A Unified Framework for Measuring Preferences for Schools and Neighborhoods

JPE, 2007

Patrick Bayer, Fernando Ferreira, Robert McMillan

Research Question

RQ

How do economists get reliable estimates of household preferences for school and neighborhood attributes?

This paper develops a framework for recovering household preference for a broad set of school and neighborhood attributes in the presence of sorting.

LITERATURE

- Preference for schools and neighborhoods shapes the way that household sort in the housing market. (Tiebout, 1956; Epple and Zelenitz, 1981;...)
- Discrete choice model (McFadden, 1978) for sorting with household unobservables (Berry et al., 1995).

ENDOGENEITY

Hedonic price regression in traditional sense:

$$P_h = \beta \boldsymbol{X}_h + \xi_h$$

Sorting correlates to the household and neighborhood attributes, and induces to correlations among a host of neighborhood attributes, including those unobserved.
Hard to isolate variation in neighborhood sociodemographics uncorrelated with unobserved aspects of neighborhood and housing quality.

⇒ Acknowledge the limitation. (Cutler et al., 1999; Bajari and Kahn, 2005)
⇒ Identification in broader regions.
⇒ *Boundary discontinuity design* (BDD) with boundary fixed effect, embedded in our sorting model.

Structure of this paper

Mainly two parts:

1. Provide a strategy (Boundary discontinuity design, BDD) for addressing endogeneity of the hedonic price regression.
2. Build up and estimate a model of residential sorting. This provides an adjustment to the hedonic price regression.

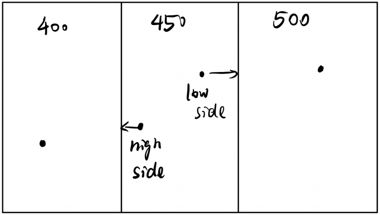
Contribution

1. It provides a strategy for addressing the endogeneity of school and neighborhood attributes in the context of heterogeneity sorting model.
2. It provides new estimates of household preference on schools and neighbors.

Data

- Restricted access version of the 1990 census.
- Household information + their house characteristics on block level.
- School attendance zone and school quality.
- House transaction data for robustness check.

How to define better/worse school attendance zone?



DISCONTINUITY

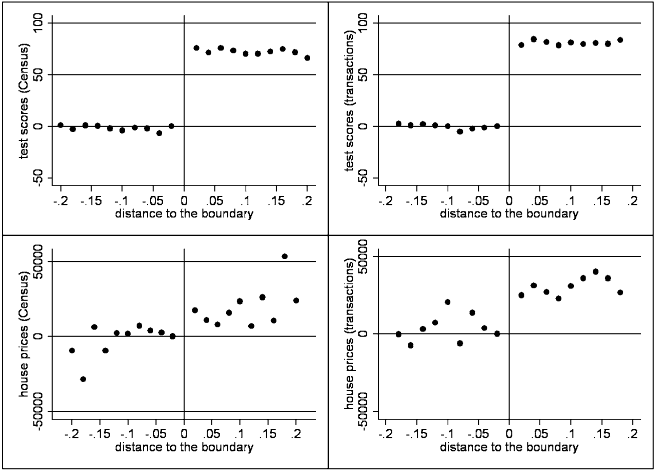


Fig1. Test scores and house prices around the boundary.

DISCONTINUITY

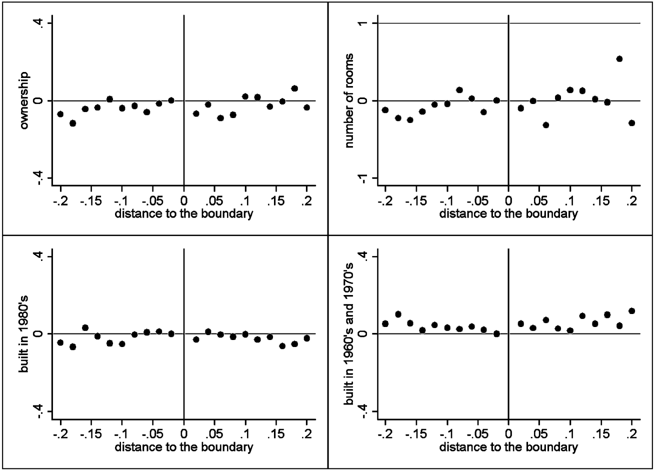


Fig2. Census housing characteristics around the boundary.

DISCONTINUITY

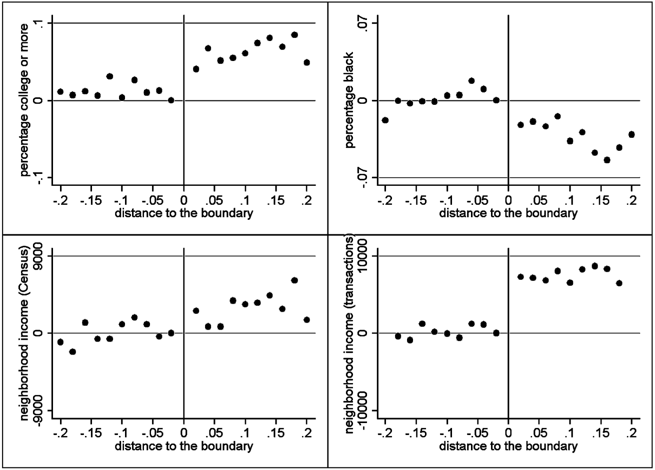


Fig4. Neighborhood sociodemographics around the boundary.

HEDONIC PRICE REGRESSION

$$p_h = \beta \boldsymbol{X}_h + \theta_{bh} + \xi_h \tag{1}$$

- p_h: price of house h.
- $\boldsymbol{X}_h$: housing and neighborhood characteristics.
- θ_{bh}: boundary fixed effects, if house h is within a specified distance of boundary b and zero otherwise.

TABLE 3
KEY COEFFICIENTS FROM BASELINE HEDONIC PRICE REGRESSIONS

	SAMPLE			
	Within 0.20 Mile of Boundary (N = 27,548)		Within 0.10 Mile of Boundary (N = 15,122)	
Boundary fixed effects included	No	Yes	No	Yes
	A. Excluding Neighborhood Sociodemographic Characteristics			
	(1)	(2)	(5)	(6)
Average test score (in standard deviations)	123.7 (13.2)	33.1 (7.6)	126.5 (12.4)	26.1 (6.6)
R^2	.54	.62	.54	.62

- When boundary fixed effect is included, the coefficient of test score declines 75%
- This implies the majority of the observed correlation is driven by he correlation of school quality with **other aspects of housing or neighborhood quality.**

	B. Including Neighborhood Sociodemographic Characteristics			
	(3)	(4)	(7)	(8)
Average test score (in standard deviations)	34.8	17.3	44.1	14.6
	(8.1)	(5.9)	(8.5)	(6.3)
% census block group black	−99.8	1.5	−123.1	4.3
	(33.4)	(38.9)	(32.5)	(39.1)
% block group with college degree or more	220.1	89.9	204.4	80.8
	(39.9)	(32.3)	(40.8)	(39.7)
Average block group income (/10,000)	60.0	45.0	55.6	42.9
	(4.0)	(4.6)	(4.3)	(6.1)
R^2	.59	.64	.59	.63

- After including neighborhood sociodemographic characteristics, coefficients reduce 50%.
- Coefficients on neighborhood race becomes insignificant after including neighborhood sociodemographics.
- However, this evidence **DOES NOT** implies households do not have strong racial preferences.

So far...

- So far, the evidence of sorting across school attendance zone boundaries is clear.
- This suggests that households vary in their willingness to pay for some features of schools and neighborhoods.

$$\Downarrow$$

How the coefficients in the hedonic price regressions relate to underlying household preferences?

The sorting model

Specification

Choice is discrete. Households choose the location of residence (House) by maximize their utility.

$$\max_{(h)} V_h^i = \alpha_X^i \boldsymbol{X}_h - \alpha_p^i p_h - \alpha_d^i d_h^i + \theta_{bh} + \xi_h + \epsilon_h^i. \qquad (2)$$

Each household's valuation of choice characteristics ($\alpha_j^i, j \in \{X, Z, d, p\}$) is allowed to vary with its own characteristics (z^i):

$$\alpha_j^i = \alpha_{0j} + \sum_{k=1}^{K} \alpha_{kj} z_k^i \qquad (3)$$

This essentially assumes household preferences are **heterogeneous**.

THE SORTING MODEL

ESTIMATION

A two-step estimation. (Barry et al., 1995)

Rewrite (2) as,

$$V_h^i = \delta_h + \lambda_h^i + \epsilon_h^i \tag{4}$$

, where δ_h measures mean indirect utility by choosing house h

$$\delta_h = \alpha_{0X}\mathbf{X}_h - \alpha_{0p}p_h + \theta_{bh} + \xi_h \tag{5}$$

ϵ_h^i is the unobservable household preference.
and

$$\lambda_h^i = \left(\sum_{k=1}^{K} \alpha_{kX} z_k^i\right) \boldsymbol{X}_h - \left(\sum_{k=1}^{K} \alpha_{kp} z_k^i\right) p_h - \left(\sum_{k=1}^{K} \alpha_{kd} z_k^i\right) d_h \tag{6}$$

?

The sorting model

MLE

Assuming that ϵ_h^i follows an extreme value distribution, choice probability can be written as,

$$P_h^i = \frac{\exp\left(\delta_h + \lambda_h^i\right)}{\sum_k \exp\left(\delta_k + \lambda_k^i\right)} \tag{7}$$

The likelihood function,

$$l = \sum_i \sum_h I_h^i \ln\left(P_h^i\right) \tag{8}$$

THE SORTING MODEL

After obtaining the estimation of δ_h... Recall equation (5):

$$\delta_h = \alpha_{0X}\mathbf{X}_h - \alpha_{0p}p_h + \theta_{bh} + \xi_h \tag{5}$$

Moving the price to the LHS,

$$p_h + \frac{1}{\alpha_{0p}}\delta_h = \frac{\alpha_{0X}}{\alpha_{0p}}\boldsymbol{X}_h + \frac{1}{\alpha_{0p}}\theta_{bh} + \frac{1}{\alpha_{0p}}\xi_h \tag{10}$$

Also, recall the hedonic price regression (1):

$$p_h = \beta\boldsymbol{X}_h + \theta_{bh} + \xi_h \tag{1}$$

Consequently, in the presence of heterogeneous preferences, the mean indirect utility δ_h provides an adjustment to the hedonic price equation, so that price regression accurately returns mean preference.

HEDONIC REG AND HOUSEHOLD PREFERENCE

How the coefficients in the hedonic price regressions relate to underlying household preferences?

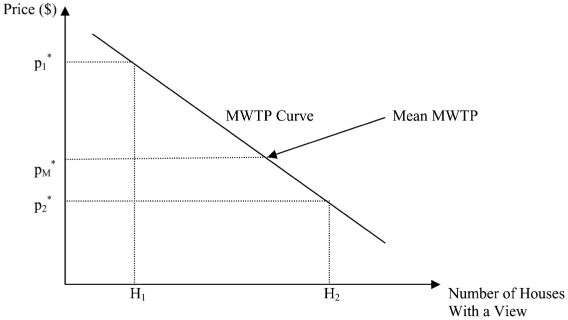


Fig5. Demand for a view of the Golden Gate Bridge

Hedonic reg and household preference

Demand curve (MWTP) ought to be horizontal if households have homogeneous preferences.

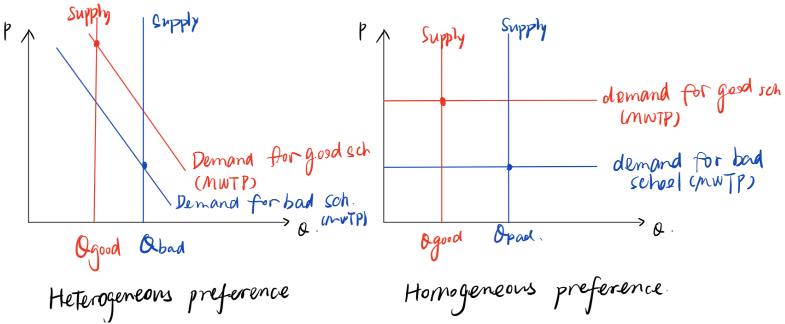


Fig5Plus. My further illustration

Hedonic reg and household preference

Hedonic price regression essentially assumes households' preferences are **homogeneous**! Why?

Recall (10),

$$p_h + \frac{1}{\alpha_{0p}}\delta_h = \frac{\alpha_{0X}}{\alpha_{0p}}\boldsymbol{X}_h + \frac{1}{\alpha_{0p}}\theta_{bh} + \frac{1}{\alpha_{0p}}\xi_h \tag{10}$$

δ_h degenerates to a constant when households' preferences are homogeneous. So we have the following equation,

$$\alpha_{0X}\boldsymbol{X}_h - \alpha_{0p}p_h + \theta_{bh} + \xi_h = K \Rightarrow p_h = \frac{\alpha_{0X}}{\alpha_{0p}}\boldsymbol{X}_h + \frac{1}{\alpha_{0p}}\theta_{bh} + \frac{1}{\alpha_{0p}}\xi_h \tag{11}$$

, which is exactly the hedonic regression in (1).

The sorting model

Estimation

In the estimation of (10) (The second step of the two-step estimation), use an IV to deal with the endogeneity. Go.

The predicted market-clearing prices for a version of the model that sets the vector of unobserved characteristics ξ_h to zero.

TABLE 7
DELTA REGRESSIONS: IMPLIED MEAN WILLINGNESS TO PAY
SAMPLE: WITHIN 0.20 MILE OF BOUNDARY (N = 27,458)

Boundary fixed effects included	No	Yes
	A. Excluding Neighborhood Sociodemographic Characteristics	
	(1)	(2)
Average test score (in standard deviations)	97.3	40.8
	(14.0)	(5.5)
	B. Including Neighborhood Sociodemographic Characteristics	
	(3)	(4)
Average test score (in standard deviations)	18.0	$\boxed{19.7}$
	(8.3)	(7.4)
% block group black	−404.8	−104.8
	(41.4)	(36.9)
% census block group Hispanic	−88.4	−3.5
% block group with college degree or more	183.5	104.6
	(26.4)	(31.8)
Average block group income (/10,000)	30.7	36.3
	(3.7)	(6.6)

- The estimated mean preferences for average test score are almost identical. ⇒ Hedonic capture the mean preference.
- Race coefficient becomes significantly negative. ⇒segregation!

Heterogeneous in preferences

TABLE 8
HETEROGENEITY IN MARGINAL WILLINGNESS TO PAY FOR AVERAGE TEST SCORE AND NEIGHBORHOOD SOCIODEMOGRAPHIC CHARACTERISTICS

		Neighborhood Sociodemographics		
	Average Test Score +1 SD	+10% Black vs. White	+10% College-Educated	Block Group Average Income +$10,000
Mean MWTP	19.69	−10.50	10.46	36.3
	(7.41)	(3.69)	(3.18)	(6.60)
Household income (+$10,000)	1.38	−1.23	1.41	.86
	(.33)	(.37)	(.21)	(.12)
Children under 18 vs. no children	7.41	11.86	−16.07	2.37
	(3.58)	(3.03)	(2.25)	(1.17)
Black vs. white	−14.31	98.34	18.45	−1.16
	(7.36)	(3.93)	(4.52)	(2.24)
College degree or more vs. some college or less	13.03	9.19	58.05	.31
	(3.57)	(3.14)	(2.33)	(1.40)

- Conditional on income, households prefer rich neighborhood.
- Household racial segregation
- Educational segregation.

Summary

- For those housing and neighborhood characteristics that vary continuously throughout the urban area, hedonic regression may be interpreted **not only a measure of implicit price of a particular attribute, but also an estimate of mean preference.**
- Hedonic price regression cannot tell the story of the role of race in the housing market, since the race do not vary continuously in the urban area due to household sorting issue (household segregation).
- The heterogeneous sorting model essentially demonstrate a compounding general equilibrium impact of exogenous school quality increase on housing price. Direct and indirect from sorting.

Appendix. 1

Robustness check for hedonic regression

- Distance to the boundary
- School characteristics versus immediate neighbors
- Top-coding of census prices
- Only owner-occupied units

Appendix. 2

The mean indirect utility

How to understand δ_h as the mean indirect utility?

$$\begin{aligned}\frac{\partial l}{\partial \delta_h} &= \sum_{i=h} \frac{\partial \ln\left(P_h^i\right)}{\partial \delta_h} + \sum_{i\neq h} \frac{\partial \ln\left(P_h^i\right)}{\partial \delta_h} \\ &= \sum_{i=h} \left(1 - P_h^i\right) + \sum_{i\neq h} \left(-P_h^i\right) \\ &= 1 - \sum_{i} \left(P_h^i\right) = 0.\end{aligned} \tag{9}$$

No house is systematically more attractive than other houses.

Appendix. 3

Hedonic regression and household preference

Why does δ_h degenerate to a constant when preference is homogeneous?

From (9), δ_h is determined for all h from MLE by

$$\sum_i (P_h^i) = 1$$

$\Rightarrow$ No house is systematically more attractive.

Additionally, if household have homogeneous preference, for any specific house, everyone's value is identical.
$\Rightarrow \delta_h = K$

Appendix. 4

Endogeneity in (10)

The key insight from the industrial organization literature is that the equilibrium price of any particular product will be affected not only by its own quality but also by the availability of products that are close substitutes for it.

The equivalent insight in a housing market context is that two identical houses in neighborhoods of identical quality may command very different prices, depending on how they are situated relative to other housing choices within the metropolitan area.