The Impact of Pollution Burden on Micro-Level Residential Sorting

Cristina Connolly
Agribusiness Department
California Polytechnic State University, San Luis Obispo
crconnol@calpoly.edu

Mitchell R. Livy
Department of Economics
California State University, Fullerton
mlivy@fullerton.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2.

The Impact of Pollution Burden on Micro-Level Residential Sorting

PRELIMINARY DRAFT: PLEASE DO NOT CITE

Abstract

The presence of environmental degradation results in significant costs to residents that are likely to affect location choices at the micro-level. This paper utilizes unique Census tract level pollution data to examine the extent that homeowners sort across measurable environmental quality, and to endogenizes the grant process for high-pollution communities. Initial results indicate that drinking water and PM 2.5 pollution burden levels significantly impact sorting behaviors.

Keywords: Residential sorting, environmental quality, location choice, pollution.

1. Introduction

The existence of preventable environmental hazards results in over $250 million in annual children's medical costs, and nearly 280,000 new child asthma cases each year in California (Public Health Institute 2015). The California Environmental Protection Agency (CalEPA) additionally recognizes spatial heterogeneity in pollution burden so severe that they have a variety of programs available to address this disparity. Given the economic ramifications of these outcomes, homeowners are likely to account for damages associated with potential environmental exposure when making location decisions. While the effect of pollution levels on residential location choices has been studied at aggregate spatial levels, there has not been significant structural research into the impacts of air and water pollution on sorting behaviors within metropolitan statistical areas (MSA). This paper addresses this gap in the literature by estimating the relationship between pollution burdens and residential sorting across Census tracts in Orange County, CA. Through the implementation of a vertical sorting analysis, we find that residents do internalize the pollution gradient when choosing a location.

Tiebout (1956) observed that households consider the complete bundle of amenities provided by each neighborhood when choosing a location. These amenities include housing attributes, and the larger set of environmental and local public goods. The theory of residential sorting behavior has been applied widely in the structural economics literature to study a variety of amenities, such as school quality (Bayer et al., 2007) and air pollution (Hamilton and Phaneuf, 2015). Despite the prevalence of sorting research, preferences for air quality have largely been studied at the macro-level, rather than the micro-level. For example, Bayer et al. (2009) utilize

MSA data to determine the impact of pollution, measured as PM 10 concentrations, on residential location choice. Hamilton and Phaneuf (2015) improve on this literature by controlling for the micro-level location choice in a model of sorting across MSAs; however, their model estimates pollution preferences at the macro-level instead of the micro-level. Our paper contributes to this body of research into preferences for environmental quality by investigating the micro-level impact pollution on location choice using a vertical model of residential sorting.

To empirically investigate this topic we use unique pollution data at the Census tract level to measure micro-level preferences for air and water pollution. This allows for improved policy evaluations of spatially targeted pollution reduction policies, as these policies typically occur at a more micro scale than has been addressed by the current literature. Implementing a vertical model of household location sorting, we find that households do sort across pollution burden levels of air and water quality. These results provide evidence that households consider micro-level pollution levels when making within-city location choices.

2. Data

We use data covering Orange County, California, an area with over 3 million residents. This region contains 575 Census tracts, which we define as the set of neighborhoods over which households may choose to locate. Neighborhood attributes are compiled using results from the 2013 American Community Survey (ACS). Compared to other southern California counties, Orange County is spatially small and contains relatively high amounts of development across most areas of the county. It therefore presents an ideal study area as most of the Census tracts are populous and uniform in size, with significant variation in both demographics and pollution measures.

The analysis combines data on housing and local public goods, which are defined as pollution levels. These pollution data are assembled using the California Office of Environmental Health Hazard Assessment (OEHHA) CalEnviroScreen 2.0 program, which is a database of disaggregated pollution variables measured at the Census tract level. This project includes information on multiple sources of pollution in California communities, including ozone, particulate matter 2.5 (PM 2.5), drinking water, impaired water bodies, and other environmental hazards. We focus on the drinking water and PM 2.5 measures as these are both easily observable and measurable. The drinking water statistic is calculated by the CalEnviroScreen 2.0 project as an aggregation of multiple pollutants such as arsenic, lead and uranium, to form a comparable measure of water quality across space. The PM 2.5 statistic is calculated as the concentration of airborne particles with a size less than 2.5 micrometers. These particles are comprised of dust, metals, and other substances which have been linked to negative health outcomes.

Under the vertical sorting model, consumer preferences for public goods are expressed through housing prices, which allows for the ranking of communities by desirability. In this study we use the American Community Survey (ACS) 2013 5-year housing price estimates to establish neighborhood housing desirability ranks. In order to convert this interval-censored data into price indices we fit a log-normal distribution to the data, weighted by the total housing stock in each neighborhood, to calculate the median housing price of each Census tract. We then use these values to sort the neighborhoods by desirability, as represented by their median house price.

Summary statistics are presented in Table 1. Census tracts, representing neighborhoods, are the units of observation. The mean housing price is $386,161.30, and the average income for individuals is $73,330.55. The mean housing prices by neighborhood are presented in Figure 1. In general, the Census tracts in the center of the county have lower prices compared to the beach and mountainous areas. Focusing on the pollution terms, PM 2.5 and drinking water contamination have means of 1.04 and 33.17, respectively. The spatial distribution of PM 2.5 concentrations is given in Figure 2. Moving from the less developed southern portions of the county to the more developed northern area of the county demonstrates a clear increase in the gradient. Figure 3 presents the drinking water contamination map. There are higher levels contamination in the northern portions of the county, with the remaining areas having relatively even levels. Both of the pollutants demonstrate significant spatial heterogeneity in concentration levels, providing the variation necessary for clear economic identification.

3. Model

In this study we adapt the Epple and Sieg (1999) pure characteristics model (PCM) of household location choice using a constant elasticity of substitution (CES) specification for utility. This framework employs mixed discrete-continuous depictions of the choice set where households choose continuous quantities of physical housing characteristics in a discrete number of residential communities. In effect, households choose a neighborhood, thereby choosing public goods, g_j, which are a composite of local public goods including water quality and air pollution measures. Conditional on that location decision, a consumer also selects a house with the optimal level of housing services given housing prices.

The population of heterogeneous households differ in preferences (α) and income (y). These households are characterized by the joint distribution of income and tastes, F(α,y). Household preferences are defined over neighborhood quality/amenities, *g*, quantity of housing consumed, *q*, and a composite private good, *b*. In order to characterize a sorting equilibrium, prices, physical housing characteristics, amenities and location choices are all defined such that no household could improve its utility by moving, and each household exactly occupies one house.

Households are assumed to choose the neighborhood that maximizes their utility. A CES specification for preferences defines the utility that household *i* obtains from living in community *j* as:

$$V_{i,j} = \left\{ \alpha_i \left(G_j \right)^{\rho} + \left[\exp\left(\frac{\left(y_i \right)^{1-\nu} - 1}{1-\nu} \right) \exp\left(-\frac{\beta P_j^{\eta+1} - 1}{1+\eta} \right) \right]^{\rho} \right\}^{\frac{1}{\rho}}, \tag{1}$$

with $F(\alpha, y) \sim$ *lognormal*. The first term in this CES specification represents the utility households receive from neighborhood amenities, while the second term encompasses utility from the private good component of housing. This specification is convenient as the CES parameters are readily interpretable and can be easily compared to estimates from the existing literature.

The index of public goods, G, is defined as a linear index of amenities provided by each community $G_j = \gamma_1 g_{1,j} + \cdots + \gamma_{R-1} g_{R-1,j} + \xi_j$. Households agree on a common set of weights for the amenities in the index $(\gamma_1, \dots, \gamma_{R-1})$ but differ in their overall preferences for amenities relative to the private good components of housing and the numeraire (α_i). Of the R amenities in the index, R-1 are observable. Then $g_{R,j} = \xi_j$ represents the composite of public goods unobserved by the analyst but observed by the households. Note that the "error term" of the model enters into the indirect utility function in a non-additively separable manner. This gives rise to the "pure characteristics" nomenclature as utility is defined solely over the characteristics of communities and there is no idiosyncratic location-household-specific shock.

For the private good component households are assumed to share the same elasticity of substitution between amenities and private goods, ρ, and the same demand parameters for the private good components of housing: price elasticity of housing, η, income elasticity of housing, ν, and demand intercept, β.

Using this indirect utility function, Epple and Sieg (1999) derive three necessary conditions for equilibrium. The increasing bundles property implies that locations with higher prices have better amenities, while boundary indifference defines the income and preference combination (α,y) that makes households exactly indifferent between neighborhoods j and $j+1$. Stratification then requires that households in locations with higher rankings of the public good have higher income and stronger preferences for amenities.

Given that a household with income and preference combination (α,*y*) makes their location decision based on amenity provision, *g*, and house price, *p*, Ellickson's (1971) *single crossing* condition ensures the sorting restrictions described above hold. Specifically, in a vertical model households agree on the ranking of locations by overall quality and differ only in their preferences for said housing "quality" relative to the numeraire. Given this assumption, if the slope of an indirect indifference curve in (*g*,*p*) space is monotonically increasing in income (y|α) and preferences (α|y) then indifference curves in the (*g*,*p*) plane will satisfy single crossing in y and α. This ensures that households will sort into neighborhoods by income and taste preferences and implies a negative value for rho (ρ), the elasticity of substitution between amenities and private goods.

Additionally, given certain constraints[1] on the utility function a sorting equilibrium can be described by a hedonic price function. Namely, equilibrium prices are functionally related to housing characteristics and amenities $P_{n_j} = P(g_j, h_{n_j})$. Unlike the traditional hedonic model, there is no requirement that households be free to choose continuous quantities of each amenity nor is the market assumed to be perfectly competitive. Thus, we can no longer translate the price function gradient into measures of the marginal willingness to pay for amenities. However, Sieg et al. (2002) show that housing expenditures can be expressed as the product of a price index and a quantity index[2]

[1] If $U_i(g_i, h_{n_j}, b, \alpha_i)$ is continuously differentiable, monotonically increasing in the numeraire, and Lipschitz continuous.

[2] As long as a h_{n_j} enters utility through a separable sub-function that is homogeneous of degree 1.

(2) $$\ln P_{n_j} = \ln q(h_n) + \ln p(g_j),$$

allowing the neighborhood level prices $P_1,\ldots,P_j$ to be estimated as fixed effects in a hedonic regression using transactions data.

Estimation proceeds using the simulated two-stage generalized method of moments estimator developed by Sieg et al. (2004). In the first stage housing price estimates are treated as known constants in order to recover all of the structural parameters

(3) $$\theta = [\beta, \eta, \nu, \rho, \mu_\alpha, \mu_y, \sigma_\alpha, \sigma_y, \lambda, G_1, \gamma_1, \ldots, \gamma_{R-1}].$$

Following Sieg et al. (2004) all parameters can be recovered using moment conditions defined over income quartiles, expenditure quartiles and public goods. These moment conditions are given as

(4) $$m_j(\theta) = \begin{Bmatrix} \tilde{G}_j - \gamma_1 g_{1,1} - \ldots - \gamma_{R-1} \cdot g_{R-1,j} \\ y_j^{25} - \tilde{y}_j^{25} \\ y_j^{50} - \tilde{y}_j^{50} \\ y_j^{75} - \tilde{y}_j^{75} \\ \ln P_{n\in j}^{25} - \ln \beta - (\eta+1)\ln p_j - \nu \ln \tilde{y}_j^{25} \\ \ln P_{n\in j}^{50} - \ln \beta - (\eta+1)\ln p_j - \nu \ln \tilde{y}_j^{50} \\ \ln P_{n\in j}^{75} - \ln \beta - (\eta+1)\ln p_j - \nu \ln \tilde{y}_j^{75} \end{Bmatrix}.$$

The first moment condition is based on the level of amenity provision, where the public good is defined using a linear relationship between school quality and food access. Given a value for the cheapest community, G_1, the sorting behavior implied by vertical differentiation allows $G_2,...,G_J$ to be defined recursively. The predictions for $G_1,...,G_J$ are then used to identify the (constant) weights in the amenity index. The residual to the moment condition defines the composite unobserved amenity in each community $(\xi_1, \ldots, \xi_J)$ as the researcher does not perfectly observe g_j but instead $g_j + \xi_j$.

The next three moment conditions are based on the model's prediction for the distribution of income. Under the maintained assumptions on preferences, the information in θ can be used to simulate community-specific income distributions. Three of the moment conditions match the 25th, 50th, and 75th quantiles from the simulated distributions of income in each community $(\tilde{y}_j^{25}, \tilde{y}_j^{50}, \tilde{y}_j^{75})$ to their empirical counterparts $(y_j^{25}, y_j^{50}, y_j^{75})$. Income data from the 2010 census was used to create income quartiles for each neighborhood. Given that the data included the number of households in a series of income brackets, coefficients from a censored interval regression were used to estimate the 10th, 25th, 50th, 75th and 90th quantiles.

The last three moment conditions use the simulated income distributions to match predicted and observed quantiles from the distribution of housing expenditures in each community. The expenditure moments are obtained by multiplying the demand function by price and taking logs.

The mechanics of the simulated GMM estimator can be implemented using a Nelder-Mead algorithm. After solving for these parameters it is possible to estimate elements such as the relationship between taste for public goods and income preferences for public goods or elasticities of substitution. Additionally, this framework allows for estimation of the effect of an exogenous change, such as a public policy, by solving for a new PCM equilibrium when amenities are exogenous. As preferences in this model are "vertical", communities will always be ordered by their equilibrium housing prices and provisions of public goods: $p_1 < p_2 < ... < p_J$. Thus the problem can be reduced to a one-dimensional root finding problem as after a policy change the new equilibrium price ranking must be identical to the new ranking by *G*, allowing the new equilibrium price of housing in community 1 to be adjusted until the market clears in community *J*.

4. Results

Combining the ACS and environmental data, we estimate a vertical sorting model. These results can be found in Table 2. All demand parameters are of the expected sign and magnitude. Nu (ν) is positive, as an increase in income will lead to an increase in demand for housing, while eta (η) is negative as a higher price should result in reduced demand. The positive estimate of beta (β) is interpreted as price increases leading to decreases in demand, and the formulation above incorporates a negative sign. Additionally, the initial results provide evidence that households do sort across pollution burden levels. Specifically, the negative value of γ_1 demonstrates a consumer preference to live in neighborhoods with lower levels of drinking water contamination. Future analysis will allow us to calculate both marginal and general equilibrium willingness to pay values, as well as endogenize the potential welfare effects of proposed pollution-related public policies.

5. Conclusion

This research is unique in applying a vertical model of residential sorting to measure preferences for pollution at the micro-level. By combining CalEnviroScreen 2.0 and Census data, we are able to investigate the impact of local water and air quality pollution burdens to determine their impact on residential location choices. The initial results indicate that these pollution measures have significant impacts on sorting behaviors. There are many pathways for extending this research beyond the discussion presented here. For example, we intend to disaggregate the pollution burden measure to determine the drivers of the preference measured in these initial results. Additionally, the use of a sorting model allows for us to simulate the welfare effects of public policies. The CalEnviroScreen 2.0 data has been used to identify the 25% most burdened California communities, which are then targeted with funds allocated from the California cap and trade program. Further study would allow us to estimate the effect of this policy on resident welfare.

6. References

Bayer, Patrick, Ferreira, Fernando, and McMillan, Robert. 2007. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy.* 115 (4), 588–638.

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.

California Office of Environmental Health Hazard Assessment. *CalEnviroScreen 2.0.* Web. 25 Sept. 2015. http://oehha.ca.gov/calenviroscreen

Ellickson, B., 1971. "Jurisdictional Fragmentation and Residential Choice." *American Economic Review* 61(2): 334-339.

Epple, D., and H. Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107:645–681.

Hamilton, T.L., and D.J. Phaneuf. 2015. "An Integrated Model of Regional and Local Residential Sorting With Application To Air Quality." *Journal of Environmental Economics and Management* (74)71-93.

Public Health Institute. 2015. "Costs of Environmental Health Conditions in California Children." California Environmental Health Tracking Program. http://www.phi.org/CEHTPKidsHealthCosts

Sieg, H., V.K. Smith, H.S. Banzhaf, and R. Walsh. 2004. “Estimating the General Equilibrium Benefits Of Large Changes In Spatially Delineated Public Goods.” *International Economic Review* 45:1047–1077.

Tiebout, C.M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64(5): 416-24.

U.S. Census Bureau; 2013 American Community Survey; generated using American FactFinder; http://factfinder2.census.gov.

U.S. Census Bureau; Census 2010, Summary File 1; generated using American FactFinder; http://factfinder2.census.gov.

7. Tables

Table 1: Summary statistics

Variable	Mean	Std. Dev.
housing price	$386,161.30	$118,877.40
income	$73,330.55	$29,052.43
household size	3.0936	0.7942286
Census tract population	5,214.69	2,035.01
PM 2.5	1.041528	0.1484326
drinking water	33.17031	10.11265
Observations	575	

Table 2: Parameter estimates

Variable	Estimate
standard deviation of income	0.50457
mean alpha	1.0691
standard deviation of alpha	0.37842
lambda	0.016211
nu	2.2413
eta	-1.7225
beta	1.7112
rho	-0.05903
g_0	0.067479
$gamma_1$	-0.01474

8. Figures

Figure 1: Average housing price (darker is higher)

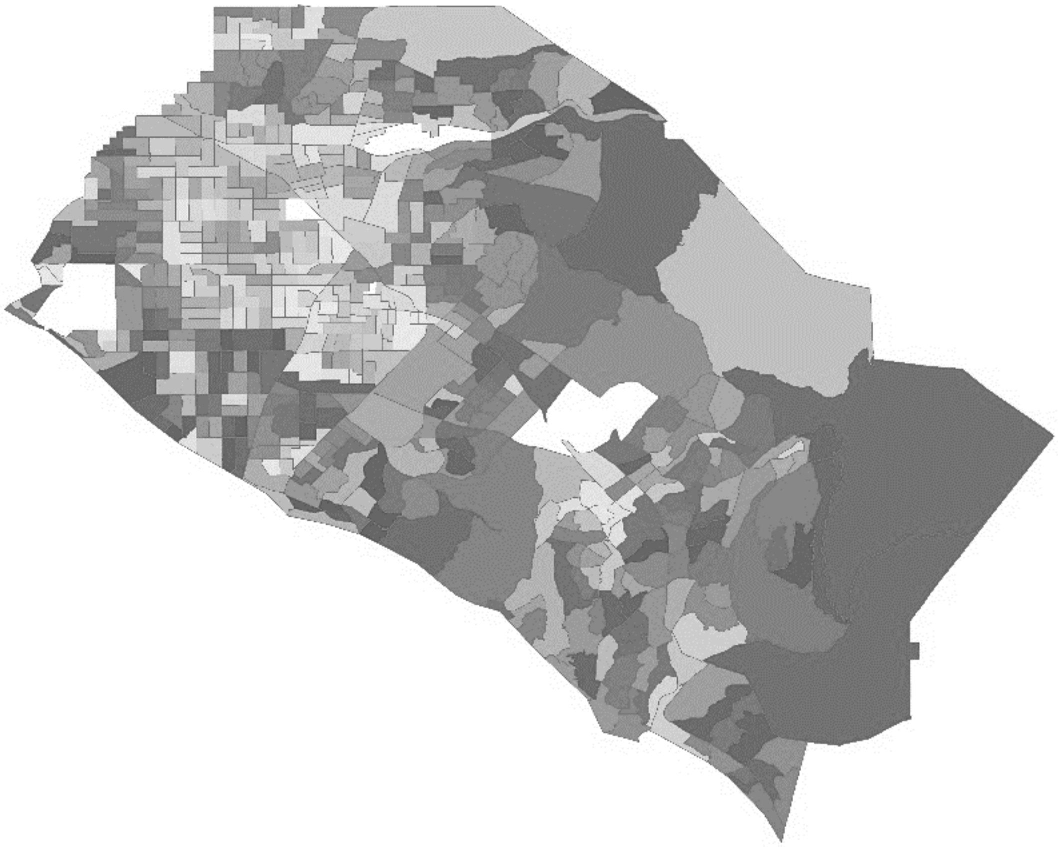

Figure 2: PM 2.5 concentrations (darker is higher)

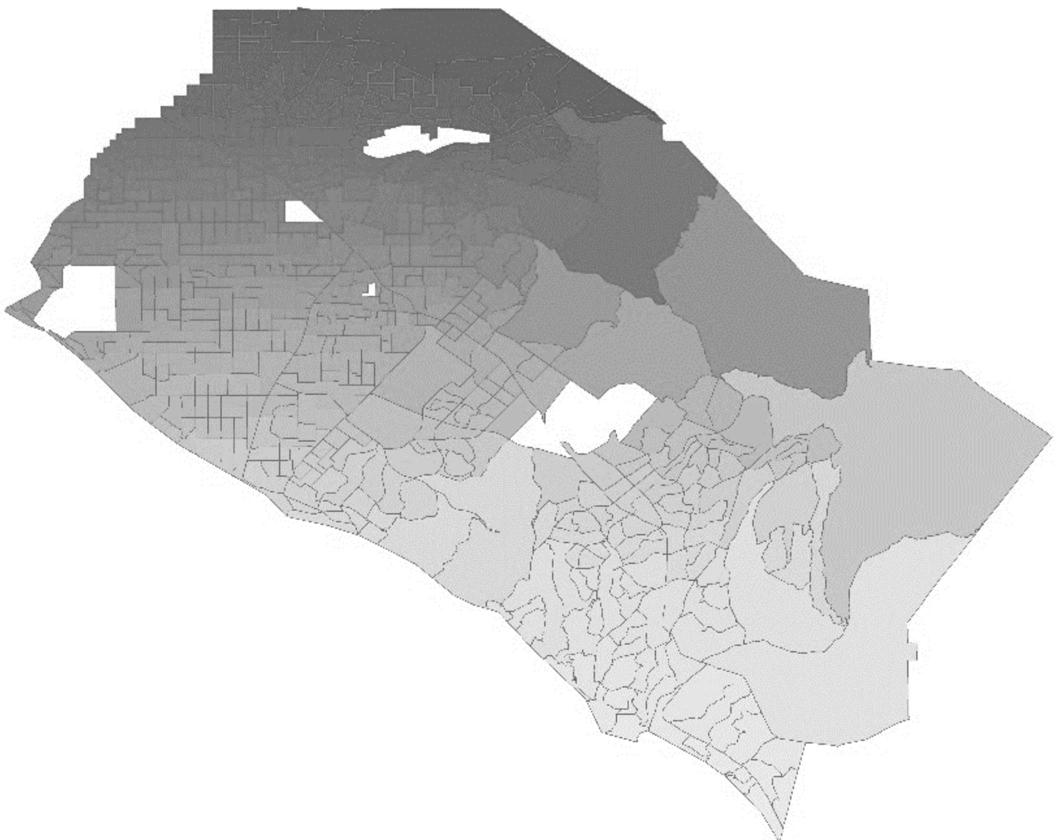

Figure 3: Drinking water contamination (darker is higher)