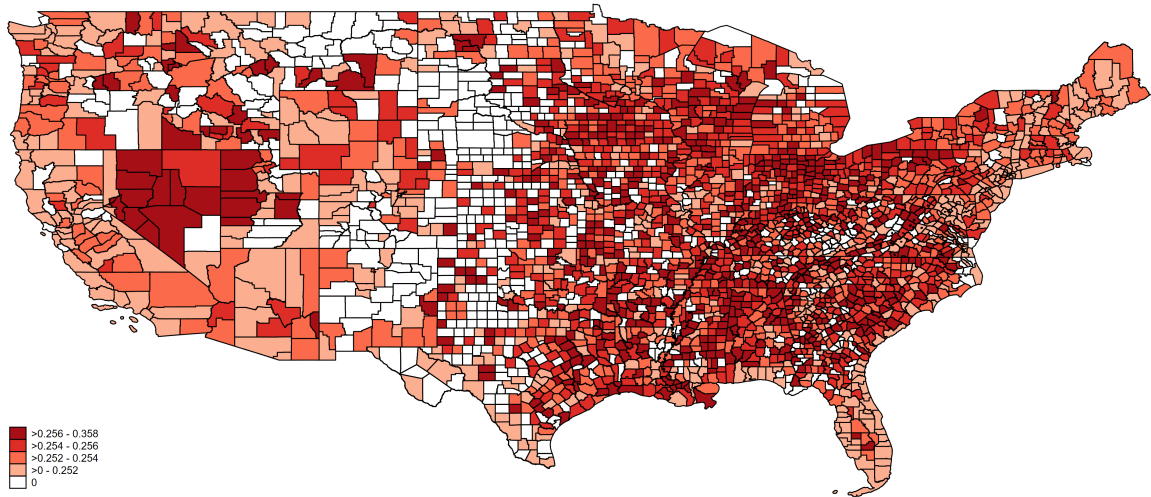


Figure 3: TRI reporting plants relative to total establishments in a county

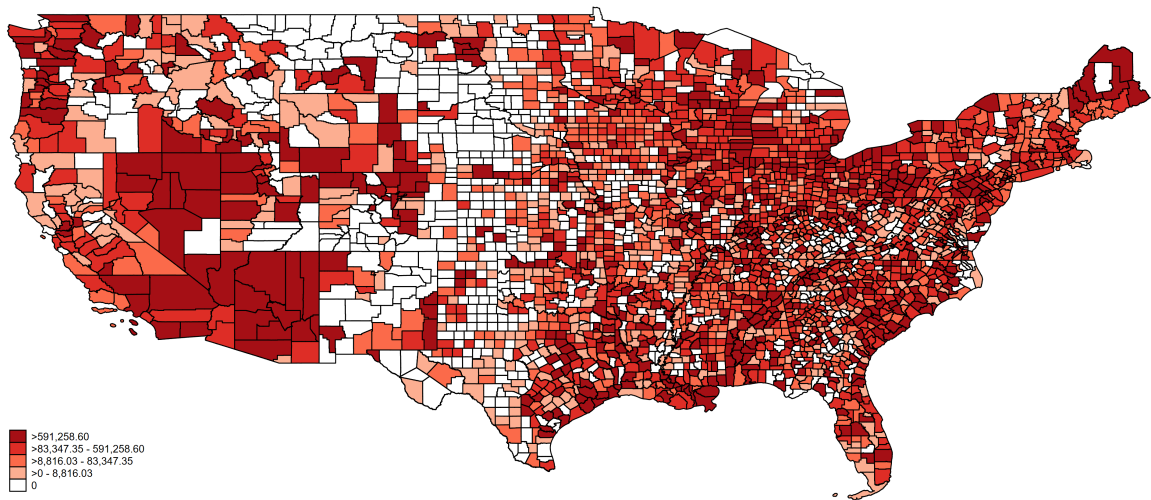


Figure 4: Toxic release patterns



# Tables

Table 1: Regression variables by County

Variable	Mean	Standard deviation
<i>Dependent Variables</i>		
Average number of households moved out	1,808.554	5,860.543
Average net migration of households	0.000	915.4711
Average relative income of out-migrants	0.717	0.235
<i>Independent Variables</i>		
Number of TRI reporters	7.037	17.613
Toxic releases (in pounds)	816,776.2	4,844,581.0
PM <sub>2.5</sub> concentration ( $\mu\text{g}/\text{m}^3$ ) <sup>a</sup>	7.635	2.271
Black population ratio	0.092	0.146
Hispanic population ratio	0.087	0.134
Household income (\$)	44,824.770	11,506.13
House price index (base year 2008)	93.381	9.574
Number of non-TRI reporting establishments	2,358.818	8,036.332
Number of manufacturing establishments	95.664	346.044
Number of amenity establishments	247.853	892.528
Population	100,303.000	320,708.600
Unemployment rate	0.085	0.031
Median age	40.753	5.164
College ratio	0.050	0.054
Metro County	0.373	0.484
Urban County	0.425	0.494
Coastal County	0.072	0.259
<i>Instrumental Variables</i>		
PM NonAttainment	0.107	0.399
Attainment County	0.090	0.286

This dataset contains 15,545 county-year observations.

<sup>a</sup>PM<sub>2.5</sub> is an ambient fine particulate matter with aerodynamic diameter less than  $2.5\mu\text{m}$ . PM<sub>2.5</sub> concentration is defined as the mass per cubic meter of air of particles with a size (diameter) generally less than 2.5 micrometers ( $\mu\text{m}$ ).

Table 2: Heterogeneity in destination choices

Variable	2010	2011	2012	2013	2014
Number of home counties	2,828	2,842	2,879	2,873	2,499
Number of destinations per home county	27.1 [2, 661]	27.9 [2, 658]	30.4 [2, 693]	31.1 [2, 719]	17.8 [2, 423]
Relative household income $_{d,h,t}$	0.70 [0.42, 1.41]	0.70 [0.42, 1.45]	0.73 [0.42, 1.86]	0.74 [0.42, 2.42]	0.71 [0.49, 1.12]
$\Delta$ PM $_{2.5d-h,t-1}$	0.22 [-1.49, 1.97]	0.23 [-1.22, 1.82]	0.18 [-1.55, 1.98]	0.19 [-1.41, 1.86]	0.17 [-0.92, 1.44]
$\Delta$ Number of TRI reporters $_{d-h,t-2}$	16.23 [-6.74, 150.43]	16.41 [-6.40, 147.69]	16.58 [-6.43, 153.40]	16.40 [-6.51, 155.71]	15.98 [-6.64, 125.14]

This table shows descriptive statistics for home counties ( $h$ ) with at least two destination counties ( $d$ ) in a given year. For those counties, we present the average and the range (in bracket) of the following variables: the number of destinations of per home county  $h$ , the relative income of out-migrant households leaving county  $h$ , the difference in PM $_{2.5}$  and the difference in the number of TRI reporters between the home and destination counties.

Table 3: Relative income of out-migrant households—all contiguous Counties in the USA

Variables	$\left(\frac{\text{Household income of out-migrants}}{\text{Household income of non-migrants}}\right)_{d,h,t}$			
	OLS		IV	
	(1)	(2)	(3)	(4)
$\Delta$ Log number of TRI reporters $_{d-h,t-2}$	-0.015*** (0.002)		-0.008*** (0.002)	
Both $d$ and $h$ counties have no TRI reporters $_{d-h,t-2}$	-0.009 (0.015)		-0.010 (0.016)	
$\Delta$ Log total releases on site $_{d-h,t-2}$		-0.002*** (0.000)		-0.002*** (0.000)
Both $d$ and $h$ counties have no releases $_{d-h,t-2}$		0.008 (0.014)		0.008 (0.014)
$\Delta$ Log of PM2.5 $_{d-h,t-1}$	-0.012 (0.008)	-0.016** (0.008)	-0.246*** (0.029)	-0.237*** (0.030)
$\Delta$ Black population ratio $_{d-h,t-1}$	-0.130*** (0.011)	-0.131*** (0.012)	0.056** (0.024)	0.047* (0.025)
$\Delta$ Hispanic population ratio $_{d-h,t-1}$	-0.063*** (0.010)	-0.062*** (0.010)	-0.082*** (0.011)	-0.087*** (0.011)
$\Delta$ Log of household income $_{d-h,t-1}$	0.216*** (0.007)	0.217*** (0.007)	0.251*** (0.008)	0.248*** (0.008)
$\Delta$ Log of house price index $_{d-h,t-1}$	-0.074*** (0.014)	-0.070*** (0.014)	0.027* (0.015)	0.026* (0.015)
$\Delta$ Log number of non-TRI reporting establishments $_{d-h,t-1}$	0.172*** (0.009)	0.182*** (0.010)	0.162*** (0.009)	0.167*** (0.010)
$\Delta$ Log number of manufacturing plants $_{d-h,t-1}$	-0.027*** (0.003)	-0.040*** (0.003)	-0.029*** (0.003)	-0.035*** (0.003)
$\Delta$ Log number of amenity establishments $_{d-h,t-1}$	-0.034*** (0.005)	-0.036*** (0.006)	-0.056*** (0.006)	-0.056*** (0.006)
$\Delta$ Log of population $_{d-h,t-1}$	-0.100*** (0.007)	-0.104*** (0.007)	-0.066*** (0.007)	-0.068*** (0.008)
$\Delta$ Unemployment rate $_{d-h,t-1}$	-0.350*** (0.070)	-0.338*** (0.070)	-0.247*** (0.070)	-0.248*** (0.070)
$\Delta$ Log of median age $_{d-h,t-1}$	0.107*** (0.012)	0.109*** (0.012)	0.123*** (0.012)	0.120*** (0.012)
$\Delta$ College ratio $_{d-h,t-1}$	-0.059* (0.031)	-0.059** (0.030)	0.055* (0.033)	0.043 (0.033)
Metro/Urban to Rural County	-0.009 (0.008)	-0.027*** (0.008)	-0.014* (0.008)	-0.031*** (0.008)
Year effects	Yes	Yes	Yes	Yes
Home county - destination county pair effects	Yes	Yes	Yes	Yes
Observations	364,944	364,944	364,944	364,944
R-squared	0.230	0.230		
Hansen $J$ -statistic (P-value)			0.615	0.617
Kleibergen-Paap Wald $F$ -statistic			972.2	872.0
First stage instruments for $\Delta$ Log of PM2.5 $_{d-h,t-1}$				
$\Delta$ PM NonAttainment $_{d-h,t-1}$			0.060*** (0.001)	0.056*** (0.001)
$\Delta$ Attainment County $_{d-h,t-1}$			-0.005*** (0.002)	-0.005*** (0.002)

Robust standard errors clustered by county pairs are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , and \*  $p < 0.1$ .

# Appendix A

Table A.1: Example of income patterns of migrating families

Home		Destination		Number of			Household income (in\$1,000)			
State	County	State	County	State (name)	County (name)	Households	Individuals	Aggregate	Average	Relative
1	1	1	1	AL	Autauga County	17,696	41,757	971,428	54.90	1.00
1	1	1	21	AL	Chilton County	78	223	2,367	30.35	0.55
1	1	1	47	AL	Dallas County	45	107	1,317	29.27	0.53
1	1	1	51	AL	Elmore County	445	1,012	17,682	39.73	0.72
1	1	1	69	AL	Houston County	12	19	477	39.75	0.72
1	1	1	73	AL	Jefferson County	73	110	1,773	24.29	0.44
1	1	1	85	AL	Lowndes County	10	27	336	33.60	0.61
1	1	1	101	AL	Montgomery County	377	739	12,591	33.40	0.61
1	1	1	109	AL	Pike County	10	15	279	27.90	0.51
1	1	1	113	AL	Russell County	11	16	237	21.55	0.39
1	1	8	41	CO	El Paso County	14	46	938	67.00	1.22
1	1	12	5	FL	Bay County	11	25	411	37.36	0.68
1	1	12	91	FL	Okaloosa County	15	48	1,039	69.27	1.26
1	1	13	63	GA	Clayton County	10	13	77	7.70	0.14
1	1	13	121	GA	Fulton County	12	15	299	24.92	0.45
1	1	17	169	IL	St. Clair County	11	47	915	83.18	1.52
1	1	47	157	TN	Shelby County	10	24	1,192	119.20	2.17
1	1	48	29	TX	Bexar County	16	36	677	42.31	0.77
1	1	51	59	VA	Fairfax County	15	59	1,491	99.40	1.81
1	1	57	5	FR	Foreign - APO/FPO ZIPs	43	118	2,785	64.77	1.18

Table A.2: Example of TRI reporting plants and toxic releases by year

Year	TRI plant ID	Total release	State	County	NAICS code	NAICS title
2008	3606WSHRMN74CUN	3.37	AL	Autauga County	327320	Ready-mix concrete manufacturing
2009	3606WSHRMN74CUN	0.15	AL	Autauga County	327320	Ready-mix concrete manufacturing
2010	3606WSHRMN74CUN	0.141	AL	Autauga County	327320	Ready-mix concrete manufacturing
2011	3606WSHRMN74CUN	0.155	AL	Autauga County	327320	Ready-mix concrete manufacturing
2012	3606WSHRMN74CUN	0.114	AL	Autauga County	327320	Ready-mix concrete manufacturing
2008	36067NNCAMP100JE	1,540,481	AL	Autauga County	322130	Paperboard mills
2009	36067NNCAMP100JE	2,888,154	AL	Autauga County	322130	Paperboard mills
2010	36067NNCAMP100JE	2,983,310	AL	Autauga County	322130	Paperboard mills
2011	36067NNCAMP100JE	2,648,291	AL	Autauga County	322130	Paperboard mills
2012	36067NNCAMP100JE	3,084,997	AL	Autauga County	322130	Paperboard mills
2008	36066TWBCL17JES	0.53	AL	Autauga County	332999	All other miscellaneous fabricated metal product
2009	36066TWBCL17JES	0.4	AL	Autauga County	332999	All other miscellaneous fabricated metal product
2010	36066TWBCL17JES	0	AL	Autauga County	332999	All other miscellaneous fabricated metal product
2011	36066TWBCL17JES	0	AL	Autauga County	332999	All other miscellaneous fabricated metal product
2012	36066TWBCL17JES	2.75	AL	Autauga County	332999	All other miscellaneous fabricated metal product

Table A.3: Regression variables for relative income estimation

Variable	Mean	Standard deviation
<i>Dependent Variables</i>		
Relative household income of migrants $_{d,h,t}$	0.761	0.765
Relative personal income of migrants $_{d,h,t}$	0.909	0.934
<i>Independent Variables</i>		
$\Delta$ Number of TRI reporters $_{d-h,t-2}$	0.170	88.111
Both $d$ and $h$ counties have no TRI reporters $_{d-h,t-2}$	0.005	0.067
$\Delta$ Log total releases on site $_{d-h,t-2}$	25,011.50	1.36e+07
Both $d$ and $h$ counties have no releases $_{d-h,t-2}$	0.008	0.088
$\Delta$ PM2.5 $_{d-h,t-1}$	-0.021	2.076
$\Delta$ Black population ratio $_{d-h,t-1}$	0.0003	0.154
$\Delta$ Hispanic population ratio $_{d-h,t-1}$	0.002	0.171
$\Delta$ Household income $_{d-h,t-1}$	12.481	17,528.89
$\Delta$ House price index $_{d-h,t-1}$	-0.097	11.731
$\Delta$ Number of non-TRI establishments $_{d-h,t-1}$	248.295	42,037.71
$\Delta$ Number of manufacturing establishments $_{d-h,t-1}$	0.548	1,937.02
$\Delta$ Number of amenity establishments $_{d-h,t-1}$	24.661	4,868.752
$\Delta$ Population $_{d-h,t-1}$	9,367.465	1,694,359
$\Delta$ Unemployment rate $_{d-h,t-1}$	-0.0004	0.027
$\Delta$ Median age $_{d-h,t-1}$	0.024	5.943
$\Delta$ College ratio $_{d-h,t-1}$	0.00001	0.054
Metro/Urban to Rural County	0.018	0.133
<i>Instrumental Variables</i>		
$\Delta$ PM NonAttainment $_{d-h,t-1}$	-0.007	0.875
$\Delta$ Attainment County $_{d-h,t-1}$	0.383	0.486

These summary statistics are based on the data sample between years 2010-2014 used to estimate equation (1). The sample has 364,944 observations.  $\Delta x_{d-h,t} = x_{d,t} - x_{h,t}$  is the difference between the value of the local attribute  $x$  at destination  $d$  and its value in the home county  $h$  in year  $t$ .

Table A.4: Within State and Out of State

Panel A: Relative income of out-migrant households			
Variables	$\left(\frac{\text{Household income of out-migrants}}{\text{Household income of non-migrants}}\right)_{d,h,t}$		
	US	Within State	Out of State
	(1)	(2)	(3)
$\Delta$ Log number of TRI reporters $_{d-h,t-2}$	-0.008*** (0.002)	-0.010*** (0.002)	-0.004 (0.003)
Both $d$ and $h$ counties have no TRI reporters $_{d-h,t-2}$	-0.010 (0.016)	-0.001 (0.015)	-0.194** (0.092)
$\Delta$ Log of PM2.5 $_{d-h,t-1}$	-0.246*** (0.029)	-0.084** (0.037)	-0.198*** (0.031)
Year effects	Yes	Yes	Yes
Home - destination county effects	Yes	Yes	Yes
Observations	364,944	154,321	210,623
Hansen $J$ -statistic (P-value)	0.615	0.238	0.185
Kleibergen-Paap Wald $F$ -statistic	972.2	258.9	934.6
First stage instruments for $\Delta$ Log of PM2.5 $_{d-h,t-1}$			
$\Delta$ PM NonAttainment $_{d-h,t-1}$	0.060*** (0.001)	0.048*** (0.002)	0.071*** (0.002)
$\Delta$ Attainment County $_{d-h,t-1}$	-0.005*** (0.002)	-0.001 (0.001)	-0.015*** (0.004)
Panel B: Relative personal income of out-migrants			
Variables	$\left(\frac{\text{Personal income of out-migrants}}{\text{Personal income of non-migrants}}\right)_{d,h,t}$		
	US	Within State	Out of State
	(1)	(2)	(3)
$\Delta$ Log number of TRI reporters $_{d-h,t-2}$	-0.026*** (0.002)	-0.020*** (0.002)	-0.024*** (0.004)
Both $d$ and $h$ counties have no TRI reporters $_{d-h,t-2}$	-0.008 (0.017)	0.002 (0.016)	-0.262** (0.123)
$\Delta$ Log of PM2.5 $_{d-h,t-1}$	-0.238*** (0.034)	-0.036 (0.048)	-0.199*** (0.037)
Year effects	Yes	Yes	Yes
Home - destination county effects	Yes	Yes	Yes
Observations	364,922	154,308	210,614
Hansen $J$ -statistic (P-value)	0.411	0.582	0.168
Kleibergen-Paap Wald $F$ -statistic	972.3	258.9	934.7
First stage instruments for $\Delta$ Log of PM2.5 $_{d-h,t-1}$			
$\Delta$ PM NonAttainment $_{d-h,t-1}$	0.060*** (0.001)	0.048*** (0.002)	0.071*** (0.002)
$\Delta$ Attainment County $_{d-h,t-1}$	-0.005*** (0.002)	-0.001 (0.001)	-0.015*** (0.004)

Robust standard errors clustered by county pairs are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , and \*  $p < 0.1$ . In all specifications we control for Black and Hispanic population ratios, median household income, house price index, population, unemployment rate, number of non-TRI establishments, number of manufacturing establishments, number of amenity establishments, median age, college ratio and Metro/Urban to Rural county dummy.



Table A.5: Relative income of out-migrant households-MSA and Urban Counties

Variables	$\left( \frac{\text{Household income of out-migrants}}{\text{Household income of non-migrants}} \right)_{d,h,t}$					
	Within		Urban/MSA		County type	
	MSA	Urban and MSA	to all	Small	Low density	High density
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta$ Log number of TRI reporters $_{d-h,t-2}$	-0.004* (0.003)	-0.008*** (0.002)	-0.008*** (0.002)	-0.001 (0.002)	-0.006** (0.003)	0.005 (0.003)
Both $d$ and $h$ counties have no TRI reporters $_{d-h,t-2}$	-0.003 (0.022)	-0.018 (0.016)	-0.020 (0.014)	-0.008 (0.014)	-0.012 (0.017)	-0.037 (0.025)
$\Delta$ Log of PM2.5 $_{d-h,t-1}$	-0.205*** (0.026)	-0.237*** (0.029)	-0.237*** (0.029)	-0.271*** (0.028)	-0.244*** (0.077)	-0.166*** (0.026)
Home - destination county effects	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	260,241	350,619	356,987	287,230	105,440	119,807
Hansen $J$ -statistic (P-value)	0.229	0.474	0.548	0.830	0.100	0.183
Kleibergen-Paap Wald $F$ -statistic	988.2	967.4	967.7	971.5	36.4	1,238.0
First stage instruments for $\Delta$ Log of PM2.5 $_{d-h,t-1}$						
$\Delta$ PM NonAttainment $_{d-h,t-1}$	0.066*** (0.002)	0.060*** (0.001)	0.060*** (0.001)	0.062*** (0.001)	0.029*** (0.003)	0.091*** (0.002)
$\Delta$ Attainment County $_{d-h,t-1}$	-0.008*** (0.002)	-0.005*** (0.002)	-0.005*** (0.002)	-0.004*** (0.001)	0.003 (0.002)	-0.012*** (0.004)

Robust standard errors clustered by county pairs are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , and \*  $p < 0.1$ . In all specifications we control for Black and Hispanic population ratios, median household income, house price index, population, unemployment rate, number of non-TRI establishments, number of manufacturing establishments, number of amenity establishments, median age, and college ratio.

## Appendix B: Households location decisions

In this section, we investigate how US households choose where to move given the trade-offs between moving costs, job opportunities, and environmental and other amenities.

### Net migration patterns

The structure of the IRS data allows us to calculate the net migration for a given county for a given year. Net migration in a county in a given year is defined as the difference between the total number of households that move in that county from all the other counties (in-migration) and the total number of households that move out (out-migration). A snapshot of the dataset used to estimate the net migration model is presented in Table B.1. If we consider the same example as in the main text (Autauga County in Alabama), we obtain that, for 2013, net migration is positive while, for 2014, it is negative.

We explore how the year-to-year changes in net migration covary with lagged pollution measures and demographic characteristics in the home county  $h$ :

$$\ln \left( \frac{in - migration}{out - migration} \right)_{h,t} = \beta \mathbf{E}_{h,t-1} + \mathbf{X}'_{h,t-1} \gamma + \alpha_h + \tau_t + \epsilon_{h,t} \quad (3)$$

We denote  $\mathbf{E}$  as the vector containing our two pollution measures at the county level.  $\mathbf{X}$  are county demographics.  $\alpha$  and  $\tau$  are county/state and time fixed effects.  $\epsilon$  is the error term. We estimate this empirical model using a simple linear regression, where *PM NonAttainment* and *Attainment County* are used to instrument  $PM_{2.5}$  concentration levels in the home county. We have 3,109 counties over the period 2010-2014.

We present our county-level net migration results in Table B.2. In columns 1 and 3, we use the log number of TRI reporters, while in column 2 and 4, we use the log of total toxic releases (in pounds). Columns 1 and 2 include state and time fixed effects,

while columns 3 and 4 show the estimation results with county and time fixed effects.

In column (1), the coefficients associated with the number of TRI reporters and PM<sub>2.5</sub> concentration in a county are negative and statistically significant. Once we allow for county fixed effects, the coefficients of all our pollution measures become insignificant. However, the IV estimates with county fixed effects have to be interpreted with caution as the first-stage  $F$ -statistics is significantly lower than the ones in columns (1) and (2), with state fixed effects. Finally, a higher unemployment rate in county  $h$  is associated with a lower net migration in county  $h$ .

Using net migration patterns and changes in local environmental quality does not allow us to identify a sorting by income across US counties. Net migration can be zero or minimal if a similar number of people move into and out of a county, but the households moving into or out of the county may have very different characteristics and come from counties with very different levels of pollution (see Depro et al. 2015). To address this issue, we formulate a destination choice model that exploits information regarding both the origin and destination counties, enabling us to explore internal migration patterns within the US.

## Destination choice model

We consider the destination choice of a household  $i$ , located in county  $h$ , who has to decide where to move among  $N$  possible destination counties. In our model, we assume that a household migration decision is based on the comparison of home and destination attributes (defined in section 3 of the paper). We express the indirect utility from destination  $d$  for household  $i$  as follows:

$$V_{i,d,h,t} = \Delta E_{d-h,t} + \Delta X_{d-h,t} + D_{dh} + \epsilon_{i,d,h,t}$$

where  $\Delta E_{d-h,t}$  contains our measures of environmental quality, while  $\Delta X_{d-h,t}$  includes other characteristics of county  $d$  (relative to county  $h$ ) that might attract migrants (e.g. differences in unemployment rates, wages, number of amenity establishments,

house price index).  $D_{dh}$  includes the proxy for moving costs. The disturbance term  $\epsilon_{i,d,h,t}$  is independent and identically distributed.

In a utility-maximization framework, household  $i$  will choose to move to county  $d$  at time  $t$  if their indirect utility associated with moving to county  $d$  is larger than the indirect utility they could obtain by moving to any other county  $k$ , i.e.  $V_{i,d,h,t} \geq V_{i,k,h,t}$  for all  $k \neq d$ , and  $d, k \in \{1, \dots, N\}$ . In order to have closed form expressions for a household's choice probabilities, we assume that  $\epsilon_{i,d,t}$  follows a Type 1 extreme value distribution. We also assume that each household knows their private costs and expected utility. This asymmetric information assumption enables us to convert the discrete actions of households into continuous location choice probabilities. Using the results of McFadden (1973), we model household  $i$ 's choice of destination location ( $d$  - county) using a conditional Logit model:

$$\begin{aligned} \Pr(V_{i,d,h,t} = \max_k V_{i,k,h,t}) &= \Pr(m_{i,d,h,t} = 1 | \Delta E_{d-h,t}, \Delta X_{d-h,t}, D_{dh}) \\ &= \frac{\exp[\Delta E_{d-h,t} + \Delta X_{d-h,t} + D_{dh}]}{\sum_{k=1}^N \exp[\Delta E_{k-h,t} + \Delta X_{k-h,t} + D_{kh}]} \end{aligned}$$

where  $m_{i,d,h,t}$  equals 1 if a household  $i$ , living in county  $h$  chooses location  $d$  and 0 otherwise. By modeling a household migration choice as a function of the difference between home and destination characteristics, we assume that households moving to the same destination county but from different home counties are treated differently. In that respect, the home county characteristics can be interpreted as the average characteristics of households from county  $h$ .

Our summary statistics in Table 1 indicate that about 1,800 households move from one county to another each year. Given that each household has 3,108 different counties to choose from, we have, on average, about 5.6 million observations for a given county in a given year. As we consider the 3,109 counties in the lower 48 states, the number of observations is about 17.4 billion for a given year or 87 billion for the whole sample period (five years). With the large number of observations in this dataset, estimating the location choice model using the conditional Logit technique becomes

computationally intensive and challenging. We overcome this issue by adopting an approach similar to the one used by Rosenthal and Strange (2004), De Silva et al. (2016) and De Silva et al. (2021) when studying firm location decisions. Due to the volume of data, they aggregate firm entry by location to obtain the number of firms entering each area in a given year, and they estimate their models using a Poisson pseudo-maximum likelihood (PPML) estimator, with year fixed effects. De Silva et al. (2016) also show that, when studying entry decisions of firms reporting releases to the TRI, the results (in terms of sign and significance of the coefficients) of the conditional Logit and PPML estimations are very similar.

We therefore aggregate household movements from a home county to a destination county. For each of the 3,109 counties in a given year, there will be 3,108 choices of destination counties. This gives us 9,662,772 ( $3,109 \times 3,108$ ) county-to-county observations per year. Since we have five years, our aggregated sample size is 48,313,860 observations. Let  $H_{dht}$  denote the total number of households moving from county  $h$  to county  $d$  at time  $t$ . We estimate the following specification using the PPML method:<sup>24</sup>

$$E[H_{dh,t} | \Delta E_{d-h,t}, \Delta X_{d-h,t}, D_{dh}] = \exp[\beta_1 \Delta E_{d-h,t} + \beta_2 \Delta X_{d-h,t} + \beta_3 D_{dh} + \tau_t] \quad (4)$$

where  $\tau_t$  is a time effect.  $D_{dh}$  includes the proxy for moving costs. We use *geographical distance*, as a proxy for the costs associated with moving from one county to another. We define geographical distance between two counties as the distance between the home and destination counties centroids. As moving to a different state is associated with additional costs (e.g. transferring a vehicle insurance and registration documentation, applying for a new driving license...), we include a dummy variable *Out of State* that takes the value 1 for interstate migration.

---

<sup>24</sup>The PPML estimator corrects for both under- and over-dispersion. Gouriéroux et al. (1984) showed that PPML is consistent and an asymptotically normal estimator can be obtained without specifying the probability density function of disturbances representing a specification error in the parameter of the Poisson distribution. Further, Silva and Teneyro (2011) showed that, even if the conditional variance is not proportional to the conditional mean, the PPML will still be consistent and is generally well-behaved when the proportion of zeros is large.

In this estimation, we maintain the integrity of the original conditional Logit setup. Initially, our interest lies in analyzing the cross-sectional variation of location choices within a given year, while accounting for household unobservable heterogeneity through household fixed effects. Given that households tend to move infrequently, each household is likely to feature only once during our five-year period. Consequently, a conditional Logit model is estimated by incorporating fixed effects by year exclusively. Similarly, in our final PPML specification, we introduce time fixed effects to capture cross-sectional variations across destination counties.

Note that we are unable to use our IV approach or incorporate county-pair fixed effects into equation (4). Despite having 48 million observations, the variable  $H_{dh,t}$  is strictly positive for only 0.79% of our dataset, resulting in estimation issues.<sup>25</sup> With only five observations per county pair and a substantial number of county pairs exhibiting zeros over the five-year period, there is insufficient temporal variation within county pairs.

As our models do not allow for county fixed effects, we are able to include an additional measure of environmental quality: the number of listed Superfund sites within a county in a given year.<sup>26</sup> We also account for Metro, urban, and rural moving patterns using dummies to identify households moving within Metro counties, urban to Metro counties, Metro to urban counties, rural to Metro counties, Metro to rural counties, urban to rural counties. Further we control for moving patterns among coastal and non-coastal counties using dummies to identify households moving from coastal to coastal counties, to a coastal county from a non-coastal county, and to a non-coastal county from a coastal county.

Table B.3 summarizes our results for the destination choice model estimated for all contiguous US states. Specifications differ in terms of the pollution measures in-

---

<sup>25</sup>99.21% of our 48,313,860 observations are zeros.

<sup>26</sup>We use the number of Superfund sites on the National Priority List (NPL), provided by the EPA. The NPL is the list of sites of national priority among the known releases or threatened releases of hazardous substances, pollutants, or contaminants throughout the United States and its territories. As for the TRI-related measures, we include a dummy when both counties do not have any Superfund sites.

cluded in the regressions. The coefficients of two of our pollution measures (i.e.  $PM_{2.5}$  and Superfund sites) have the expected sign and are statistically significant. When the concentration of  $PM_{2.5}$  or the number of Superfund sites at destination increases, compared to the home levels, less households will migrate to that county. The parameter estimated in column (3) shows that a 1% increase in the difference of  $PM_{2.5}$  concentrations between home and destination counties is associated with an 11% decrease in the number of households moving between these home and destination counties. The coefficient of our TRI-related pollution measures are not significant, but results in columns (1), (3) and (4) show that households are about twice less likely to move between counties that do not have any TRI reporters (or toxic releases), than between county pairs where one of the counties has at least one TRI reporting facility.

Households tend to avoid counties with higher unemployment rates. In addition, counties with more manufacturing establishments (relative to the home county) appear to be less attractive. Among the amenity variables included in the model, the coefficient of the house price index is statistically significant (at 10% level): households are more likely to move to places where house prices have increased less relative to the base year. In terms of demographic characteristics, our estimates reveal that if the proportion of individuals with a college degree in the home county increases compared to the proportion in the destination county (i.e.  $\Delta \text{College ratio}_{d-h,t-1}$  decreases), we observe more households moving from county  $h$  to county  $d$ . This is consistent with prior evidence that more-educated individuals are more likely to move.

Finally, moving costs are strongly associated with household decisions to move to a particular destination county. Households do not seem to move out of state and, they don't move too far as the coefficient of the distance variable is negative.<sup>27</sup>

---

<sup>27</sup>We re-estimate our models with different subsets of control variables and the main results (available upon request) are qualitatively the same.

Table B.1: Example of TRI reporting plants, toxic releases, and demographics by County and year

Year	State	County	Net migration	TRI		Minority ratio	PM2.5	Household income	Population	Unemployment	Metro	Urban	Coastal
				Reporters	Release								
2013	1	1	117	3	3,084,996.9	0.215	9.043	51,868.00	54,792	0.062	1	0	0
2014	1	1	-63	2	3,234,673.9	0.218	8.419	54,366.00	54,977	0.059	1	0	0
2013	1	3	1,141	9	19,359.2	0.126	9.043	47,539.00	195,147	0.066	1	0	1
2014	1	3	499	10	15,246.0	0.128	8.419	49,626.00	199,745	0.061	1	0	1
2013	1	5	-47	6	269,664.3	0.500	9.043	30,981.00	26,938	0.102	0	1	0
2014	1	5	-26	7	164,940.9	0.500	8.419	34,971.00	26,763	0.106	0	1	0



Table B.2: Regression results for net migration in contiguous counties

Variables	$Log \left( \frac{in-migration}{out-migration} \right)_{h,t}$			
	(1)	(2)	(3)	(4)
Log number of TRI reporters $_{h,t-2}$	-0.019** (0.008)		0.009 (0.030)	
Log of total toxic release $_{h,t-2}$		-0.001 (0.001)		0.002 (0.004)
Log of PM <sub>2.5</sub> $_{h,t-1}$	-0.405** (0.168)	-0.434** (0.172)	-0.077 (0.709)	-0.084 (0.703)
Black ratio $_{h,t-1}$	-0.086 (0.066)	-0.090 (0.067)	2.289 (2.370)	2.290 (2.369)
Hispanic ratio $_{h,t-1}$	-0.200*** (0.071)	-0.210*** (0.072)	-1.668 (2.693)	-1.659 (2.686)
Log of household income $_{h,t-1}$	0.216*** (0.043)	0.219*** (0.043)	-0.256 (0.157)	-0.256 (0.157)
Log of house price index $_{h,t-1}$	-0.120* (0.069)	-0.119* (0.070)	0.020 (0.212)	0.022 (0.213)
Log number of non-TRI reporting establishments $_{h,t-1}$	-0.186*** (0.042)	-0.184*** (0.042)	0.047 (0.223)	0.048 (0.223)
Log number of manufacturing plants $_{h,t-1}$	0.031* (0.018)	0.024 (0.017)	-0.050 (0.075)	-0.051 (0.075)
Log number of amenity establishments $_{h,t-1}$	0.029 (0.026)	0.026 (0.026)	0.093 (0.095)	0.093 (0.095)
Log of population $_{h,t-1}$	0.191*** (0.035)	0.191*** (0.035)	0.403 (0.399)	0.408 (0.400)
Unemployment rate $_{h,t-1}$	-0.354 (0.340)	-0.305 (0.342)	-1.491*** (0.574)	-1.491*** (0.574)
Log of median age $_{h,t-1}$	0.214*** (0.062)	0.218*** (0.063)	-1.229 (0.763)	-1.225 (0.764)
College ratio	0.068 (0.142)	0.052 (0.142)	0.106 (0.205)	0.106 (0.205)
Year effects	Yes	Yes	Yes	Yes
State effects	Yes	Yes		
County effects			Yes	Yes
Observations	15,545	15,545	15,545	15,545
Hansen $J$ -statistic (P-value)	0.257	0.217	0.378	0.376
Kleibergen-Paap Wald $F$ -statistic	18.73	19.05	7.034	7.171

Robust standard errors clustered by home county are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$   
Log of PM<sub>2.5</sub> is instrumented using *PM NonAttainment* and *Attainment County*.  
All columns include indicators for Metro, Urban, and coastal county effects.

Table B.3: Location choice results for out migration patterns—all contiguous counties

Variables	Number of households moving from $h$ to $d$			
	(1)	(2)	(3)	(4)
$\Delta$ Log number of TRI reporters $_{d-h,t-2}$	0.005 (0.007)		0.006 (0.007)	
Both $d$ and $h$ counties have no TRI reporters $_{d-h,t-2}$	-2.128*** (0.042)		-2.128*** (0.042)	
$\Delta$ Log total releases on site $_{d-h,t-2}$				0.001 (0.001)
Both $d$ and $h$ counties have no releases $_{d-h,t-2}$				-1.889*** (0.029)
$\Delta$ Log of PM2.5 $_{d-h,t-1}$		-0.113** (0.053)	-0.114** (0.053)	-0.113** (0.053)
$\Delta$ Log number of superfunds $_{d-h,t-1}$	-0.030*** (0.010)	-0.030*** (0.010)	-0.032*** (0.010)	-0.031*** (0.010)
Both $d$ and $h$ counties have no superfunds $_{d-h,t-1}$	-1.282*** (0.017)	-1.321*** (0.017)	-1.282*** (0.017)	-1.268*** (0.017)
$\Delta$ Black ratio $_{d-h,t-1}$	-0.077** (0.037)	-0.045 (0.040)	-0.047 (0.040)	-0.046 (0.040)
$\Delta$ Hispanic ratio $_{d-h,t-1}$	-0.026 (0.102)	-0.040 (0.101)	-0.038 (0.102)	-0.037 (0.102)
$\Delta$ Log of household income $_{d-h,t-1}$	0.023 (0.032)	0.033 (0.032)	0.033 (0.032)	0.032 (0.032)
$\Delta$ Log of house price $_{d-h,t-1}$	-0.155** (0.076)	-0.137* (0.073)	-0.139* (0.073)	-0.140* (0.074)
$\Delta$ Log number of non-TRI reporting establishments $_{d-h,t-1}$	0.037 (0.035)	0.038 (0.034)	0.044 (0.034)	0.042 (0.035)
$\Delta$ Log number of manufacturing plants $_{d-h,t-1}$	-0.031** (0.013)	-0.026** (0.011)	-0.031** (0.013)	-0.027** (0.012)
$\Delta$ Log number of amenity establishments $_{d-h,t-1}$	-0.019 (0.019)	-0.026 (0.019)	-0.028 (0.020)	-0.029 (0.020)
$\Delta$ Log of population $_{d-h,t-1}$	0.019 (0.024)	0.026 (0.025)	0.022 (0.025)	0.023 (0.025)
$\Delta$ Unemployment rate $_{d-h,t-1}$	-0.626** (0.257)	-0.478* (0.274)	-0.473* (0.275)	-0.483* (0.275)
$\Delta$ Median age $_{d-h,t-1}$	0.045 (0.043)	0.027 (0.044)	0.029 (0.045)	0.029 (0.044)
$\Delta$ College ratio $_{d-h,t-1}$	-0.395*** (0.139)	-0.364*** (0.137)	-0.372*** (0.140)	-0.370*** (0.140)
Log distance between counties $d$ and $h$	-1.359*** (0.009)	-1.359*** (0.009)	-1.359*** (0.009)	-1.358*** (0.009)
Out of state	-1.699*** (0.018)	-1.694*** (0.018)	-1.699*** (0.018)	-1.701*** (0.018)
MSA, urban, and rural controls	Yes	Yes	Yes	Yes
Coastal and non-coastal county controls	Yes	Yes	Yes	Yes
Time effects	Yes	Yes	Yes	Yes
Observations	48,313,860	48,313,860	48,313,860	48,313,860
Loglikelihood	-6.240e+07	-6.260e+07	-6.240e+07	-6.230e+07

Robust standard errors are in parentheses, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

We also account for Metro, urban, and rural moving patterns using dummies to identify moving within Metro counties, urban to Metro counties, Metro to urban counties, rural to Metro counties, Metro to rural counties, urban to rural counties. Further we control for moving patterns among coastal and non-coastal counties using dummies to identify household moving from coastal to coastal counties, to a coastal county from a non-coastal county, and to a non-coastal county from a coastal county.