DISCUSSION PAPER

MARCH 2006 ■ RFF DP 06-10

Do People Vote with Their Feet?

An Empirical Test of Environmental Gentrification

H. Spencer Banzhaf and Randall P. Walsh

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

Do People Vote with Their Feet?
An Empirical Test of Environmental Gentrification

H. Spencer Banzhaf and Randall P. Walsh

Abstract

Tiebout's (1956) suggestion that people "vote with their feet" to find the community that provides their optimal bundle of taxes and public goods has played a central role in the theory of local public finance over the past 50 years. Given the central importance of Tiebout's insights, there have been surprisingly few direct tests of his premise. In this paper, we use a Tiebout equilibrium model to derive testable hypotheses about changes in local community demographics. The model clearly predicts increased population density in neighborhoods that experience an exogenous increase in public goods but yields only tentative predictions about the effect on neighborhood composition. To test these hypotheses, we use a difference-in-difference model to identify the effect of initial pollution levels and changes in local pollution on population and demographic composition.

Our results provide strong empirical support for the notion that households "vote with their feet" in response to changes in environmental quality. This result has two implications. First, and most broadly, it provides direct empirical support for the assumptions underlying the Tiebout model. Second, in our particular application, the potential for what we call "environmental gentrification" has important implications both for the analysis of environmental equity and for the design of environmental policies aimed at benefiting the less-advantaged elements of society.

Key Words: Tiebout, gentrification, air quality

JEL Classification Numbers: J6, Q5, R2

Contents

Do People Vote with Their Feet? An Empirical Test of Environmental Gentrification

H. Spencer Banzhaf and Randall P. Walsh*

I. Introduction

Tiebout's (1956) suggestion that people "vote with their feet" to find the community that provides their optimal bundle of taxes and public goods has played a central role in the theory of local public finance over the past 50 years, motivating such diverse literatures as capitalization and "hedonics," fiscal federalism, and the formation of endogenous public goods. More recently, a new and growing empirical literature has leveraged the equilibrium properties of the Tiebout model to identify general equilibrium models of household sorting (e.g. Bayer et al. 2005, Ferreyra 2005; Sieg et al. 2004; Timmins 2003).

Given the central importance of Tiebout's insights, there have been surprisingly few direct tests of his premise. Existing tests of the Tiebout model can be grouped into two broad categories.[1] Indirect or implicit tests, the most common, have focused on deductive implications of the model. For example, Wallace Oates's (1969) seminal article on the link between local tax and service packages and property values introduced a hedonic model as an implicit test of a Tiebout equilibrium. Brueckner (1982) tested implications of the model related to the efficient provision of public goods. In a recent paper reflecting on the impact of Tiebout's model today, Oates (2005) highlights the fact that many tests have focused on issues of stratification in demand for public goods and the link between diversity across communities in income and public good provision (e.g., Gramlich and Rubinfeld 1982; Epple and Sieg 1999).

---

* Banzhaf is Fellow, Resources for the Future, and Walsh is Assistant Professor, Department of Economics and Institute of Behavioral Studies, University of Colorado. Support for this research was provided by the National Science Foundation, NSF SES-03-21566. We thank Lisa Crooks and Joshua Sidon for excellent research assistance. We also appreciate the helpful comments of Patrick Bayer, Lori Bennear, Trudy Cameron, Nicholas Flores, Wayne Gray, Michael Greenstone, Michael Hanemann, Erik Johnson, Michael Margolis, Craig McIntosh, Kenneth McConnell, Wallace Oates, Kate Probst, David Simpson, Kerry Smith, Christopher Timmins, Kip Viscusi, Roberton Williams, Ann Wolverton, and participants of the 2004 AERE Workshop, the 2004 CU Environmental and Resource Economics Workshop, the 2005 ASSA meetings, the 2005 NBER Summer Institute, the 2005 Occasional California Workshop on Environmental and Natural Resource Economics, and seminar participants at Duke University, George State University, Rice University, the University of Maryland School of Public Policy, and the University of Wisconsin School of Business.

[1] For reviews and discussion, see Rubinfeld (1987) and Oates (2005).

Direct tests of actual migratory responses to public good provision—Tiebout's mechanism of people voting with their feet—have been less common. Graves and Waldman (1991) found that the elderly retire in counties where public goods are capitalized more into wages than into land prices. Kahn (2000) found migration into California counties with improving air quality. These tests have identified county differences on a national or regional scale. Yet given constraints on mobility related to family, career, and other networks, residential responses to changes to public goods are most likely to occur within, rather than across, metropolitan areas.

Finally, a small number of papers in the "environmental justice" literature focus on migratory responses associated with changes in environmental quality. This literature has consistently found a correlation between various pollutants and poor or minority populations (e.g. United Church of Christ 1987; Goldman and Fitton 1994, Mohai and Bryant 1992), as well as a measure of political organization (Hamilton 1993). Most policy responses to this finding proceed on the premise that it arises from discrimination on the part of firms.[2] However, this correlation could also arise from Tiebout sorting on the part of households. In exploring this hypothesis, Been (1997) and Wolverton (2002) did not find evidence of a correlation at the time of the siting of the polluting facility. Inasmuch as this correlation does occur in contemporary data, this finding is consistent with ex post Tiebout sorting.

Been (1997) also provides the first attempt to evaluate directly the neighborhood composition effects from changes in pollution exposure, studying 544 "communities" that hosted active commercial hazardous waste treatment storage and disposal facilities. She finds little direct evidence that the siting of facilities led to changes in demographics.

Been's insights represent a significant advance in the environmental justice literature. Nevertheless, four specific concerns raise doubts about the inference that can be drawn from her analysis of sorting after the siting of the facility. First, she uses only a very sparse set of econometric controls, omitting exiting facilities, lagged effects of older facilities, and demographic variables other than baseline racial composition. Second, her use of all national nonhost Census tracts as controls is potentially problematic. If subregions of the country that host facilities are experiencing demographic shifts that vary systematically from those of

[2] For example, concerns about "environmental racism" have prompted the U.S. Environmental Protection Agency to create a national Environmental Justice Office and several states to craft their own legislative responses.

subregions that do not host facilities, the use of all other tracts as controls could lead these systematically varying demographic shifts to mask environmentally driven composition changes that are occurring within the host subregions.[3] Third, the study uses the presence of a hazardous waste facility only, without regard to the heterogeneity in the quantity and toxicity of wastes or differences across tract-facility pairs in the average distance that tract households are located from the facility.[4] Finally, she uses the host Census tract as the affected "community." As Been points out, however, it is not clear that the Census tract is the appropriate definition. While relatively homogeneous in population, Census tracts can vary significantly in size and are often quite large, making it possible for their use as a "community" definition to mask demographic shifts that occur within each tract. Additionally, because facilities are often located along Census tract boundaries, attaching facility effects to their host tract introduces significant noise to the empirical model.

Several of these issues are addressed in more recent work by Cameron and Crawford (2004). They evaluate how demographic composition, measured at the Census block level, varies across time and distance from six Superfund sites—finding heterogeneous demographic shifts across the six facilities. This approach yields much smaller neighborhood definitions, uses neighboring Census blocks located farther from the facility as controls, and directly accounts for variation in risk associated with varying distances from the facility. Their analysis provides an excellent case study for these six sites. However, because of the limited number of facilities studied, it is difficult to draw broader inferences from their work.

In this paper, we first provide a theoretical model that clearly predicts increased population density in neighborhoods that experience an exogenous increase in public goods but yields only tentative predictions about its effect on neighborhoods' composition. To test these hypotheses, we use a difference-in-difference model, identifying the impact of entry and exit of facilities that are required to report their use of chemicals for the Toxics Release Inventory (TRI), as well as changes in toxicity-weighted emissions levels, on population and demographic composition. We also test for lagged responses to differing baseline exposure levels, which were not publicly announced until 1989. As our unit of analysis, we use a set of "communities"

[3] This concern will hold whether the appropriate "subregion" is a metropolitan area, urban core, state, or any region whose spatial definition is larger than that of a Census tract.

[4] The second factor is driven both by heterogeneity in the location of facilities within tracts (i.e., center versus border) and by differences in tract size.

defined randomly in space by equally spaced half-mile circles. We control for demographics and other location-specific effects using linear regression and, because migratory responses may be highly nonlinear functions of demographics, using a bias-corrected nonparametric matching estimator (Abadie and Imbens 2005). We find clear evidence of migration correlated with TRI facility emissions and their arrival or exit from a community. Furthermore, we find evidence that TRI facilities cause the composition of a community to become poorer and less white over time.

Thus, we find support for the fundamental mechanism underlying the Tiebout model: households do appear to vote with their feet in response to changes in public goods. Moreover, our results, given their context in an environmental amenity, have a specific implication for the environmental justice literature: namely, that correlations between pollution and poor or minority population may be due to Tiebout sorting rather than discrimination.

II. Model

To motivate the empirical work that follows, we begin in the spirit of Tiebout (1956) and explore the impacts of changing environmental quality on community composition within a general equilibrium model of locational choice. In particular, we use a model of vertically differentiated communities introduced by Epple et al. (1984), a more general version of which was recently applied to environmental improvements by Sieg et al. (2004).

Assume a continuum of households that are characterized by their income y and demographic group t. The joint distribution of types and income is given by $f(y,t)$. The marginal distribution of income is given by $f_y(y)$, and the distribution of income conditional on type t is given by $f_y^{\,t}(y)$. Household preferences are defined over housing with price P, a numeraire whose price is normalized to 1, and environmental quality G. Household i's preferences are represented by the indirect utility function

$$Vi = V(yi,P,G). \qquad (1)$$

Each household chooses to live in a community $j \in J$ and, conditional on community choice, chooses a quantity of housing D_i. Each community is characterized by its supply of housing S_j and level of environmental quality G_j, both of which are exogenously determined. To facilitate a characterization of the equilibrium sorting of households across communities, we further assume that household preferences satisfy the “single crossing” property. This condition requires that the slope of an indirect indifference curve in the (*G,P*) plane be increasing in *y*. [5] Although household demand for public goods in this simple model is differentiated only by differences in income, the model can be extended to include heterogeneity in tastes, as in Epple and Sieg (1999), without altering the primary insights derived here.

Given the assumption of single crossing, equilibrium can be characterized by an ordering of communities that is increasing in both *P* and *G*. That is, there is a clear ordering of communities from low-price, low-quality communities to high-price, high-quality communities. Further, for each pair of "neighboring" communities (as sorted by this ranking), there will exist a set of boundary households (defined by an income level) that are indifferent between the two communities. Households whose income is below the boundary income will prefer the lower-ordered community, and those whose income is above the boundary income will prefer the higher-ordered community. This leads to perfect income stratification of households across communities.[6] Equilibrium prices P_j and boundary incomes $\overline{Y}_{j,j+1}$ are implicitly defined by the equilibrium conditions of equation (2):

$$
\begin{aligned}
&V(\overline{Y}_{j,j+1}, P_j, G_j) = V(\overline{Y}_{j,j+1}, P_{j+1}, G_{j+1}) \quad \forall j \in \{1,...,J-1\} \\
&M \int_{y \in C_j} D(y, P_j, G_J)\, f_y(y)\, dy = S_j \quad \forall j \in \{1,...,J\}
\end{aligned}
\qquad (2)
$$

,

where M is the total mass of households, D(.) is housing demand, and Cj is the set of incomes locating in community j. These equations formalize the J-1 boundary indifference conditions and the requirement that the land markets clear in each of the J communities, yielding 2J-1 equations to identify the 2J-1 endogenous variables.

We use the model to consider two issues important for the analysis of migration and environmental gentrification. First, we consider the implied distribution of households across communities for demographic groups with different income distributions, $f_y^t(y)$. Second, we evaluate how the predicted demographic compositions of communities change in response to changes in environmental quality. Consider two demographic groups, Type 1 and Type 2.

---

[5] For a discussion of the single crossing property in this context, see Epple and Sieg (1999).

[6] It is straightforward to relax this assumption by introducing heterogeneity in tastes so that there is heterogeneity of income within each community, but perfect stratification by tastes for each income (see Epple and Sieg 1999 and Sieg et al. 2004). Accordingly, this assumption is not critical for the following implications of the model.

Assume that their conditional income distributions $f_y^{\,1}(y)$ and $f_y^{\,2}(y)$ are such that the mean income for Type 1 individuals is less than the mean income for Type 2 individuals. Figure 1 provides a graphic representation of the distribution of these demographic types in a system of two communities, with Community 1 having the lower (P, G) pair and Community 2 the higher. All households to the left of $\overline{Y}_{1,2}$ sort into Community 1, all to the right sort into Community 2. Obviously, Community 1 has a lower average income than Community 2. As shown in the figure, in this example Community 2 will have a much higher concentration of Type 2 individuals and Community 1 a much higher concentration of Type 1 individuals. Thus, Tiebout sorting with heterogeneity in income can induce correlations between race and pollution. (Nevertheless, although this correlation is guaranteed under the central case where the two distributions have identical higher moments, it need not occur in general.)

The above results are completely expected given the model. However, the comparative statics associated with a change in environmental quality in one of the communities is more subtle. Consider the impact of an increase in the environmental quality in the lowest G community in a system of two communities. Evaluating the resulting demographic responses requires identifying the shift in the income boundary, $d\overline{Y}_{1,2} \,/\, dG_1$. To evaluate this shift, we assume that housing demand is separable from G and apply the implicit function theorem to the boundary indifference condition and two market-clearing conditions from equation (2). This yields the following comparative static relationship:

$$\frac{d\overline{Y}_{1,2}}{dG_1} = \frac{-V_{G_1}^{\;1}}{(V_y^{\;1} - V_y^{\;2}) - f_y(\overline{Y}_{1,2}) \left[\dfrac{D(P_1, \overline{Y}_{1,2}) V_{P_1}^{\;1}}{\int_0^{\overline{Y}_{1,2}} D_{P_1}(P_1, y) f(y)dy} + \dfrac{D(P_2, \overline{Y}_{1,2}) V_{P_2}^{\;2}}{\int_{\overline{Y}_{1,2}}^{\infty} D_{P_2}(P_2, y) f(y)dy} \right]}, \quad (3)$$

where

$$V^1 = V(\overline{Y}_{1,2}, P_1, G_1),$$
$$V^2 = V(\overline{Y}_{1,2}, P_2, G_2),$$

$D(.)$ is the household demand function, and subscripts denote partial derivatives.

The key to signing the derivative in equation (3) is to recognize that the single crossing property implies that $(V_y^1 - V_y^2) < 0$ implying that $d\overline{Y}_{1,2} / dG_1$ is positive.[7] Figure 2 illustrates the impact of an increase in G1 on the equilibrium sorting. In response to the change, the indifference boundary $\overline{Y}_{1,2}$ moves to the right and the set of households in the shaded region A relocate from Community 2 to Community 1. If G1 were to fall instead, or if G2 were to increase instead of G1, an opposite shift would occur. At the aggregate level this change leads to an increase in population for Community 1 and a decrease in population for Community 2. (And to an increase in prices in Community 1 and a decrease in Community 2.)

What is the change in the community 1's composition relative to community 2? Surprisingly, the model does not offer clear predictions. Consider first the effect on income distributions. As the bordering households move from Community 2 to Community 1, Community 1 gets richer. We call this an absolute composition effect. But meanwhile Community 2 loses its lowest-income residents and therefore also experiences an increase in average income. Thus, we have the counterintuitive result that increasing the level of G1 leads to an increase in average income for both communities! So the relative composition effect is indeterminate. Not surprisingly, the effect on relative racial composition is also indeterminate. In fact, so is the absolute effect: as Community 1 increases in average income, it does so by gaining the richer Type I individuals as well as poorer Type II individuals. In general, the ratio of new Type II to new Type I individuals can be either greater or lesser than the existing ratio, so the percentage of Type I individuals can increase or decrease.[8]

We thus have three propositions.

Proposition 1 (scale effect). For any two communities, a marginal increase in public goods in one community relative to the other will cause population to rise in the community experiencing the improvement and to fall in the other community.

---

[7] By the definition of $\overline{Y}$, $V(\overline{Y}_{1,2}, P_1, G_1) = V(\overline{Y}_{1,2}, P_2, G_2)$. Since all those with incomes higher than $\overline{Y}$ prefer Community 2, $V(\overline{Y}_{1,2} + \varepsilon, P_1, G_1) < V(\overline{Y}_{1,2} + \varepsilon, P_2, G_2), \forall \varepsilon > 0$.

[8] As an example, consider a set of three minorities with incomes {10k, 30k, 50k} and a set of whites with incomes {40k, 60k, 80k}. These are two symmetric distributions differing only in the location parameter, with a higher mean for whites. Yet if $\overline{Y}_{1,2}$ is initially at \$45k and shifts to \$55k, the composition of Community 1 will change from two-thirds minorities to three-quarters minorities.

Proposition 2 (absolute composition effect). Ceteris paribus, a marginal increase in public goods in any community will increase its average income. The effect on the share of racial or other demographic groups is indeterminate.

Proposition 3 (relative composition effect). The change in the average income, or mean share of demographic groups positively correlated with income, in a community experiencing a marginal increase in G, relative to another community, is indeterminate.

These results suggest that for small changes in Community 1's environmental quality, there are no clear predictions for the relative change in community compositions. This negative conclusion is mitigated by two factors. First, if we consider a larger change—one that raises the level of environmental quality in Community 1 above that of Community 2—clearer predictions arise. Such a change will cause the populations in Communities 1 and 2 to switch places, resulting in an increase in average income in Community 1, while average income drops in Community 2. Second, when we consider many communities instead of just two, the affect of a change in public goods in Community 1 will intuitively have the largest effects on the composition of close substitutes, with effects on other communities dampening out in the rank ordering (a pattern we have confirmed in simulations). If these more distant communities act as a control group, we would expect to find a relative composition effect. Thus, despite our inability to predict relative composition effects in general, there remain plausible reasons for expecting an increase in income among communities experiencing an exogenous improvement in public goods. One might also expect a similar increase in the share of demographic groups correlated with income, such as race, but this remains purely an empirical matter.

In summary, our strongest hypothesis is that in a Tiebout model, changes in amenities should cause a scale effect. A somewhat weaker hypothesis, based on the mitigating factors discussed above, is that it would have a relative composition effect in terms of income. We do not have a theoretically based hypothesis for composition effects in terms of race or other groups correlated with income.

Empirical Strategy

We test for such demographic scale and composition effects in this paper. Our work is most similar in spirit to Kahn (2000), who presents general evidence for the Tiebout hypothesis, documenting over-all county-level population growth in California that is correlated with ozone improvements. However, his work did not consider composition effects, and the fairly aggregate county-level analysis bears further scrutiny at more local levels.

Our analysis is at a much finer level of aggregation, focusing on communities defined by half-mile-diameter circles. To test scale effects, we relate 1990–2000 changes in community populations (in level terms and percentage terms) and changes in racial composition to changes in exposure to air pollution. For composition effects, we focus on changes in median income, race, and single-parent households. Environmental quality is measured as the toxicity-weighted exposure to air pollution released from sites listed in the Toxics Release Inventory.

We employ a difference-in-difference design for the effect of 1990–2000 changes in pollution on 1990–2000 changes in these demographic measures. One potential weakness with this analysis is that if firms site their facilities based on racial composition or other confounding factors, the relationship between changes in pollution and changes in demographics would be endogenous. To address this problem, we also identify lagged demographic responses to TRI sites that predated the demographic changes and can therefore be treated as exogenous. We discuss these and other identification issues in more detail following a description of the data.

III. Data

Constructing the data set necessary to test for environmentally induced migration requires three related tasks. First, we identify a set of spatially delineated "communities." Second, we construct demographic composition measures for each community for 1990 and 2000. Finally, for each community we construct measures of the toxicity-weighted level of exposure to air pollution in 1990 and 2000 based on data from the Toxics Release Inventory of the U.S. Environmental Protection Agency (EPA). This section of the paper begins by discussing the spatial definition and data matching for the communities in our data set. We then discuss in more detail the demographic and pollution measures utilized in the empirical analysis.

Definition of Communities

Our analysis requires a set of communities whose boundaries remain fixed between 1990 and 2000. One approach would be to use Census tracts, block groups, or blocks as our community definition. This approach is problematic because these definitions change. Been (1997) found that nationally, one-fifth of tracts had changed boundaries between decennial censuses. Been (1997) and Wolverton (2002) address this problem by aggregating up to the greatest common boundaries, but in the process they still eliminate some areas and end up with higher levels of aggregation.

An alternative approach has recently become available with the release of Geolytic's Neighborhood Change Database. This product provides access to 1960 through 2000 Census data consistently aggregated to 2000 Census tract definitions. Although this new data set solves the problem of changing definitions by providing aggregation to the 2000 tract boundaries, the use of Census tracts as "communities" may itself be problematic for three reasons. First, Census tracts are locally defined to create relatively homogenous entities. Although some see this as a virtue because it gives more integrity to the concept of community, such gerrymandering may also bias results. For example, if polluting sites have an impact on the demographics of only the most local neighborhoods, and these neighborhoods are conjoined with other neighborhoods with similar characteristics to form Census tracts, it would induce correlation between the polluting site and the wider geographic entity (the tract). Second, although roughly equal in population, Census tracts range greatly in size. For example, in California, tracts range from less than a tenth of a square mile to more than a thousand square miles. This creates problems in controlling for this large degree of heterogeneity when estimating migration models. Finally, previous research on the correlation between race and environmental quality has shown that results can be quite sensitive to community definitions (Anderton et al. 1994; Hersh 1995). Census tracts may be too aggregate a unit and in any case preclude sensitivity analysis along these dimensions.

For those reasons we take a different approach to neighborhood definitions. We define neighborhoods as a set of half-mile-diameter circles (alternatively one-mile-diameter circles) evenly distributed across our study area. Using the GIS software package ARCVIEW, we construct weights that are used to attach environmental quality data from the TRI and demographic data from Census blocks to our communities.

The specifics of the data construction are as follows. First, to keep the data construction task manageable, we restrict our analysis to California. California is attractive because of its racial heterogeneity and because of its relative size. To further restrict the size of the data task and to reduce the heterogeneity between different communities, we limit our analysis to locations that were denoted as urban in the 1990 census. We construct our communities by first placing an equidistant grid across our study area. Grid points are one-half mile apart for the half-mile circles and one mile apart for the one-mile circles. If the study area were small—a single county, for instance—it would be possible to treat lines of latitude and longitude as an equally spaced grid. However, because of the size of California, the distance between lines of longitude varies by approximately 13 percent as one moves from its southern border to the northern border. To avoid an uneven sampling density between northern and southern portions of the state, it is necessary to account for this variation.[9]

Once grids that cover the entire state have been constructed (one each for the half-mile and one-mile circles) a quarter-mile (half mile) buffer is placed around each point in the grid, yielding a set of circles a half-mile (one mile) in diameter that are evenly distributed across the state. The set of circles that fall within Census's 1990 urban area boundaries are then selected, and all circles that lie across water are dropped. This process yields 6,218 "communities" within one-mile circles and 25,166 "communities" based on half-mile circles. Figure 3 shows the distribution of communities across the study area.

Census Data

As noted above, we aggregate demographic data from the 1990 and 2000 censuses into our circle-communities. We collected block-level data on the total populations of each racial group, both as individuals and as households, and economic variables including homeownership rates, rental rates, and self-assessed home values. We also collected block-group-level data including average incomes, educational attainment, and workforce descriptors.

Demographic count data on numbers of individuals or households by race and other categories are assigned to our communities. Specifically, for each block, a share of each demographic count is assigned to communities based on the percentage of the block's geographic area lying within each community.[10] Even for our half-mile communities, most blocks are assigned entirely to a single community, and 99 percent are assigned to five or fewer. Table 1 summarizes the opposite mapping, the number of blocks assigned to each community. Because of the splitting of blocks between 1990 and 2000, there are more blocks assigned per circle in 2000 than in 1990. The 50th percentile ranges from a low of 10 blocks per circle for 1990 half-mile circles to 38 for 2000 one-mile circles.

Table 2 provides descriptive statistics of the demographic characteristics of the half-mile communities. Note that they are easily comparable, since they are all circles of equal size

---

[9] The grids are constructed using the following factors: 1 degree of latitude = 69.172 miles and 1 degree of longitude = cos (latitude) * 69.172.

[10] For block group–level data, the values were distributed to the blocks based on population shares, then distributed to the communities as for the block-level data.

(approximately 0.785 square miles). Most communities have small populations, with an average 1990 population of about 772 persons, but with a wide interquartile range of 98 to 1162. Most communities are also largely non-Hispanic white, with the mean share of 67 percent across communities. However, this share decreased on average by 9 percent from 1990 to 2000. Moreover, whites are disproportionately located in less dense communities; weighting by population reduces the mean white share to 54 percent. Hispanics are the largest and fastest growing minority, with an average 19 percent share in 1990, 27 percent when weighted by population. Blacks and Asians represent smaller minorities, with average 1990 shares of 5 and 8 percent, respectively. Again, there is substantial variation in these data. Among communities with more than 100 residents, each racial group is completely absent from at least one, and each racial group makes up 90 percent or more of the population of at least one.

TRI Data

As a measure of pollution exposure we use EPA's Toxics Release Inventory to find releases of air pollution at facilities throughout the United States. These sites may be perceived as a disamenity through several channels. First, they may simply be visually unattractive sites. Second, their air pollution may have an unpleasant odor or may cause respiratory irritation. Finally, since 1987, facilities handling more than 10,000 pounds each year of certain hazardous chemicals have been required to report these emissions. These data first became publicly available in 1989 and are routinely publicized by environmental activists.[11] Because these data first became available immediately prior to our 1990–2000 study period, we might also expect to find lagged migratory responses to earlier emissions.

TRI facilities have been a focus of the environmental justice literature.[12] After early work looking only at the presence of a facility or total tons of emissions of all kinds, recent work has begun looking at emissions weighted by *toxicity*, so that more potent pollutants are given more weight than less potent ones. In the first attempt at such weighting, Sadd et al. (1999) estimated an ordered probit model on Los Angeles neighborhoods with no TRI releases, with noncarcinogenic releases, and with carcinogenic releases. In followup work, Morello-Frosch et al. (2001) use a linear index of cancer risks. Both these papers make the arbitrary assumption that

[11] The 1987 data were released with a two-year lag, thus becoming available in 1989.

[12] E.g., Arora and Cason (1996), Brooks and Sethi (1997), Kriesel et al. (1996), Morello-Frosch et al. (2001), Rinquist (1997), Sadd et al. (1999), Wolverton (2003).

carcinogenic releases are more important than noncarcinogenic releases. However, such is not necessarily the case, as some pollutants such as particulates have strong correlations with cardiovascular deaths and some carcinogens may have very weak potency.

As an alternative, we use a toxicity-weighted index of all emissions in the TRI[13] based on toxicity weights developed by EPA and available in its Risk Screen Environmental Indicators model (RSEI).[14] These weights have been used in recent work by Ash and Fetter (2002), who also go further to look at modeled exposure based on atmospheric dispersion. In sensitivity analyses, we also consider nonweighted emissions and the simple presence of a polluting site. Because of the TRI reporting threshold, these data all involve some measurement error. As we discuss below, as in the usual case, this is likely to have a conservative effect on our results, biasing them to zero.

The latitude and longitude for each facility were taken from a recent careful quality control analysis by EPA.[15] This geographic information allows a match of facilities to our communities. We construct buffers (quarter-mile and half-mile) around each TRI site and then assign emissions from a given TRI site to the communities that lie within the given buffer. The sample of TRI sites is the 2,311 California TRI sites located such that a half-mile buffer intersects at least one community. Figure 4 illustrates the approach used to assign emissions for the case of half-mile TRI buffers and one-mile circles. In the figure, the shaded circles are half-mile buffers around four TRI sites. The unshaded circles are communities. Emissions from a given TRI site are assigned to communities based on the percentage of their buffers that lie within a given community. For instance, in Figure 4, 3.1 percent of the emissions from TRI site A are assigned to community N1, 17.9 percent to community N2, and 52.4 percent to community N3. In this way, we can consistently aggregate emissions levels from multiple TRI sites into a total exposure in each community.

Table 1 summarizes the assignment of all TRI sites, active at some time during the 1988 to 2000 time period, to communities for each community-size, buffer-size pair. For those communities that are exposed to at least one TRI site, the table summarizes the distribution of

---

[13] The list of reporting chemicals greatly expanded in 1994. To maintain a consistent comparison of TRI emissions over time, we have limited the data to the common set of chemicals used since 1988.

[14] Information about this model is available at http://www.epa.gov/opptintr/rsei/.

[15] This quality control analysis provides a predicted accuracy for each site's location data. Fifteen sites are dropped because of poor-quality geo-coding data.

the number of sites to which each community is exposed. In all cases, the 50th percentile is either 2 or 3, with the maximum exposure going as high as 69 sites in the case of one-mile circles and one-mile TRI buffers. On a community basis, Table 2 indicates that 10 percent of half-mile communities were exposed in the baseline period (1988–1990), with 4 percent losing exposure by 1998–2000 and 1 percent gaining exposure. It also shows mean toxicity-weighted exposure among all communities and among those exposed.[16]

Additional Spatial Variables

Several factors, including other spatially distributed amenities, are likely to drive sorting across communities and should be controlled for as well as possible. As controls, we include in some models coarse controls for location effects, including distance to the coast and degrees latitude. However, our main approach to controlling for unobserved spatial amenities is local fixed effects. We use two sets of fixed effects: school districts and zip codes. Both are very local measures that are consistent with the notion that households are likely to choose a larger area based on other factors and then sort within that area based on the most local amenities. School districts have the advantage of mapping directly into an important local public good whose quality is otherwise notoriously difficult to measure. We find the share of each community that falls within each of the 226 school districts in our urban areas and assign a continuous variable on [0,1] to that community for each school district. Seventy-six percent of half-mile communities lie entirely within one school district, 21 percent within two, and the remainder within three or four. Table 1 shows the number of communities falling within each school district. Zip codes are even more local measures. Here, we assign each community to one of the 883 zip codes in our area based on the zip code of the community centroids. Table 1 reports the distribution of communities across zip codes. The median zip code is assigned 21 half-mile circles and 6 one-mile circles.

IV. Estimation and Results

Using those data, we test for differential changes in community population and demographics that are induced by changes and baseline differences in TRI emissions. Our

---

[16] This is in contrast to the approach that has typically been taken in the literature of assigning the pollution from a given facility to the Census tract in which it resides. The traditional approach is particularly problematic for handling facilities located on or near the boundary of their Census tracts. The ability to overcome this problem is one of the advantages of the methodology used here.

primary results center on models using half-mile diameter communities and half-mile diameter buffers around TRI facilities. We consider respective one-mile diameters in sensitivity analyses.

Contemporaneous Patterns in Exposure

We begin by confirming, in our own data, the premise that there is a correlation between pollution and minorities, as found in the environmental justice literature. To this end, we first test for correlations between 1990 TRI locations and community composition. For each composition outcome of interest (median income, percentage nonminority, percentage single-parent households), we regress the groups' representation in a given community on the presence of a TRI site and TRI emissions, as well as density and school district fixed effects. The basic model is presented in equation (4) for the income case.

$$INC_i = \beta_{0\mathrm{r}} + \beta_{\mathrm{yr}} y_i + \beta_{Dr} D_i + \beta_{Lr} L_i + e_{ir.} \qquad (4)$$

where i indexes communities, *Inc* is median income, y_i is the observed toxicity-weighted TRI exposure of a community, D is a vector of other demographic variables, and L is a vector of locational variables (including school district fixed effects).

The results, shown in Table 3, indicate that both the presence of a TRI site and TRI emissions are associated with poorer and minority populations, though the relationship with single-parent households is not the expected sign. Having confirmed this pattern in a single cross-section, our next goal is to develop a model that allows us to test for migratory responses associated with changes in toxic emissions from TRI sites.

Estimation Strategy

The strongest prediction of the Tiebout model is that the introduction of a TRI facility should cause individuals to leave the community (and that the exit of a facility should cause them to enter). To test this hypothesis, we regress both level changes and percentage changes in population from 1990 to 2000 on baseline TRI exposure, changes in TRI exposure, and other controls.[17] To test for composition effects, we regress changes in median income and changes in the share of white residents and single-parent households on the same variables. As discussed

[17] To develop an operable definition of a percentage change, we use the average of the 1990 and 2000 levels in the denominator. Within our data, this measure is approximately normally distributed and is bounded above and below by +2 and –2, respectively.

below, we also consider a nonparametric matching estimator as an alternative to linear regression.

TRI exposure is measured as the three-year lagged average, anchored respectively on 1990 and 2000, of the toxicity-weighted emissions of the 1988-defined chemicals, allocated to each community as described previously. As exposure variables, we include measures of both shocks and baseline exposure. As shocks, we include discrete indicators for when a community changes status from exposed to not exposed (or vice versa),[18] plus continuous measures of the change in emissions levels (which picks up the magnitude of those entering or exiting facilities as well as changes at those continuously emitting). Because in reality populations do not adjust instantaneously, we also include an indicator and continuous measure of 1990 exposure to pick up lagged reactions to previous exposure. These lagged responses may be particularly important when we consider that emissions levels were not publicly available before 1987.

The model for the scale effect regression analysis is presented in equation (5):

$$\begin{aligned}\Delta POP_i = {} & \delta_0 + \delta_{BL} I_i^{BL} + \delta_{NEW} I_i^{NEW} + \delta_{EXIT} I_i^{EXIT} \\ & + \delta_y y_i^{1990} + \delta_{\Delta y+}(\Delta y_i | \Delta yi > 0) + \delta_{\Delta y-}(\Delta y_i | \Delta yi < 0) \\ & + \delta_D D_i + \delta_L L_i + u_i \qquad (5)\end{aligned}$$

where ΔPOP is the change (or percentage change) in population from 1990 to 2000; I^{BL}, I^{NEW}, and I^{EXIT} are indicator variables for whether the community had any 1990 baseline exposure, went from no exposure to some exposure, or went from some exposure to no exposure; y_i^{1990} is the level of baseline toxicity-weighted exposure; $\Delta y_i | \Delta yi > 0$ is the change in toxicity-weighted exposure, if positive; and $\Delta y_i | \Delta yi < 0$ is the change in toxicity-weighted exposure, if negative. For regression models of composition effects, ΔPOP_i is simply replaced by the appropriate measure (change in median household income, change in share of whites, change in share of single-parent families).

To test our hypotheses, we focus on three treatment effects. The "average effect of baseline TRI exposure" estimates the average differential effect on a neighborhood's 1990 to 2000 population change from being exposed to TRI emissions in 1990. The "average effect of new TRI exposure" estimates the average differential effect on a previously unexposed neighborhood of becoming exposed to TRI emissions. And the "average effect of exiting TRI

[18] Note that the discrete variables indicate proximity of a community to any facilities over time, which is related to but not the same as the entry and exit of firms.

exposure" estimates the average differential effect on a previously exposed neighborhood of losing all its TRI exposure. These treatment effects are calculated as combinations of the estimated coefficients on both indicator and continuous variables. Specifically,

$$\text{Average baseline treatment} = \hat{\delta}_{BL} + \hat{\delta}_y \left(\frac{1}{N_{BL}} \sum_{i \in BL} y_i^{1990} \right), \qquad (6)$$

$$\text{Average new treatment} = \hat{\delta}_{NEW} + \hat{\delta}_{\Delta y+} \left(\frac{1}{N_{NEW}} \sum_{i \in NEW} \Delta y_i \right), \text{ and}$$

$$\text{Average exit treatment} = \hat{\delta}_{EXIT} + \hat{\delta}_{\Delta y-} \left(\frac{1}{N_{EXIT}} \sum_{i \in EXIT} \Delta y_i \right),$$

where, for example, N_{BL} is the number of communities with baseline exposure. Thus, the estimated effect of average baseline TRI exposure, relative to no exposure, is the estimated indicator variable for exposure, plus the estimated coefficient on the continuous measure of exposure times the average baseline exposure among those communities with baseline exposure. Similar logic holds for the effect of new and exiting exposure. Note that the first two treatments are relative to communities that never experience exposure, while the exit treatment is relative to the set of communities that had baseline exposure.

We estimate four basic regression models with different levels of control for confounding factors (the *D* and *L* variables). Our first model includes no controls. While clearly lacking any pretense to identifying causality, this model does give a signal as to the correlation between migration and changed exposure. Our second model controls for the baseline demographic variables listed in Table 2, including squares of these terms. As an important spatial amenity, it also includes the FBI crime rate imputed from overlapping political jurisdictions and spatial effects measured by latitude and distance to the coast in kilometers. Our third and fourth models contain the same demographic controls but replace the spatial variables with school district fixed effects and zip code fixed effects, respectively. Finally, all models are estimated with and without baseline population weights for the communities.

Before turning to the results, we consider some of the potential limitations associated with these data and with our experimental design, and we discuss their implication for interpreting our results. In most cases, the potential limitations will tend to bias our results toward zero, so that our tests can be described as conservative. First, note that we are essentially comparing treatment communities within our TRI buffers to control communities outside them, but we cannot know the *true* area of impact of the TRI facilities. If a TRI facility's actual impact is narrower than our buffers, the treatment communities will be contaminated by control areas, diluting the differential. On the other hand, if its actual impact is wider than our buffers, the TRI facilities will have some impact on control communities, again diluting the identified differential. Moreover, this latter effect is accentuated by our local fixed effects, since the nearest control communities are the most likely to be affected by the TRI facility.

Second, as noted above, TRI emissions are censored based on threshold quantities of individual chemicals handled by the facilities. This suggests that some low-level emissions go undetected. To the extent that this in turn means that some communities diagnosed as controls are in fact exposed, the estimated differential is again diluted. For these reasons, if we find migration effects with these data, using this design, we have reason to be confident in the existence of Tiebout effects related to pollution.

Finally, consider the potential issues related to endogeneity of the treatments. Our baseline treatment has the strongest claim to exogeneity, since historical exposure cannot be the consequence of future changes in local demographics. However, two potential problems arise with endogeneity for our new treatment and exit treatment. First, changes in a neighborhood's TRI emissions are likely associated with changes in that neighborhood's economic conditions. Such changes in economic conditions can reasonably be expected to be associated with changes in the neighborhood's population and/or demographic mix, leading to problems of endogeneity and biased estimates. Similarly, firms' decisions about opening and closing polluting facilities may be made partly in response to changes in local labor market conditions. Our response to these concerns is the inclusion of local spatial fixed effects for school districts or zip codes. The relevant spatial scale for considering economic conditions and labor market opportunities in a locational choice is likely much wider than the scale of environmental amenities. If labor and other economic opportunities are roughly equal within a given fixed effect, but if the environmental disamenities (actual pollution exposure, smell, or sight) of a polluting facility differ by our half-mile communities within the area of the fixed effect, this strategy will effectively eliminate the endogeneity problem.

A more troublesome problem with endogeneity arises for our new and exit treatments if polluting firms make entry and exit decisions based on the demographics of the community affected by the disamenity per se, as suggested in some of the environmental justice literature. Note, however, that the relevant issue is not the baseline demographics of the affected community but would have to be firms' response to *changes* in demographics from 1990 to 2000. Moreover, as noted previously, studies of actual firm siting decisions have found no strong correlation with community composition (Been 1997; Wolverton 2002), suggesting this kind of mechanism is not likely to pose a problem for our composition effects. On the other hand, Been (1997) and Wolverton (2002) have found that such decisions are correlated with population density, with facilities entering low-density communities, suggesting endogeneity may remain a concern for the scale effects in the case of these two treatments. Again, however, this problem is not present for the baseline treatment.

Migration: Scale Effects

The results from our scale-effect models are presented in Tables 4a and 4b. Both the weighted and the unweighted models fit reasonably well given the cross-sectional nature of the data, with R^2s of 0.04 to 0.18 for models with controls but no fixed effects and 0.09 to 0.58 for the fixed-effect models. Aside from the important impact of the TRI sites, we find that denser communities gain more people from 1990 to 2000, as do communities with lower housing prices but higher rental rates. We also find statistically significant nonlinear adjustments to baseline racial composition.

Table 4a presents the estimated scale effects associated with toxic emissions from TRI sites from the un-weighted regressions. The table includes estimates for models with both changes in population level and percentage changes in population as the dependent variable, for each of the three treatment effects. Both the change in level and percentage change models provide statistically significant, policy-relevant evidence of migratory scale effects consistent with the Tiebout hypothesis. Focusing on the percentage change model, on average, baseline exposure to TRI emissions is associated population declines that range from 10 to 16 percent, depending on the model. Likewise, the appearance of new toxic emissions in a previously untreated neighborhood is associated with population declines between 5 and 9 percent. Finally, the model predicts consistent responses in the opposite direction for communities that lose exposure. On average these communities are predicted to experience population gains of 5 to 7 percent.

The unweighted models take as their unit of analysis communities. They tell a story about what is happening at different places. As such, we view this approach as appropriate for evaluating the effect of Tiebout forces across spatially differentiated neighborhoods. However, from a policy perspective, we might be equally interested in understanding the average effect of these changes on the population. To better understand how populations are behaving, we rerun these regressions weighting by the baseline population. These weighted regressions are reported in Table 4b. The table indicates a similar qualitative pattern of migratory responses but with level effects somewhat higher and percentage effects much lower than the unweighted regressions.

This result is not unexpected, since the weighting scheme downweights less-populated areas, where larger percentage changes in populations are more likely to occur. In general, the effects continue to be statistically significant—with the exception of the estimated effect for new TRI exposure, which remains negative but loses significance in some models. We interpret these results as strong evidence in support of the scale effects predicted by our simple theory model.

To verify the robustness of these results, we employ a large number of sensitivity analyses. First, as an alternative to changes in population, which could be partly affected by family size and age, we used changes in the number of households as the dependent variable, continuing to find qualitatively similar and statistically significant effects. Second, we replicated the reported models with one-mile diameter (versus half-mile) communities. For baseline exposure, the effects are of greater magnitude (even in percentage terms) and greater significance for unweighted models and, for weighted models, are likewise greater for the models with no controls and basic controls, but quite similar for the models with fixed effects. The estimated effects for new and exiting TRI exposure are also similar.

Third, we tested many alternative definitions of the exposure variable. In particular, we used one-mile buffers around TRI facilities instead of half-mile buffers. We also used 1990 and 2000 emissions only (rather than three-year averages), raw emissions levels unweighted by toxicity, and a measure of emissions that treated each facility equally (so that communities differed only in the number of TRI sites to which they were exposed and their proximity to those communities). None of these sensitivity analyses changed the qualitative nature of the results, although the latter model did lower the magnitude of the effects, suggesting that actual pollution levels are important. Fourth and finally, we also estimated separate models on subsets of the data: on only those communities with no baseline exposure, to estimate the effect of a new exposure; on only those communities with baseline exposure, to estimate the effect of losing exposure; and on only those communities that do not change status over time, to estimate the effect of baseline exposure. None of these variations changed our results. Thus, our evidence is highly robust and strongly consistent with the Tiebout hypothesis.

Migration: Composition Effects

Although our theory model provides strong predictions regarding scale effects, it does not provide strong predictions about composition effects, relative to other communities, except for large changes in public goods that affect the relative rankings of the communities. Nonmarginal changes in exposure caused by exiting and entering TRI facilities may well qualify as such changes. In any case, these composition effects remain of empirical interest because they may partly explain the observed correlations between race and pollution exposure in the environmental justice literature.

To test for these composition effects, we reestimate the scale effects model using as dependent variables the change in median household income from 1990 to 2000, the change in the share of each community's population made up of non-Hispanic whites, and following Cameron and Crawford (2004), the change in the share of each community's population made up of single-parent families. (In other models, available from the authors, we also examined shares of individual racial minorities.) The fit is reasonably good with R^2s of 0.05 to 0.36 for models with statistical controls but no fixed effects and 0.17 to 0.68 for models with fixed effects. Baseline observables are typically significant. More urban areas are becoming more white and more Hispanic but less black and Asian. Neighborhoods with more expensive baseline housing are also becoming more white. Finally, highly white areas are generally becoming less white over time, suggesting a structural shift to less segregation, or simply randomness or entropy causing regression to the mean.

Table 5a presents the effects of average TRI exposure using the unweighted model. (The specification is identical to that of equation (6), with changes in community composition now on the left-hand side.) There is mixed evidence of a composition effect broadly consistent with intuition. Baseline and new TRI exposure cause communities to have a differential growth in median income of about $2,000 or $3,000 less than that experienced by the controls and to become 1 to 2 percent less white, statistically significant changes.[19] These treatments are also associated with increases in the share of a community with single-parent families, but the effect is generally not significant. Losing TRI exposure generally has the opposite effects, but the changes are generally not statistically significant. The population-weighted regressions, reported in Table 5b, generally show the same trends but with small magnitudes and generally statistically insignificant effects. Even though these estimated composition effects are less robust than those for the scale effects, they provide additional evidence in support of the general Tiebout model. The results are most consistent with expectations for the effect on income for the baseline treatment, where the model yields stronger predictions and where the treatment has the best claim at exogeneity. Again, we subjected these models to the sensitivity tests described above for the scale effects, and again found qualitatively similar results.

[19] Auxiliary models on minority groups indicate that most of the offsetting effect on whites is captured by Hispanics.

Before moving on to the matching estimates, we consider the effectiveness of our fixed-effects approach to controlling for unobserved heterogeneity. Underlying the regression analysis presented in Tables 4a through 5b is the assumption that the inclusion of school district and/or zip code fixed effects adequately controls for the presence of unobserved confounding variables. Although the inclusion of these controls is an improvement over the existing literature, it remains an open question whether these fixed effects provide adequate control for unobservable covariates. To help evaluate the effectiveness of these controls, Table 6 compares the predicted treatment effects from a model that includes as controls only zip code fixed effects to a model that includes all of our controls in addition to the zip code fixed effect. If the zip code fixed effect is successfully controlling for unobserved confounding factors, one might expect it to also effectively control for the impact of our observed spatially varying covariates. Therefore, if the treatment effects vary little between a model that includes only the zip code fixed effects and a model that includes zip code fixed effects and the observable controls, one might have more faith in the ability of these fixed effects to control for spatially varying unobservables. Table 6 presents 95 percent confidence interval for the three treatment effects under these two models. As the table shows, the zip code fixed effects appear to control quite effectively for the observable covariates, with point estimates close and confidence intervals overlapping.

Nearest-Neighbor Matching

As noted above, we find evidence of nonlinear migratory responses to baseline racial composition, suggesting it may be difficult to control for these effects parametrically. These controls are important, since the estimated composition effects are sensitive to the presence of such controls. (Note, for example, the large changes in point estimates from the “No Controls” to “Basic Controls” models in Tables 5a and 5b).

To better account for this nonlinearity and uncertainty about functional form, we also nonparametrically match each community receiving a TRI “treatment” to a set of control communities with similar observable characteristics, and then compare their migration patterns. Symmetrically, we match each control community to a set of treatment communities with similar characteristics. Under controlled experiments, a treatment is given randomly so that, by design, the expected values of unobserved variables are the same in the treatment and control groups. In a quasi experiment, treatment and nontreatment observations are grouped by other observed variables and compared conditional on those variables. In our case, the three treatments are the presence of baseline TRI exposure among the set of communities that do not change exposure status over time; the move to exposure among those communities that did not experience baseline exposure; and the ending of exposure among those communities exposed in the baseline. These three treatment definitions mirror the estimated treatment effects from the simple linear models presented in Tables 4 and 5.

This approach relaxes the need for functional form assumptions about the controlling variables. Further, it weakens the necessary assumptions regarding the error term, requiring only that, conditional on the observables, the expected value of the error term is equal for the treatment and control cases. This is in contrast to the classical assumption that, conditional on the observables (including the treatment variable), the expected value of the error term is zero.[20]

Under the standard matching model, observations are grouped by values of the observables (baseline racial composition, density, and other locational amenities or proxies) and, within each cell, differences in outcomes between treated and untreated observations are calculated. However, the number of cells required to do this can be prohibitively large. Rosenbaum and Rubin (1983) showed that when a large number of observed variables create too many empty cells, one can instead match on a univariate index representing overall distance in the space of observables between a treatment observation and its matched control(s).

We estimate the Euclidian distance between communities in the space of the same set of observables from the regression models. The universe of potential control communities is defined to eliminate confounding treatments. For each community receiving one of these respective TRI treatments, we compare its scale and composition effects with the average among four nearest (most similar) control communities. The difference, the estimated treatment effect, corresponds to the first terms in equation (7) (without the continuous terms). To account for the potential bias associated with a comparison with control communities having slightly different observables, we adjust for these differences using linear regression, as in Abadie and Imbens (2005).

The last row in each section of Tables 4 and 5 reports the sample average treatment effect from the nearest-neighbors matching estimator. Standard errors are computed using the approach proposed by Abadie and Imbens (2005). Qualitatively, most outcomes are unchanged from the regression models, with two exceptions. For the "new exposure" treatment, the outcomes for the raw scale effect (but not percentage changes) and for the change in the share of whites in the

[20] See Heckman et al. (1997, 1998) and Dehejia and Wahba (2002). Greenstone (2004) has recently applied this approach to air quality changes.

community reverse signs.[21] Overall, however, the results provide more evidence of the scale effects predicted by the model, and for the presence of composition effects posited by Been (1997).[22]

V. Conclusions

Tiebout's suggestion that people vote with their feet to find the community that provides their optimal tax and public goods pair has played a central role in the theory of local public finance. Despite its importance, however, this model has been subjected to few direct tests of its basic mechanism: households voting with their feet in responses to changes in these amenities.

Toward this end, we use changes in the emissions of toxic air pollutants across spatially delineated neighborhoods to test for environmentally motivated migration patterns associated with increased demand for land in improving neighborhoods, and which may alter the demographic mix between richer and poorer households. Our analysis follows on a small number of studies that have explored the link between changes in environmental quality and the prevalence of the poor and minorities. Using a new approach to community definition that overcomes the problems associated with the use of Census tracts, in conjunction with better controls for potentially confounding factors than have been used in previous studies, we provide the strongest evidence to date of the link between changes in environmental quality and local changes in community demographics. We find strong evidence of scale effects of a magnitude that is both statistically significant and empirically relevant. We also find evidence of composition effects that suggest that pollution in a given location is associated with the emigration of richer households and/or immigration of poorer households. Evidence of racial composition effects is weak. This pattern of results is consistent with our theoretical predictions based on a simple Tiebout model.

Beyond the ratification of a significant model in local public finance, these findings have potentially profound implications for environmental policy. In a world where households sort in

---

[21] These anomalies are eliminated when focusing on the average effect among treated communities only, in contrast to the average effect on the entire sample. Thus, they may be due to the small number of "new" treatments available for matching to the control communities.

[22] In addition to the simple nearest-neighbors estimates reported here, we also estimate the treatment effect by further restricting the match to those control communities located within the same school district as the treated community, an approach analogous to our fixed effects regression models. The results remain unchanged when we restrict the matches to those within the same school district.

response to changes in environmental quality, the bulk of the benefits of a policy that cleans up dirtier neighborhoods where the poor live may actually be captured by rich households. As the neighborhood amenity improves, wealthier households may move in, driving up rents. If the poor do not own their homes, landlords would capture the capital appreciation of the local housing, while the poor would pay higher rents. This “environmental gentrification” may more than offset the direct gain of the environmental improvement, leaving the original residents worse off. Such outcomes have been demonstrated in simulation models of air quality improvements in Los Angeles (Sieg et al. 2004) and increases in protection of open space in Raleigh, North Carolina (Walsh 2004).

References

Abadie, Alberto, and Guido Imbens. 2005. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica.* Forthcoming.

Anderton, Douglas L., Andy B. Anderson, John Michael Oakes, and Michel R. Fraser. 1994. Environmental Equity: The Demographics of Dumping. *Demography* 31(2): 229–48.

Arora, Seema, and Timothy N. Cason. 1999. Do Community Characteristics Influence Environmental Outcomes? Evidence from the Toxics Release Inventory. *Southern Economic Journal* 65(4): 691–716.

Ash, Michael, and T. Robert Fetter. 2002. Who Lives on the Wrong Side of the Environmental Tracks? Evidence from the EPA's Risk-Screening Environmental Indicators Model. Working Paper 50. Political Economy Research Institute, University of Massachusetts at Amherst.

Bayer, Patrick, Robert McMillan, and Kim Reuben. 2005. An Equilibrium Model of Sorting in an Urban Housing Market. Mimeo. New Haven, CT: Yale University.

Been, Vicki, with Francis Gupta. 1997. Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims. *Ecology Law Quarterly* 34: 3–56.

Brooks, Nancy, and Rajiv Sethi. 1997. The Distribution of Pollution: Community Characteristics and Exposure to Air Toxics. *Journal of Environmental Economics and Management* 32: 233–50.

Brueckner, Jan K. 1982. A Test for Allocative Efficiency in the Local Public Sector. *Journal of Public Economics* 14: 311–32.

Cameron, Trudy Ann, and Graham D. Crawford. 2004. Superfund Taint and Neighborhood Change: Ethnicity, Age Distributions, and Household Structure. Mimeo.

Dehejia, Rajeev H., and Sadek Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics* 84(1): 151–61.

Epple, Dennis, Radu Filimon, and Thomas Romer. 1984. *Journal of Public Economics* 24: 281–308.

Epple, Dennis, and Holger Sieg. 1999. Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy* 107: 645–81.

Ferreyra, Maria M. 2005. Estimating the Effects of Private School Vouchers in Multi-District Economies. Mimeo. Pittsburgh: Carnegie Mellon University.

Goldman, Benjamin A., and Laura Fitton. Toxic Wastes and Race Revisited. Washington, DC: Center for Policy Alternatives.

Gramlich, Edward M., and Daniel L. Rubinfeld. 1982. Micro Estimates of Public Spending Demand Functions and Tests of the Tiebout and Median-Voter Hypotheses. *Journal of Political Economy* 90(3): 536–60.

Graves, Philip E., and Donald M. Waldman. 1991. Multimarket Amenity Compensation and the Behavior of the Elderly. *American Economic Review* 81(5): 1374–81.

Greenstone, Michael. 2004. Did the Clean Air Act Amendments Cause the Remarkable Decline in Sulfur Dioxide Concentrations? *Journal of Environmental Economics.* Forthcoming.

Hamilton, James T. 1993. Politics and Social Costs: Estimating the Impact of Collective Action on Hazardous Waste Facilities. *Rand Journal of Economics* 24: 101–25.

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies* 64: 605–54.

———. 1998. Matching as an Econometric Evaluation Estimator. *Review of Economic Studies* 65: 261–94.

Hersh, Robert. 1995. Race and Industrial Hazards: A Historical Geography of the Pittsburgh Region, 1900–1990. Discussion Paper 95-18. Washington, DC: Resources for the Future.

Kahn, Matthew E. 2000. Smog Reduction's Impact on California County Growth. *Journal of Regional Science* 40(3): 565–82.

Kriesel, Warren, Terence J. Centner, and Andrew G. Keller. 1996. Neighborhood Exposure to Toxic Releases: Are There Racial Inequities? *Growth and Change* 27: 479–99.

Mohai, Paul, and Bunyan Bryant 1992. Environmental Racism: Reviewing the Evidence. In *Race and the Incidence of Environmental Hazards: A Time for Discourse,* edited by Bunyan Bryant and Paul Mohai. Boulder, CO: Westview Press.

Morello-Frosch, Rachel, Manuel Pastor, and James Sadd. 2001. Environmental Justice and Southern California's "Riskscape": The Distribution of Air Toxics Exposures and Health Risks among Diverse Communities. *Urban Affairs Review* 36(4): 551–78.

Oates, Wallace E. 1969. The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis. *Journal of Political Economy* 77(6): 957–71.

———. 2005. The Many Faces of the Tiebout Model. Working Paper.

Rinquist, Evan. 1997. Equity and Distribution of Environmental Risk: The Case of TRI Facilities. *Social Science Quarterly* 78: 811–29.

Rosenbaum, P., and D. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70: 41–55.

Rubinfeld, Daniel L. 1987. The Economics of the Local Public Sector. In *Handbook of Public Economics,* edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: North-Holland, 571–645.

Sadd, James L., Manuel Pastor, J. Thomas Boer, and Lori D. Snyder. 1999. "'Every Breath You Take …": The Demographics of Toxic Air Releases in Southern California. *Economic Development Quarterly* 13(2): 107–23.

Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh. 2004. Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods. *International Economic Review* 45(4): 1047–77.

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

Timmins, Christopher. 2003. If You Can't Take the Heat, Get Out of the Cerrado … Recovering the Equilibrium Amenity Cost of Non-Marginal Climate Change in Brazil. Mimeo. Durham, NC: Duke University.

United Church of Christ. 1987. Toxic Wastes and Race in the United States. Report of the Commission for Racial Justice.

Walsh, Randall P. 2004. Analyzing Open Space Policies in a Locational Equilibrium Model with Endogenous Landscape Amenities. Mimeo. Boulder: University of Colorado.

Wolverton, Ann. 2002. Does Race Matter? An Examination of a Polluting Plant's Location Decision. Mimeo. Washington, DC: National Center for Environmental Economics, U.S. Environmental Protection Agency.

———. 2003. The Demographic Distribution of Pollution: Does Neighborhood Composition Affect Plant Pollution Behavior? Mimeo. Washington, DC: National Center for Environmental Economics, U.S. Environmental Protection Agency.

Figures and Tables

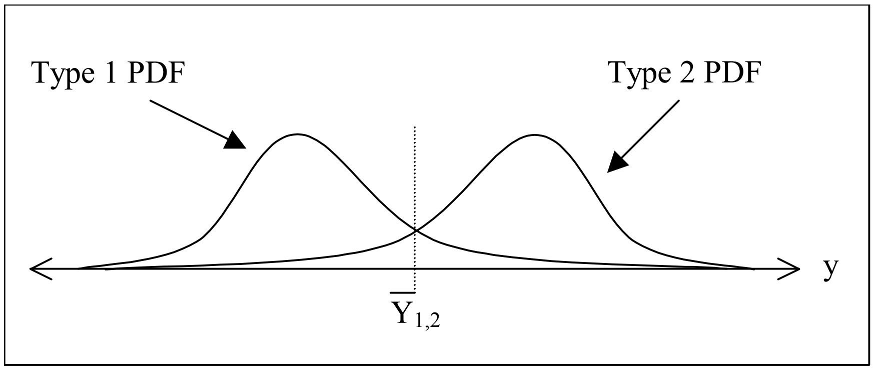

Figure 1. Density of income for two household types and community income boundary.

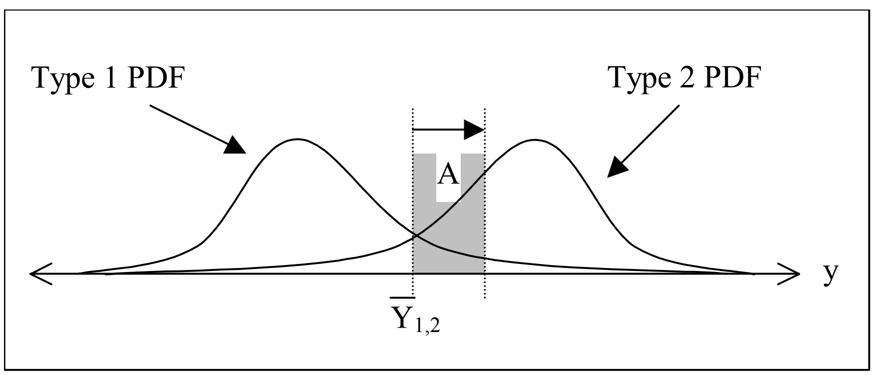

Figure 2. Shift in community income boundary after improvement in G1.

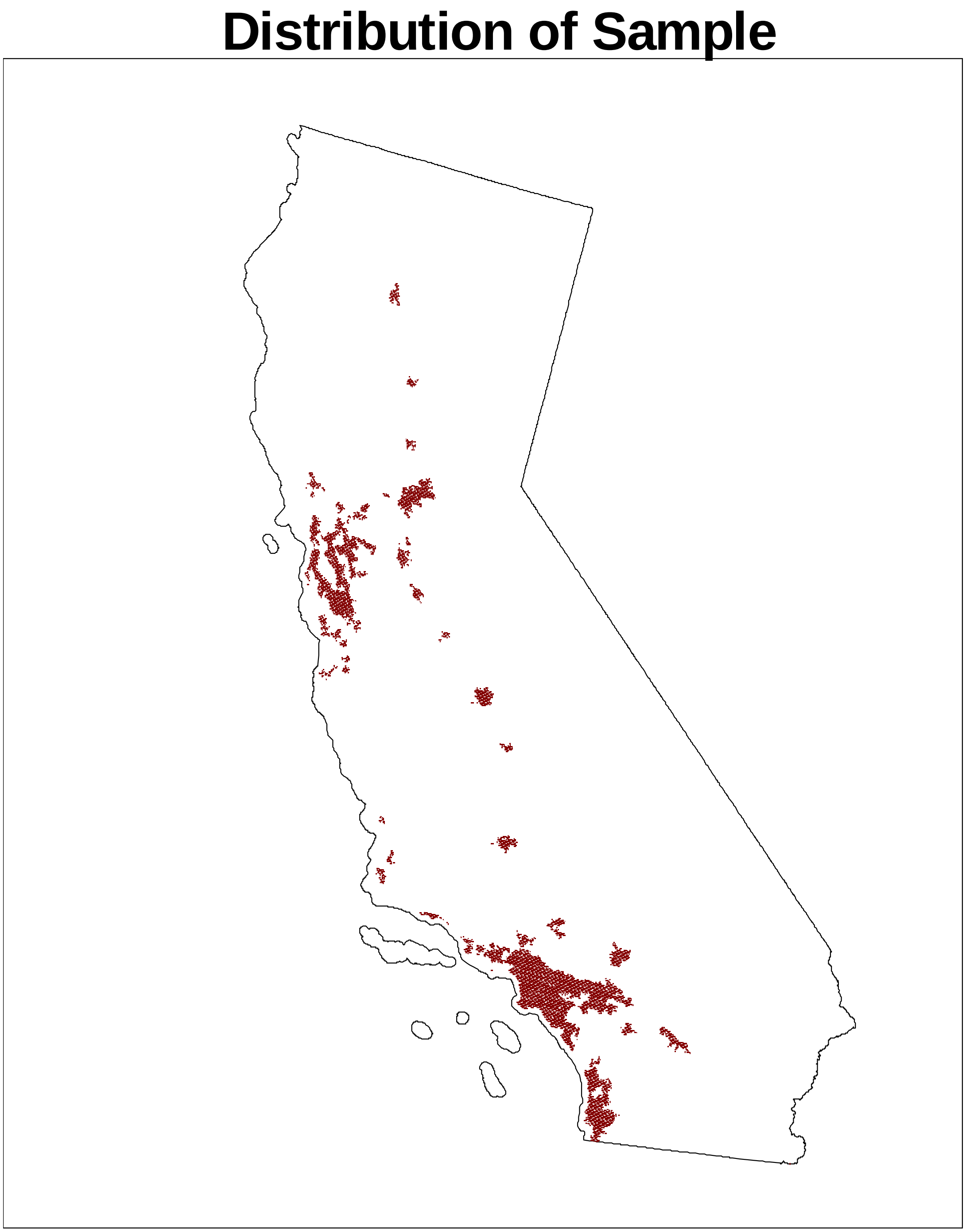

Figure 3. Distribution of "communities" in the study area, California.

Mapping TRI Site Emmisions to Neighborhoods

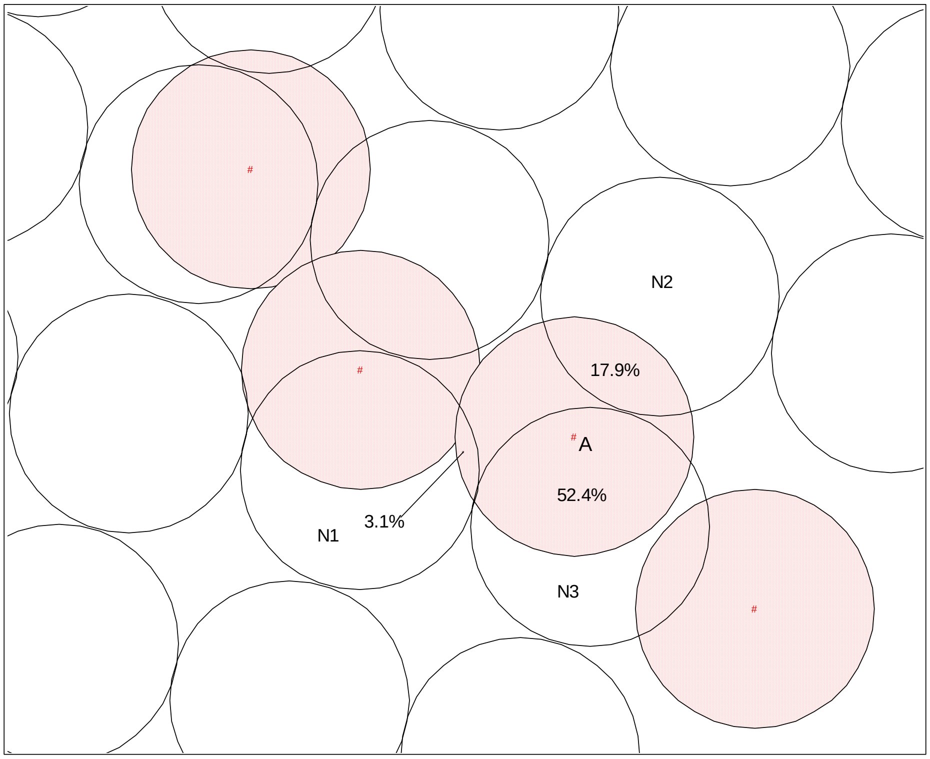

Figure 4. Half-mile buffers (shaded) and circle communities (unshaded) around four TRI sites.

Table 1. Assignment of Census blocks and TRI emissions to circle communities.

	Half-mile circles	*One-mile circles*
Count	*25,166.00*	*6,218.00*
1990 blocks per circle		
25th percentile	4	11
50th percentile	10	29
75th percentile	19	55
Max	132	383
2000 blocks per circle		
25th percentile	6	17
50th percentile	13	38
75th percentile	22	64
Max	136	408
Circles with TRI exposure		
1/4-mile buffer	3,109	1,295
1/2-mile buffer	5,179	1,795
TRI sites for exposed circles		
1/4-mile buffer		
25th percentile	1	1
50th percentile	2	2
75th percentile	3	4
max	19	25
1/2-mile buffer		
25th percentile	1	1
50th percentile	2	2
75th percentile	4	5
max	27	34
Circles per school district		
25th percentile	45	14
50th percentile	93.5	27
75th percentile	169	47
Max	2,352	620
Circles per zip code		
25th percentile	11	3
50th percentile	21	6
75th percentile	35	9
max	190	49

Table 2. Descriptive statistics of the data for half-mile circle communities.

Baseline demographic data (1990)	*Mean*	*Standard deviation*
Population (density)	772	930
Share black	0.05	0.11
Share Hispanic	0.19	0.20
Share Asian	0.08	0.10
Share other minority	0.01	0.02
Percentage households with single-parent families	0.08	0.07
Mean rental rate ($)	689	263
Mean housing value ($)	229,872	138,199
Share owning their homes	0.66	0.27
Percentage employed	0.94	0.05
Percentage of employed in manufacturing, if employed	0.15	0.08
Percentage not graduating from high school	0.10	0.07
Percentage with bachelor's degree	0.49	0.14
Median household income ($)	46,461	21,551
Changes in demographics (1990–2000)		
Population	92	256
Percentage change in population	0.09	0.67
Share non-Hispanic white	–0.09	0.13
Share black	0.01	0.06
Share Hispanic	0.05	0.11
Share Asian	0.03	0.08
Change in percentage of single-parent households	0.01	0.06
TRI data		
Share with baseline TRI exposure (1988–1990)	0.10	NA
Share with new TRI exposure (1998–2000)	0.01	NA
Share losing TRI exposure (1998–2000)	0.04	NA
Baseline emissions	300,714	4,718,020
Baseline emissions, among those exposed	3,006,542	1.46e7
Locational data		
1990 FBI crime index	0.08	0.28
Change in crime index	–0.03	0.14
Distance to coast	47.2	45.3
Degrees latitude	35.41	2.03
School or zip code fixed effects	NA	NA

Table 3. Baseline composition as function of exposure (OLS).†

	Dependent variable					
	Median household income		*Share white*		*Share single parent*	
Presence of TRI site	–13210	***	–0.141	***	–0.0064	***
Toxicity-weighted TRI emissions	–4.87e-5	**	–1.03e-9	***	–1.70e-10	*
Density	–7.582	***	–1.46e-6	***	–3.07e-6	***
R^2	0.47		0.49		0.06	

† All models contain school district fixed effects.

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

Table 4a. Estimated scale effects: Unweighted.†

	Average effect of baseline TRI exposure		*Average effect of new TRI exposure*		*Average effect of exiting TRI exposure*		R^2
Population levels							
No controls	–30	***	–13		43	***	0.00
Basic controls	–54	***	–35	**	39	***	0.07
School district fixed effects	–59	***	–35	**	42	***	0.11
Zip code fixed effects	–71	***	–36	**	45	***	0.26
Matching estimator	–32	**	27		31	***	—
Percentage change in population							
No controls	–15.6	***	–5.3		7.1	**	0.00
Basic controls	–11.7	***	–7.3	**	5.0	**	0.04
School district fixed effects	–10.3	***	–8.3	**	6.1	**	0.09
Zip code fixed effects	–12.0	***	–9.3	***	6.3	***	0.19
Matching estimator	–10.7	***	–12.11	*	4.3	**	—

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

†See equation (7) for definition of the treatment effects. Standard errors for matching models based on Abadie and Imbens (2005).

Table 4b. Estimated Scale Effects: Population-Weighted. †

	Average effect of baseline TRI exposure		*Average effect of new TRI exposure*		*Average effect of exiting TRI exposure*		R^2
Population levels							
No controls	–46	*	–18		81	***	0.00
Basic controls	–81	***	–39	***	71	***	0.18
School district fixed effects	–84	***	–31		78	***	0.25
Zip code fixed effects	–108	***	–42	**	78	***	0.58
Matching estimator	–43	**	42	***	46	***	—
Percentage change in population							
No controls	–2.6		0.8		3.0	**	0.00
Basic controls	–3.6	**	–0.7		2.6	*	0.05
School district fixed effects	–4.0	***	–1.0		3.0	**	0.10
Zip code fixed effects	–4.7	***	–1.6		2.9	***	0.24
Matching estimator	–1.1		–3.9	**	1.7	***	—

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

†See equation (7) for definition of the treatment effects. Standard errors for matching models based on Abadie and Imbens (2005).

Table 5a. Estimated composition effects: Unweighted†

	Average effect of baseline TRI exposure		*Average effect of new TRI exposure*		*Average effect of exiting TRI exposure*	R^2
Median income						
No controls	–3,132	**	–2,730	*	830	0.00
Basic controls	–2,798	**	–1,038		106	0.05
School district fixed effects	–2,619	**	–2,802	*	96	0.17
Zip code fixed effects	–2,044	*	–2,027		–648	0.35
Matching estimator	–4,188	***	–2,699	*	688	—
Share White						
No controls	–1.8	***	–1.6	**	0.4	0.00
Basic controls	–1.1	*	–1.0		0.1	0.21
School district fixed effects	–1.3	**	–1.6	**	0.3	0.35
Zip code fixed effects	–2.1	***	–0.5		0.2	0.45
–Matching estimator	0.2		3.6	***	0.2	—
Share single parent						
No controls	0.21		0.25		–0.22	0.00
Basic controls	0.22		0.19		–0.21	0.05
School district fixed effects	0.23		0.30		–0.23	0.37
Zip code fixed effects	0.02		0.19		–0.08	0.43
Matching estimator	–0.19		0.33		–0.01	—

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

†See equation (7) for definition of the treatment effects. Standard errors for matching models based on Abadie and Imbens (2005).

Table 5b. Estimated composition effects: Population-weighted†

	Average effect of baseline TRI exposure		Average effect of new TRI exposure		Average effect of exiting TRI exposure		R^2
Median income							
No controls	–2,548	***	–2,681	**	652		0.00
Basic controls	–1,347	*	–719		–219		0.11
School district fixed effects	–1,129		–1,948	*	–262		0.21
Zip code fixed effects	–743		–1,142		–603		0.35
Matching estimator	–2,252	***	–2,279	***	369		—
Share white							
No controls	1.3	***	–0.0		–1.1	***	0.00
Basic controls	0.3		–0.2		–0.0		0.36
School district fixed effects	0.0		–0.4		0.2		0.52
Zip code fixed effects	–0.5	*	–1.0	***	0.3		0.68
Matching estimator	0.4	*	0.1	***	–0.1		—
Share single parent							
No controls	–1.33	***	–0.82	***	0.45	***	0.01
Basic controls	0.21		0.33	*	–0.00		0.36
School district fixed effects	0.31	**	0.43	**	–0.14		0.44
Zip code fixed effects	0.38	***	0.31	*	–0.18	*	0.54
Matching estimator	–0.02		0.15		0.03		—

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

†See equation (7) for definition of the treatment effects. Standard errors for matching models based on Abadie and Imbens (2005).

Table 6. Sensitivity of zip code fixed effect models to inclusion of other observables: 95 percent confidence intervals for each case.

Outcome variable	*Population weighted?*	*Base treatment*		*New treatment*		*Exiting treatment*	
		Without controls	*With controls*	*Without controls*	*With controls*	*Without controls*	*With controls*
ΔPop	Y	(–125, –64)	(–136, –79)	(–75, 9)	(–81, –3)	(56, 110)	(53, 103)
	N	(–99, –56)	(–97, –46)	(–68, –7)	(–69, –4)	(33, 76)	(22, 67)
%ΔPop	Y	(–.060, –.015)	(–.070, –.025)	(–0.41, .022)	(–.047, .015)	(.012, .052)	(.009, .049)
	N	(–.212, –.090)	(–.179, –.061)	(–.159, –.001)	(–.169, –.018)	(.033, .145)	(.010, .116)
ΔMedInc	Y	(–2154, 724)	(–1871, 1025)	(–3046, 931)	(–2844, 1191)	(–1861, 683)	(–1862, 722)
	N	(–4224, 128)	(–3198, 1169)	(–4794, 779)	(–4418, 1274)	(-2598, 1319)	(–2772, 1234)
Δ%White	Y	(–.005, .007)	(–.010, .001	(–.006, .011)	(–.017, –.002)	(–.009, .002)	(–.002, .008)
	N	(–.010, .015)	(–.023, –.002)	(–.017, .014)	(–.030, –.003)	(–.021, .001)	(–.007, .012)