Estimating Preferences for Neighborhoods: Discussion of Bayer, Ferreira, and McMillan, JPE 2007

Nathan Schiff
Shanghai University of Finance and Economics

Graduate Urban Economics, Week 13
May 16, 2016

Administration

Midterm: referee reports due next class (5/23)

ALSO for 5/23: outline due for research idea

6/13: research proposal presentation

How do we measure preferences for local goods?

Use hedonic regressions, under assumption value is capitalized into housing prices

In many countries most important local good is school quality

$ln(price_{ia}) = \alpha + X'_{ia}\beta + \gamma * testScore_a + \epsilon_{ia}$
where i is house and a is school attendance zone

What is problem with this approach? What is Black and Bayer et. al. strategy?

Black, QJE, 1999

In a famous paper, Black (1999) shows that a border discontinuity approach can identify MWTP for school quality

In US, children go to school based on location; the set of locations corresponding to one school are called “attendance zones”

Basic idea of Black is to compare houses on both sides of attendance zone boundary–like RDD

Uses boundary fixed effects *and* test scores to identify MWTP

$$ln(price_{iab}) = \alpha + X'_{iab}\beta + K'_b\phi + \gamma * testScore_a + \epsilon_{iab}$$

Control-based method: key assumption is that unobservable neighborhood characteristics correlated with test scores are same on each side of border

Black 1999: streets and attendance districts

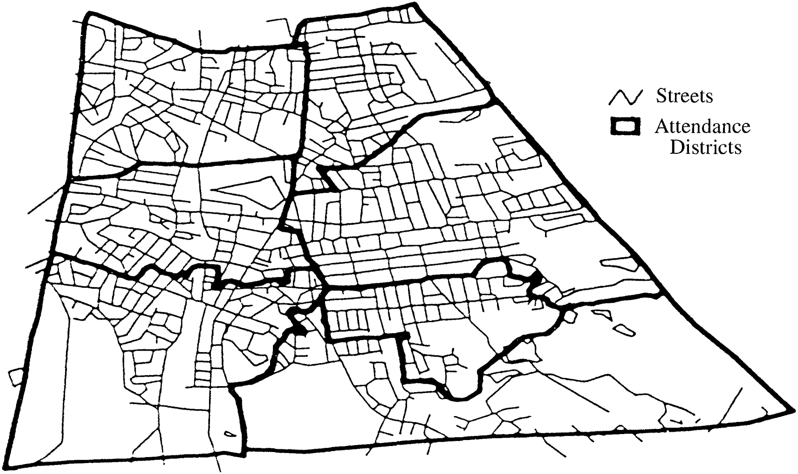


FIGURE I
Example of Data Collection for One City: Melrose
Streets, and Attendance District Boundaries

Black 1999: block groups and attendance districts

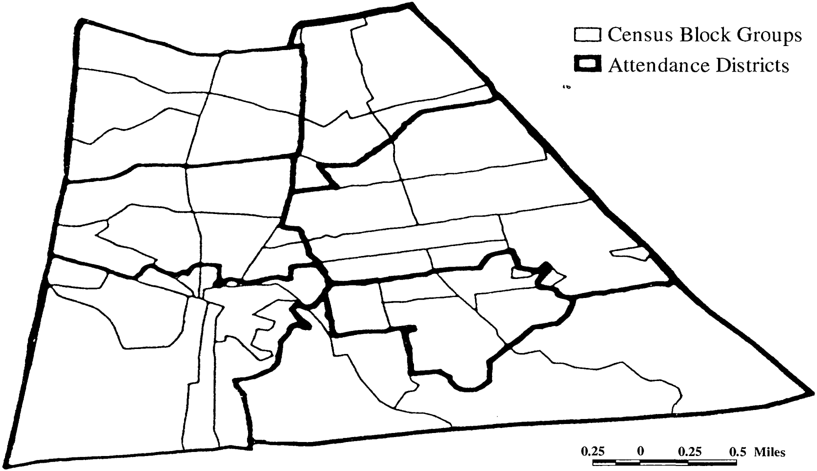


FIGURE II
Example of Data Collection for One City: Melrose
Census Block Groups and Attendance District Boundaries

Bayer, Ferreira, and McMillan, JPE 2007

BFM extend Black idea to estimate both 1) MWTP for school quality 2) MWTP for neighborhood demographics

BFM note that if demographics are still different along two sides of border (in narrow bands) then Black strategy leads to biased school quality coefficients

Two part paper:

First: estimate MWTP using hedonic regression

Estimate with very detailed, confidential, micro data of households (education, race, family structure) and houses (prices, rent, and housing), along with school characteristics and attendance zone boundaries in San Francisco area

Second: use structural model of location choice to adjust estimates to get *average* MWTP

Identification of MWTP for Demographics

An important question in the US is how much people value demographic characteristics of neighbors

For example, if whites hold prejudice against blacks then they will pay less to live in a neighborhood with more blacks

Another example: how much are people willing to pay to live with others of same education level?

Difficult questions to answer:

$$ln(price_{ij}) = \alpha + X'_{iab}\beta + \gamma * Demographic_j + \epsilon_{ij}$$

Demographics may always be correlated with unobserved neighborhood quality

How do BFM identify MWTP for demographics?

Using School Quality as Observable Source of Sorting

A key idea of BFM: school attendance zones cause demographic sorting; by controlling for observable school quality authors can control for unobservable neighborhood characteristics associated with demographics

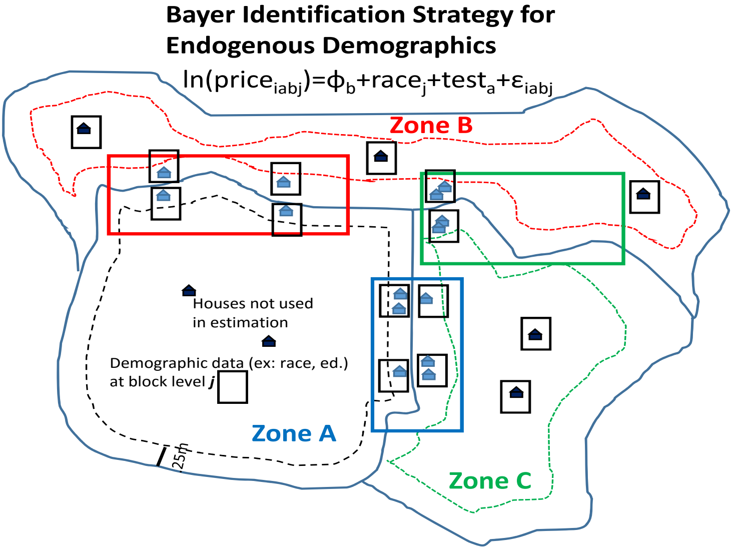
Ex: blacks in US have lower incomes and education on average than whites

This may lead to more blacks on lower test score side of school attendance zone (within same district)

By comparing value of houses along both sides of attendance zone border, where lower side has more blacks, and controlling for test scores, difference in housing value can give MWTP for living with higher black population

$$ln(price_{iabj}) = \alpha + X'_{iab}\beta + K'_b\phi + \gamma_1 * Demographic_j + \gamma_2 * testScore_a + \epsilon_{ij}$$

BFM 2007: Illustration of Border Discontinuity Design



BFM 2007: Border Discontinuity Design

$$ln(price_{iabj}) = \alpha + X'_{iab}\beta + K'_b\phi + \gamma_1 * Demographic_{ji} + \gamma_2 * testScore_a + \epsilon_{ij}$$

Key assumption: controlling for boundary fixed effects, test scores, and other area characteristics, demographic variables are no longer correlated with unobserved neighborhood characteristics affecting house values

First authors present evidence showing there is sorting of demographics on either side of attendance zone boundary

Then show how estimates of MWTP vary when include demographics and boundary fixed effects

Find that MWTP for school quality declines significantly when including boundary FE; declines even more when controlling for demographics

However, some demographics (% Black) are no longer significant when include boundary FE

Test Score RDD

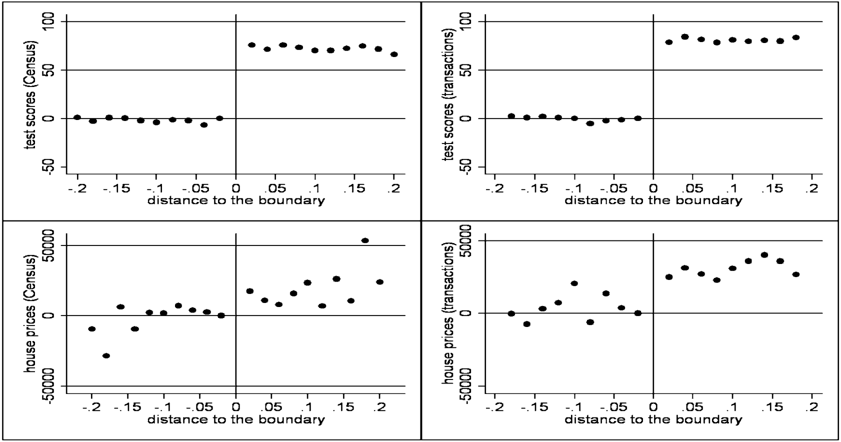
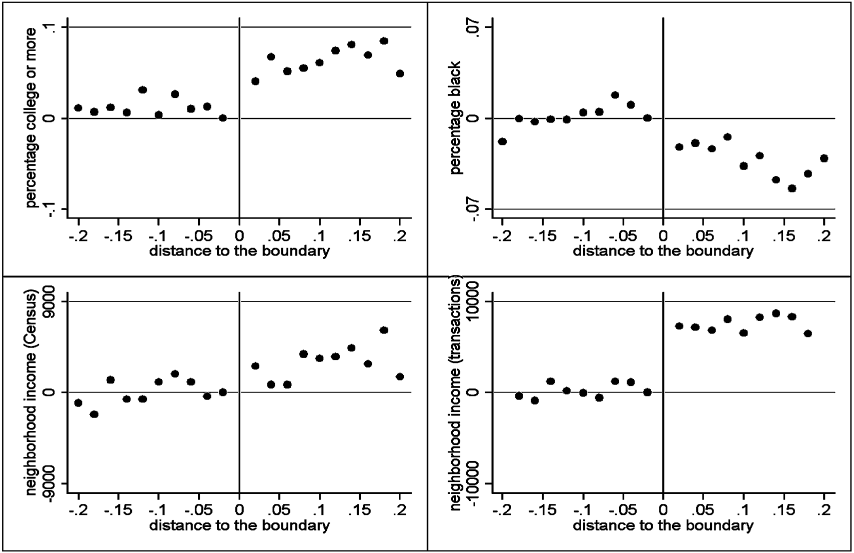


FIG. 1.—Test scores and house prices around the boundary. Each panel is constructed using the following procedure: (i) regress the variable in question on boundary fixed effects and on 0.02-mile band distance to the boundary dummy variables; (ii) plot the coefficients on these distance dummies. Thus a given point in each panel represents this conditional average at a given distance to the boundary, where negative distances indicate the low test score side.

Demographic sorting along boundary



MWTP Estimates

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	33.1	126.5	26.1
	(13.2)	(7.6)	(12.4)	(6.6)
R^2	.54	.62	.54	.62
	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

NOTE.—All regressions shown in the table also include controls for whether the house is owner-occupied, the number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and land use (% industrial, % residential, % commercial, % open space, % other) in 1-, 2-, and 3-mile rings around each location. The dependent variable is the monthly user cost of housing, which equals monthly rent for renter-occupied units and a monthly user cost for owner-occupied housing, calculated as described in the text. Standard errors corrected for clustering at the school level are reported in parentheses.

MWTP Estimates of Demographics

TABLE 4
HEDONIC PRICE REGRESSIONS: AVERAGE TEST SCORE, ALTERNATIVE SAMPLES
SAMPLE: WITHIN 0.20 MILE OF BOUNDARY

	NEIGHBORHOOD SOCIODEMOGRAPHICS			
	Excluded		Included	
	(1)	(2)	(3)	(4)
Boundary fixed effects included	No	Yes	No	Yes
Baseline results (N = 27,548)	123.7	33.1	34.8	17.3
	(13.2)	(7.6)	(8.1)	(5.9)
Schools versus immediate neighbors:				
A. Including school peer and teacher measures (N = 27,548)	95.0	32.1	31.5	22.6
	(17.9)	(10.4)	(9.3)	(8.5)
Alternative measures of neighborhood characteristics:				
B. Including block and block group measures (N = 27,548)			36.0	19.8
			(7.8)	(5.7)
C. Including block and alternative block group measures (N = 27,548)			33.7	23.8
			(7.3)	(5.6)
Other robustness checks:				
D. Dropping top-coded houses (N = 26,579)	86.6	29.5	20.3	16.1
	(9.9)	(6.6)	(7.7)	(5.7)
Only owner-occupied housing units:				
E. Using census-reported house value (N = 15,139)	64,891	14,874	27,883	9,376
	(7,474)	(3,197)	(5,047)	(2,460)
F. Using prices from transactions sample (N = 10,171)	34,262	12,210	14,208	9,176
	(4,958)	(3,108)	(2,886)	(2,738)

NOTE.—The dependent variable in specifications A–D is the monthly user cost of housing, which equals monthly rent for renter-occupied units and a monthly user cost for owner-occupied housing, calculated as described in the text; the dependent variable in specification E is the market value of the house self-reported in the census; the dependent variable in specification F is the transaction price reported in our transactions data set. Specifications A–E are based on our census sample and include controls for whether the house is owner-occupied, the number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and land use (% industrial, % residential, % commercial, % open space, % other) in 1-, 2-, and 3-mile rings around each location. Specification F is based on our transactions data set and includes the same controls as in the other specifications along with additional controls for square footage and lot size. Standard errors corrected for clustering at the school level are reported in parentheses.

Heterogeneity and MWTP

Coefficients on hedonic price regressions represent MWTP of *marginal* consumer

If consumers are heterogeneous then coefficients on a given attribute may represent MWTP of consumer who most values that attribute, not mean MWTP

BFM attempt to back out mean MWTP by using a model to first estimate heterogeneity of location choices

Illustration of MWTP Heterogeneity

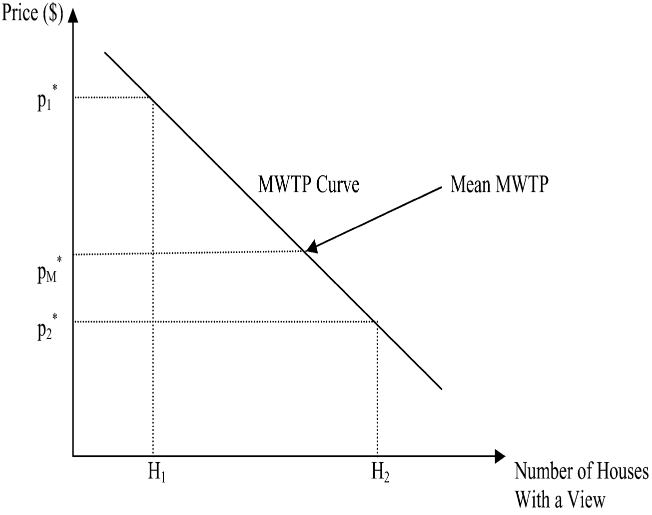


Fig. 5.—Demand for a view of the Golden Gate Bridge

Model of Residential Sorting

Household i chooses house h to maximize indirect utility:

$$\max_h V_h^i = \alpha_X^i X_h - \alpha_p^i p_h - \alpha_d^i d_h^i + \theta_{bh} + \xi_h + \epsilon_h^i \tag{2}$$

X_h represents vector of house characteristics (age, size) and neighborhood characteristics (demographics, crime)

p_h is price of house, d_h is distance from house h to work location of household i

θ_{bh} are boundary FE, equal to one if house h is within given distance of boundary b

ξ_h is unobserved characteristic of house h affects everyone equally; ϵ_h^i is EV Type 1 i.i.d. error

Preferences vary with household observables

$$\max_h V_h^i = \alpha_X^i X_h - \alpha_p^i p_h - \alpha_d^i d_h^i + \theta_{bh} + \xi_h + \epsilon_h^i \qquad (2)$$

Each coefficient on all characteristics of vector X_h, price p_h, and distance d_h allowed to vary with household characteristics (ex: race, education)

Specifically, for each characteristic j and household characteristics Z they allow:

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

Estimation

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

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

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

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

Two step estimation: first estimate 7) then estimate 5) with IV

Mean utility

Variable δ_h represents mean utility to all individuals of house *h*; it was estimated by first conditioning on individual observables

BFM show that by re-arranging it can yield a hedonic that gives *mean* MWTP

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

By estimating 10) coefficients represent mean MWTP across all different groups (population estimate)

Next class

Details of estimation: 1) how do they estimate first stage? How do they identify 2nd stage?

Results: how do mean MWTP compare to hedonic estimates?

How do racial preferences show up in estimates? Education preferences?

Implications or results?